Journal of Decision Systems

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tjds20

Anchoring bias in eliciting attribute weights and values in multi-attribute decision-making

Jafar Rezaei

To cite this article: Jafar Rezaei (2021) Anchoring bias in eliciting attribute weights and values in multi-attribute decision-making, Journal of Decision Systems, 30:1, 72-96, DOI: 10.1080/12460125.2020.1840705

To link to this article: https://doi.org/10.1080/12460125.2020.1840705

Published online: 03 Nov 2020.

JOURNAL OF DECISION SYSTEMS
2021, VOL. 30, NO. 1, 72–96
https://doi.org/10.1080/12460125.2020.1840705

Anchoring bias in eliciting attribute weights and values in multi-attribute decision-making

Jafar Rezaei

Technology, Policy and Management, Delft University of Technology, Delft, The Netherlands

ABSTRACT

The aim of this study is to look at anchoring bias – one of the main cognitive biases – in two multi-attribute decision-making methods, SMART and Swing. First, the existence of anchoring bias in these two methods for eliciting attribute weights and attribute values is theorised. Then, a special experiment is designed to compare the results estimated by the respondents and the *actual* results to measure potential anchoring bias. Data were collected from a sample of university students. The statistical analyses indicate the existence of anchoring bias in the two methods. It is also interesting to see that the impact of anchoring bias in estimates provided by the decision-makers on the obtained weights and values depends on the method that is used. These findings have significant implications for the actual decision-makers. Future research may consider the potential existence of cognitive biases in other multi-attribute decision-making methods and focus on developing mitigation strategies.

ARTICLE HISTORY
Received 3 June 2020
Accepted 17 October 2020

KEYWORDS
Multi-attribute decision-making; anchoring bias; SMART; Swing; debiasing

1. Introduction

Decisions are usually made on the basis of *evaluations* by decision-makers. Usually, these evaluations take the form of statements like 'this candidate is better than the others in communication', or 'this car generates the highest level of pollution'. Decision-makers draw conclusions by comparing alternatives in terms of different dimensions. In most cases, decision-makers have no access to (or do not use) objective figures, but instead rely on their subjective evaluations. Behavioural phycologists have found that people reduce the complexity of this task by using certain heuristics (Gilovich et al., 2002). For instance, representativeness is a heuristic (Kahneman & Frederick, 2002), which explains how, for instance, people use *categories* to decide whether or not a restaurant serves healthy food. There are several heuristics that people use (Gigerenzer & Gaissmaier, 2011; Gilovich et al., 2002). Most of the time, people are not aware of the role of heuristics in the way they make their decisions and they cannot deliberately control them. However, it is possible for people to identify and correct the resulting biases (Tversky & Kahneman, 1975). While generally speaking, using heuristics is extremely helpful in making decisions, sometimes they cause significant errors that can be very costly (Arkes, 1991). Those errors are called cognitive biases and they lead to biased decisions.

CONTACT Jafar Rezaei j.rezaei@tudelft.nl

Whereas cognitive biases have been discussed extensively in the areas of psychology (Gigerenzer, 1991; Hilbert, 2012; Kahneman et al., 1982; West et al., 2008), marketing (Fisher & Statman, 2000; Thomas et al., 2007), healthcare (Phillips-Wren et al., 2019), organisational studies (Das & Teng, 1999; Schwenk, 1984; Tetlock, 2000), business intelligence (Ni et al., 2019), and political science (Arceneaux, 2012; Rouhana et al., 1997), surprisingly enough, as also acknowledged by Montibeller and Von Winterfeldt (2015), we were able to identify only a few number of studies in the area of multi-attribute decision-making, most of which are theoretical.

Weber and Borcherding (1993) studied behavioural influences on eliciting weights in several weighting methods including SMART (simple multi-attribute rating technique) (Edwards, 1977), Swing (Von Winterfeldt & Edwards, 1986), and Tradeoff (Keeney & Raiffa, 1976) and argued that the attributes weights could be influenced by the choice of weighting method, the hierarchical structure of the problem, and the reference points. Their findings are supported by empirical evidence in the works they reviewed. They concluded that there is no criterion to determine the true weight as all weighting methods are biased and we cannot find which method is least biased. One of their recommendations to avoid systematic biases is to rely on multiple assessments. Buchanan and Corner (1997) investigated the effects of anchoring in two interactive weight elicitation methods (the Zionts and Wallenius method and the Naïve method) and reported significant anchoring bias in the former one. One important conclusion they made is that 'anchoring bias is affected by the structure of the solution method'. They found that the initial point in the two methods they studied matter and it biases the decision-maker. This fact, although seems undesirable, one could get benefit from it as one could bias the decision-makers in 'the right direction' (Buchanan & Corner, 1997). Jacobi and Hobbs (2007) studied the biases in eliciting weights by value tree and found that compared to hierarchical assessment, non-hierarchical assessment results in flatter, less varied weights which is a result of anchoring bias. In a similar vein, Hämäläinen and Alaja (2008), studied the influence of splitting the main attributes to different number of sub-attributes and found that by splitting an attribute to more sub-attributes, the main attribute in question gets higher weight, which is called splitting bias. Deniz (2020) studied three cognitive biases namely framing bias, loss aversion and status quo in a multi-attribute supplier selection problem. She proposed two pilot filters as debiasing strategies and showed the effectiveness of this strategy. Lahtinen and Hämäläinen (2016) studied the accumulation of biases in different stages of a multi-attribute decision-making problem (path dependence) and they showed that path dependency can occur in Even-Swaps method. In a later study, Lahtinen et al. (2020) proposed four debiasing techniques for multi-attribute decision-making: (i) 'Introducing a virtual reference alternative in the decision problem' (ii) 'Introducing an auxiliary measuring stick attribute' (iii) 'Rotating the reference point' (iv) 'Restarting the decision process at an intermediate step with a reduced set of alternatives'. They applied these techniques to Even-Swaps method to show the effectiveness of these techniques and recommended the same techniques for debiasing in SWING and Tradeoff. 'Rotating the reference point' is similar to the 'multiple anchors' debiasing strategy proposed by Montibeller and Von Winterfeldt (2015). This technique (rotating the reference point) is also embedded in best-worst method (BWM) (Rezaei, 2015, 2016) where a decision-maker conducts the pairwise comparison based on two opposite reference points (best and worst). As the two sets of pairwise comparisons are later used in one optimisation problem, possible anchoring bias and loss aversion bias could be cancelled-out (Rezaei, 2020). Ferretti (2020) provides a review of cognitive biases in spatial multi-attribute decision-making where the author investigates potential cognitive and motivational biases in designing spatial multi-attribute decision-making models and in the interpretation of their outcomes. One of the most recent and comprehensive studies in this area is conducted by Montibeller and Von Winterfeldt (2015), who reviewed a long list of cognitive biases in decision-making and risk analysis. They identified biases in different decision-making problems and recommended some debiasing strategies. They particularly discussed some biases (including splitting bias, equalising bias, gain-loss bias, proxy bias, range bias, insensitivity bias, desirability of options bias, and the affect influenced bias) in weight elicitation and recommended some debiasing strategies for attribute weight elicitation including cross-checking weights with different methods, using a group of decision-makers (if possible) instead of a single decision-maker, and avoiding the use of proxy attributes. They recommend more research to better understand the effect of cognitive biases on decision-making and risk analysis (two intertwined fields).

Multi-attribute decision-making has its roots in economics, mathematics, computer science, and behavioural psychology, and many of the assumptions and rules that are used in multi-attribute decision-making methods come from behavioural psychology (Köksalan et al., 2013). Although heuristics are usually considered as an alternative to rational decision-making (Gigerenzer & Gaissmaier, 2011), there are (claimed) rational decision-making tools that use some heuristics as their building blocks. Most existing multi-attribute decision-making methods need inputs from humans (decision-makers) who are prone to different cognitive biases, which suggests the conclusions we get from those methods are affected by those biased inputs. As truly pointed out by Weber and Borcherding (1993), being aware of the biases is the first and probably the most important step towards remedy the biases, which motivates us to examine the biases that might occur in these methods when using the heuristics mentioned earlier.

The ultimate aim of multi-attribute decision-making methods is to evaluate a number of alternatives with respect to a set of attributes. The evaluation task is performed in various ways in different methods. The way by which the evaluation task is performed might lead to some cognitive biases. The evaluation process has a starting point which could affect the following steps. This calls for investigating possible anchoring bias, as an important cognitive bias. The main contribution of this study is to theorise and examine anchoring bias, one of the most important cognitive biases, in eliciting attribute weights and values in two basic multi-attribute decision-making methods, SMART (Edwards, 1977) and Swing (Von Winterfeldt & Edwards, 1986). We selected these two methods because they have opposite starting points (SMART is using a low anchor while Swing is using a high anchor), which make them more interesting for our study. By theorising and experimentally investigating anchoring bias in these two methods, we contribute to the existing literature on cognitive bias in multi-attribute decision-making by finding that the weights assigned to the attributes are effected by the anchoring bias of the estimates provided by the decision-makers. We also found that although the two methods use two opposite anchors, the weights obtained by these two methods are affected by anchoring bias in a similar direction. The hypotheses and propositions which are used to formalise investigating anchoring bias can be extended for other multi-attribute decision-making methods. Studying anchoring bias in these methods paves the way for investigating those biases in other – more complex – methods. Measuring bias in multi-attribute decision-making methods has at least two important implications: (i) if a method is suffering from bias, its conclusions should be interpreted more carefully, and (ii) the research community should try to identify mitigation strategies to remedy any biases and develop methods that are less susceptible to them.

In the next section, we discuss anchoring bias. In Section 3, the two methods SMART and Swing are presented, and the research hypotheses are formulated in Section 4. Section 5 contains an experimental analysis to test the hypotheses and check the bias of the methods, followed by a discussion of the results in Section 6. Section 7 contains some suggestions for debiasing. The conclusions and avenues for further research are addressed in Section 8.

2. Anchoring bias

Anchoring-and-adjustment or briefly anchoring bias is a cognitive bias that refers to the tendency people have to heavily rely on the first piece of information (anchor) they receive or focus on, when judging and making decisions (Tversky & Kahneman, 1974, 1975). Decision-makers use that first piece of information to make an estimation, with subsequent adjustments being made on the basis of that initial estimation. The problem is that the involved adjustments tend to be *insufficient* and biased towards the initial estimation, which implies that, by changing the initial estimation, the decision-maker may come up with different adjustments. The initial estimation has two general sources: either decision-makers use an objective measure for the initial estimation, or they make a subjective estimation without relying on any other sources.

Several experiments in the literature of behavioural psychology show anchoring bias based on the two sources used to make the estimation. With regard to the first source, for instance, an experiment by LeBoeuf and Shafir (2006) involves a group of respondents who are shown a line and who are told that the line is shorter than 89 mm. They are then asked to extend the line so that it looks like it is 89 mm long. In a separate task, the respondents are shown a line and are informed that the line is longer than 89 mm and are asked to shorten it, so that it looks like it is 89 mm long. The respondents are not allowed to use any equipment, like a ruler. The results interestingly show that the average of the produced line extending the short anchor is 61.2 mm, while the average of the produced line shortening the longer anchor is 74.7 mm. This shows that anchoring bias, as the initial point, affects the final outcome. That is to say, when respondents base their adjustment on a shorter anchor, the final outcome is shorter than when they start with a longer line.

While in the study conducted by LeBoeuf and Shafir (2006), the respondents were provided with an initial estimation, the question is what happens if that initial estimation is not provided and people basically have to start from scratch, and make the estimation themselves and then perform the task at hand (expanding or shortening the line). One of the most well-known experiments of this kind was conducted by Tversky and Kahneman (1975), who asked one group of respondents to estimate (in five seconds) the product $8 \times 7 \times \ldots \times 1$, while the other group of respondents were asked to estimate (in five seconds) the product $1 \times 2 \times \ldots \times 8$. Interestingly, because the first estimation started with a higher number than the second one, the estimation of the first product was also higher than that of the second product. The ascending order of the estimates of the first group had a median of 512, while the descending order of the estimates of the second group had a median of 2250 (the correct answer is 40,320). So, again, we see an anchoring bias caused by *insufficient* adjustments.

The effects of anchoring have been studied in the context of many problems, including forecasting (Campbell & Sharpe, 2009; Cen et al., 2013; Hess & Orbe, 2013), voting (Yang et al., 2013) and negotiating (Wilson, 2012). For a review of different interesting experiments and real-world problems, we recommend the review paper by Furnham and Boo (2011). There also exist several debiasing strategies for anchoring bias (George et al., 2000; Montibeller & Von Winterfeldt, 2015; Mumma & Wilson, 1995).

3. Multi-attribute decision-making

Suppose that we have m alternatives $(i = 1, 2, \ldots, m)$ and let $a_k = (a_{k1}, a_{k2}, \ldots, a_{kn})$ be an alternative characterised by its outcomes with respect to n attributes $(j = 1, 2, \ldots, n)$. If $v_{kj}(a_{kj})$ represents the normalised value of a_{kj}, and w_j represents the weight of attribute j ($w_j > 0$ and $\sum_{j=1}^{n} w_j = 1$), an additive value function shows the total value of alternative a_k denoted as $v(a_k)$, as follows (Keeney & Raiffa, 1976).

$$v(a_k) = \sum_{j=1}^{n} w_j v_{kj}(a_{kj}). \tag{1}$$

Alternative a_k is preferred to alternative a_l if and only if $v(a_k)$ is greater than $v(a_l)$ or:

$$a_k \succ a_l \Leftrightarrow v(a_k) > v(a_l) \tag{2}$$

Determining the weights (w_j) is an essential part of any multi-attribute decision-making problem. Several attribute weight elicitation methods have been developed in the last decades including SAW (simple additive weighting) (Churchman & Ackoff, 1954), Tradeoff (Keeney & Raiffa, 1976), SMART (Edwards, 1977), AHP (analytic hierarchy process) (Saaty, 1977), Swing (Von Winterfeldt & Edwards, 1986), ANP (analytic network process) (Saaty, 1990), and BWM (best-worst method) (Rezaei, 2015, 2016). To see more attribute weight elicitation methods we refer to (Riabacke et al., 2012; Stewart, 1992; Triantaphyllou, 2000; Weber & Borcherding, 1993). Another essential part of multi-attribute decision-making problem is the identification of the attribute values a_{kj}, and their normalised values $v_{kj}(a_{kj})$. There exist several approaches to find the (normalised) attribute values, which are mainly discussed under value functions in literature (French, 1989; Ghaderi & Kadziński, 2020; Kakeneno & Brugha, 2017; Keeney & Raiffa, 1976; Kirkwood, 1997; O'Brien & Brugha, 2010; Rezaei, 2018; Von Winterfeldt & Edwards, 1986). Methods such as AHP, ANP, BWM, SMART, and Swing are also used to find the normalised attribute values $v_{kj}(a_{kj})$, when an attribute is evaluated subjectively (e.g., comfort) or when there is no access to the attribute values when they are objective (e.g., price). Several researchers have conducted comparative analysis between different sets of these methods and they have found that the methods produce different attribute weights or values for a same decision-maker and a same problem (Belton, 1986; Borcherding et al., 1991; Bottomley & Doyle, 2001; Olson et al., 1995; Pöyhönen & Hämäläinen, 2001; Schoemaker & Waid, 1982).

For the purpose of this study, we have decided to use SMART and Swing, two basic multi-attribute decision-making methods that can be used to identify the weights of the attributes and the normalised values of attributes based on the decision-maker's perception.

3.1. SMART

SMART (Edwards, 1977), in its original version, is used for the whole decision-making process, i.e. for determining the weights of the attributes, and also for ranking the alternatives. We are aware of the extensions of the method (Edwards & Barron, 1994); however, for the purpose of our study, the other extensions are not suitable, as we need to control the range effect.

In SMART, the decision-maker starts by ranking the attributes (or alternatives with respect to – w.r.t. – an attribute) in an ascending order. The decision-maker then assigns a score, like 10, to the least important attribute (or alternative w.r.t. an attribute), and higher numbers to the subsequent other attributes (or alternatives w.r.t. an attribute), after which a simple normalisation of the scores provides the weight of each attribute (or the normalised value of an alternative w.r.t. an attribute). That is to say, with n attributes (or m alternatives), if we consider the score the respondent gives to each attribute as $r_j, j = 1, \ldots, n$ (or to alternative k w.r.t. criterion j as $a_{kj}, k = 1, 2, \ldots, m, j = 1, 2, \ldots, n$) the weight of attribute f, w_f (or the normalised value of alternative k w.r.t. attribute $f, v_{kf}(a_{kf})$) is identified as follows.

$$w_f = \frac{r_f}{\sum_{j=1}^{n} r_j}, \tag{3}$$

$$v_{kf}(a_{kf}) = \frac{a_{kf}}{\sum_{i=1}^{m} a_{if}}. \tag{4}$$

SMART, as it is considered in our study, is similar to Min10 studied by Bottomley and Doyle (2001).

Example 1. Suppose that a company uses five attributes to select a supplier: price, quality, delivery, lead time, and commitment. If the company provides the following scores for the attributes:

$$r_{leadtime} = 10; r_{delivery} = 25; r_{quality} = 60; r_{committment} = 90; r_{price} = 100.$$

Then, the weight of 'lead time' is calculated as follows:

$$w_{leadtime} = \frac{10}{10 + 25 + 60 + 90 + 100} = 0.035.$$

The weight of the other attributes is identified in a similar way. So we have:

$$w_{leadtime} = 0.035; w_{delivery} = 0.088; w_{quality} = 0.210; w_{committment} = 0.316; w_{price} = 0.351.$$

If the company is evaluating three suppliers, the same procedure can be used to find the normalised attribute values for the suppliers. For instance, if the company provides the following scores for 'commitment' of the three suppliers:

$$a_{Supplier\ 1\ committment} = 10; a_{Supplier\ 2\ committment} = 50; a_{Supplier\ 3\ committment} = 90.$$

Then, the normalised values of commitment of the suppliers will be:

$$v_{Supplier\ 1\ committment}\left(a_{Supplier\ 1\ committment}\right) = \frac{10}{10 + 50 + 90} = 0.067,$$

$$v_{Supplier\ 2\ committment}\left(a_{Supplier\ 2\ committment}\right) = \frac{50}{10 + 50 + 90} = 0.333,$$

$$v_{Supplier\ 3\ committment}\left(a_{Supplier\ 3\ committment}\right) = \frac{90}{10 + 50 + 90} = 0.600.$$

So, as demonstrated above, SMART can be used to find the two main components of additive value function, Equation (1).

3.2. SWING

Swing (Von Winterfeldt & Edwards, 1986) considers the range of the attributes to determine the attribute weights or the normalised attribute values. For instance, to find the attribute weights, first the decision-maker is faced with an alternative which is characterised by the worst levels of the attributes. The decision-maker is then asked to change the level of an attribute to its best level, an attribute that she/he first would like to change. This attribute is labelled as the most important one and gets 100 points. The decision-maker is then asked to change the level of an attribute to its best level, a change which is the second most desirable for the decision-maker. The decision-maker assigns a point less than 100 to this attribute. The procedure continues until changing an attribute level to its best level which has the lowest desirability for the decision-maker. Normalising the points, the weight of each attribute is calculated. The same can be done to find the normalised attribute values. In this study, in order to control the range effect and to have a comparable method to SMART, we use a rather more straightforward version of the method (without considering the range), which makes it similar to Max100 as it has been considered in previous studies (Bottomley & Doyle, 2001).

When using Swing (the way we consider it in our study), the decision-maker first ranks the attributes (or alternatives w.r.t. an attribute) in descending order, and starts with the most important attribute (or alternative w.r.t. an attribute), assigning 100 points to that attribute (or alternative w.r.t. an attribute) and assigning lower scores to the subsequent attributes (or alternatives w.r.t. an attribute). Again, a simple normalisation provides the weights of the attributes (see Equation 3), or the normalised attribute values for the alternatives (see Equation 4).

Obviously, there is no difference in the calculation of the weights or the normalised attribute values in the two methods. The only difference is in the way we score the attributes or alternatives w.r.t. an attribute, which makes them interesting for our study.

With both methods, the decision-maker starts with a certain attribute (or alternative): the least important one (for SMART) and the most important one (for Swing), which draws our attention to the potential existence of anchoring bias in these methods.

4. Hypotheses development

In Section 2, anchoring bias explains that, if we start with a small anchor, the estimated score is smaller than its actual score, and if we start with a large anchor, the estimated score is larger than its corresponding actual score. Furthermore, several studies in the field of behavioural psychology show that, when the actual value is increasing (or decreasing) with respect to the anchor, the magnitude of anchoring bias will increase (or decrease) as well (Griffiths et al., 2015; Lieder et al., 2018). These scholars argue that anchoring bias is not caused by human irrationality, it is instead due to human resource-rationality, with the bias being a result of a rational trade-off between the time one needs to spend adjusting and the cost of error due to insufficient adjustment. This means that, if the error cost is low or the time-related cost is high, the magnitude of anchoring bias will increase even more as the actual value is increasing (decreasing) with respect to the anchor, compared to when the time cost is low or the error cost is high. In this study, we do not look at costs related to time or error, but to the increasing magnitude of anchoring bias as the actual value increases (or decreases).

In the case of SMART, respondents start with a small anchor. The expectation is that they will give the other attributes or attribute values a lower score than their actual values compared to the anchor. This bias is amplified as we move from less important attributes or smaller attribute values to more important attributes or larger attribute values. Because of this, we want to test the following hypothesis for the SMART method:

Hypothesis 1: *Using the SMART method, the estimated scores of the attributes or attribute values is smaller compared to their corresponding actual values and the difference increases as the actual value increases.*

In the case of Swing, on the other hand, the respondents start with a larger anchor, so they are expected to assign higher scores to the other attributes or attribute values compared to their corresponding actual values, a bias that is amplified as we move from more important attributes or larger attribute values to less important attributes or smaller attribute values. Consequently, we want to test the following hypothesis for the Swing method:

Hypothesis 2: *Using the Swing method, the estimated scores of the attributes or attribute values are higher compared to their corresponding actual values and the difference increases as the actual values decrease.*

If these two hypotheses are supported, we want to see the impact of anchoring bias in estimating the attributes or attribute values on the final normalisations that we use to obtain the attribute weights or the normalised attribute values in the two methods. In other words, in SMART, if the estimated scores are lower than their corresponding actual values (if we consider the actual score as r_j, the estimated score is $k_j r_j$, with a multiplier k_j, $0 < k_j \leq 1$), and as we move from the least important attribute (or the smallest attribute value) to the most important (or the largest) one, the proportional difference is becoming larger, the impact of this anchoring bias on the scores is:

Greater weights (or normalised values) for the less important attributes (or smaller attribute values) and lower weights (or normalised values) for the more important attributes (or larger attribute values) (see Figure 1).

As far as Swing is concerned, on the other hand, if the estimated scores are higher than their corresponding actual values (if we consider the actual score as g_j, the estimated score is $l_j g_j$, with a multiplier l_j, $1 \le l_j$), and as we move from the most important attribute (or the largest attribute value) to the least important (or the smallest) one, the proportional difference is becoming larger, the impact of this anchoring bias on the scores is:

Greater weights (or normalised values) for the less important attributes (or smaller attribute values) and lower weights (or normalised values) for the more important attributes (or larger attribute values) (see Figure 1).

Propositions 1 and 2 are presented to explain the impact of 'anchoring bias in the scores' on the final estimated attribute weights resulting from SMART and Swing, respectively.

Proposition 1: Having a set of ascending scores $r_j, j = 1, \ldots, n$, if we multiply each score r_j by a multiplier $0 < k_j \le 1$ belonging to a descending set, where $k_1 = 1$, there exists an index $j = b$ such that, for any $j \le b$, $\frac{k_j r_j}{\sum_{j=1}^{n} k_j r_j} \ge \frac{r_j}{\sum_{j=1}^{n} r_j}$, while, for $j > b$, $\frac{k_j r_j}{\sum_{j=1}^{n} k_j r_j} < \frac{r_j}{\sum_{j=1}^{n} r_j}$.

Proof: When the multipliers are all $0 < k_j \le 1$ with at least one $k_j < 1$, then $\sum_{j=1}^{n} k_j r_j < \sum_{j=1}^{n} r_j$ or $\sum_{j=1}^{n} k_j r_j = \rho \sum_{j=1}^{n} r_j$, where $\rho < 1$. We also have $k_n \sum_{j=1}^{n} r_j < \sum_{j=1}^{n} k_j r_j < k_1 \sum_{j=1}^{n} r_j$, which means that $k_1 > \rho > k_n$ or $1 > \rho > k_n$. Therefore, we have the descending vector $(k_1, \ldots, k_{b-1}, \rho, k_{b+1}, \ldots, k_n)$, which means that for $k_j \ge \rho$ (or equivalently $j \le b$) we have $\frac{k_j r_j}{\sum_{j=1}^{n} k_j r_j} \ge \frac{r_j}{\sum_{j=1}^{n} r_j}$, otherwise $\frac{k_j r_j}{\sum_{j=1}^{n} k_j r_j} < \frac{r_j}{\sum_{j=1}^{n} r_j}$.

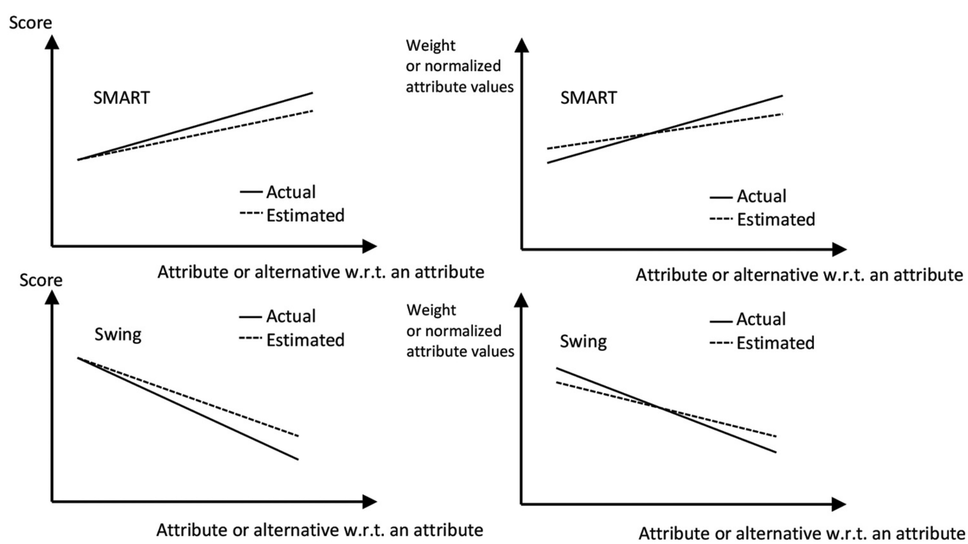

Figure 1. Actual vs estimated scores and weights or normalised attribute values in SMART and Swing.

Thus, we complete the proof of Proposition 1.

Corollary 1. As we obtain the actual attribute weights for SMART using $w_j^{act} = \frac{r_j}{\sum_{j=1}^{n} r_j}$, and the estimated weights by $w_j^{est} = \frac{k_j r_j}{\sum_{j=1}^{n} k_j r_j}$, using Proposition 1, we can see that, for the less important attributes we have $w_j^{est} \geq w_j^{act}$, while, for the more important attributes we have $w_j^{est} \leq w_j^{act}$.

Proposition 2: Having a set of descending scores $g_j, j = 1, \ldots, n$, if we multiply each score g_j by a multiplier $l_j \geq 1$ belonging to an ascending set, where $l_1 = 1$, there exists an index $j = c$ such that for any $j \leq c$, $\frac{l_j g_j}{\sum_{j=1}^{n} l_j g_j} \leq \frac{g_j}{\sum_{j=1}^{n} g_j}$, while for $j > c$, $\frac{l_j g_j}{\sum_{j=1}^{n} l_j g_j} > \frac{g_j}{\sum_{j=1}^{n} g_j}$.

Proof: When the multipliers are all $l_j \geq 1$ with at least one $l_j > 1$, then $\sum_{j=1}^{n} l_j g_j > \sum_{j=1}^{n} g_j$ or $\sum_{j=1}^{n} l_j g_j = \tau \sum_{j=1}^{n} g_j$, where $\tau > 1$. We also have $l_1 \sum_{j=1}^{n} g_j < \sum_{j=1}^{n} l_j g_j < l_n \sum_{j=1}^{n} g_j$, which means that $l_1 < \tau < l_n$ or $1 < \tau < l_n$. Therefore, we have the ascending vector $(l_1, \ldots, l_{c-1}, \tau, l_{c+1}, \ldots, l_n)$, which means that for $l_j \leq \tau$ (or equivalently $j \leq c$) then $\frac{l_j g_j}{\sum_{j=1}^{n} l_j g_j} \leq \frac{g_j}{\sum_{j=1}^{n} g_j}$, otherwise $\frac{l_j g_j}{\sum_{j=1}^{n} l_j g_j} > \frac{g_j}{\sum_{j=1}^{n} g_j}$.

Thus, we complete the proof of Proposition 2.

Corollary 2. As we obtain the actual attribute weights for Swing using $w_j^{act} = \frac{g_j}{\sum_{j=1}^{n} g_j}$, and the estimated attribute weights by $w_j^{est} = \frac{l_j g_j}{\sum_{j=1}^{n} l_j g_j}$, using Proposition 2, we can see that, for the less important attributes we have $w_j^{est} \geq w_j^{act}$, while, for the more important attributes we have $w_j^{est} \leq w_j^{act}$.

Propositions 3 and 4 are presented to explain the impact of 'anchoring bias in the scores' on the final normalised estimated attribute values resulting from SMART and Swing, respectively.

Proposition 3: Having a set of ascending scores $a_{if}, i = 1, \ldots, m$, if we multiply each score a_{if} by a multiplier $0 < q_i \leq 1$ belonging to a descending set, where $q_1 = 1$, there exists an index $i = d$ such that, for any $i \leq d$, $\frac{q_i a_{if}}{\sum_{i=1}^{m} q_i a_{if}} \geq \frac{a_{if}}{\sum_{i=1}^{m} a_{if}}$, while, for $i > d$, $\frac{q_i a_{if}}{\sum_{i=1}^{m} q_i a_{if}} < \frac{a_{if}}{\sum_{i=1}^{m} a_{if}}$.

Proof: Similar to the proof of Proposition 1.□

Corollary 3. As we obtain the actual normalised attribute values for SMART using $v_{kf}(a_{kf})^{act} = \frac{a_{kf}}{\sum_{i=1}^{m} a_{if}}$, and the estimated normalised attribute values by $v_{kf}(a_{kf})^{est} = \frac{q_i a_{if}}{\sum_{i=1}^{m} q_i a_{if}}$, using Proposition 3, we can see that, for the alternatives with a smaller attribute values we have $v_{kf}(a_{kf})^{est} \geq v_{kf}(a_{kf})^{act}$, while, for the alternatives with larger attribute values we have $v_{kf}(a_{kf})^{est} \leq v_{kf}(a_{kf})^{act}$.

Proposition 4: Having a set of descending scores $a_{if}, i = 1, \ldots, m$, if we multiply each score a_{if} by a multiplier $p_1 \geq 1$ belonging to an ascending set, where $p_i = 1$, there exists an index $i = e$ such that for any $i \leq e$, $\frac{p_i a_{if}}{\sum_{i=1}^{m} p_i a_{if}} \leq \frac{a_{if}}{\sum_{i=1}^{m} a_{if}}$, while for $i > e$, $\frac{p_i a_{if}}{\sum_{i=1}^{m} p_i a_{if}} > \frac{a_{if}}{\sum_{i=1}^{m} a_{if}}$.

Proof: Similar to the proof of Proposition 2.□

Corollary 4. As we obtain the actual normalised attribute values for Swing using $v_{kf}(a_{kf})^{act} = \frac{a_{kf}}{\sum_{i=1}^{m} a_{if}}$, and the estimated normalised attribute values by $v_{kf}(a_{kf})^{est} = \frac{p_i a_{if}}{\sum_{i=1}^{m} p_i a_{if}}$, using Proposition 4, we can see that, for the alternatives with smaller attribute values we have $v_{kf}(a_{kf})^{est} \geq v_{kf}(a_{kf})^{act}$, while, for the alternatives with larger attribute values we have$v_{kf}(a_{kf})^{est} \leq v_{kf}(a_{kf})^{act}$.

Figure 1 shows the expected scores and weights or normalised attribute values based on the developed hypotheses for the two methods.

We can conclude from the propositions that, for both methods, anchoring bias should follow the same behaviour for both types of scores – attribute scores or alternative scores w.r.t. an attribute. It should also behave the same for both types of normalised values – attribute weights or normalised attribute values for the alternatives. This similarity is also depicted in Figure 1.

5. Experimental studies

A real-world decision-making problem usually involves qualitative and quantitative attributes, with an attribute weight elicitation method identifying the weights based on the preferences provided by the decision-maker. In many cases, it is usual for different decision-makers to present different weights, based on their different preferences. For instance, while one decision-maker may assign a high importance to price when deciding on buying a car, that same attribute may be of less importance to another decision-maker. This is the same when evaluating the attribute values. That is to say, while a buyer might give 50 (out of 100) points to the comfort of a particular car, another buyer might give 30 to the comfort of the same car. This implies that there is not a universal actual weight or a universal attribute value for different decision-makers. For the purpose of our experimental setting, however, we need to have universal actual weights or attribute values, which is why we designed a special experiment, such that we are able to compare the attribute values (and the normalised attribute values) obtained from the respondents to the *actual* attribute values (and the normalised attribute values) and measure potential anchoring bias. Using such references instead of creating a case study could avoid potential principal-agent issues in interpreting the values (Grossman & Hart, 1992). As a natural consequence, we consider objective attribute (in our experiment, size) for alternatives. However, in our view, if we determine anchoring bias in such cases, it stands to reason to expect a similar bias in comparing attributes and also alternatives with respect to subjective attributes in real-world decision-making problems. Our study takes the first step in this regard and future research could build up on this to explore all the other possibilities (i.e. valuation and trade-off tasks).

In this study, we used students as the participants in the experiment, which is very common for this type of research (see, for instance (Buchanan & Corner, 1997; Hämäläinen & Alaja, 2008; Schoemaker & Waid, 1982)) as we could somehow control variables such as age, cognitive abilities, prior knowledge, and experience. It is also almost impossible to find a common real decision for a relatively large sample of real decision-makers like company managers. Two hundred and sixty-eight (mostly second year undergraduate) students from Delft University of Technology participated in the experiment. The students participated voluntarily with no extra credit for their participation, which decreases the chance of participation by unmotivated students to the case (Hämäläinen & Alaja, 2008). The experiment follows a Within-Subjects design implying that a student participates in both methods. Data were collected in two separate sessions, where the students were randomly assigned to one of the two methods (SMART or Swing) in the first session, while they were subsequently assigned to the other method in the following session.

In the experiment, we give the respondents a geographical map including five European countries (Figure 2) and ask them to score the countries based on their size (surface area) following the steps of the two methods. The students were asked to consider only the size in terms of surface area that they see on the picture and not other measures such as population. On the page containing the SMART method, the respondents are asked:

Select the smallest country and give it the minimum score 10, then give a greater score to the second smallest country, and continue until you assign a score to the biggest country.

On the page containing the Swing method, the respondents are asked:

Select the biggest country and give it the maximum score 100, then give a smaller score to the second biggest country, and continue until you assign a score to the smallest country.

Two hundred and sixty-eight students participated in the experiment, excluding the ones with missing data, 207 responses were found useful for the analysis.

6. Results

6.1. Results considering the scores

This section contains the analysis based on the collected data (the scores, not the normalised values). As mentioned before, in SMART, a score a_{if} is given to each alternative (country) with respect to an attribute (size). We also have the exact values for these alternatives. Table 1 shows the exact actual values.

The numbers in Table 1 are the actual numbers, which are not shown to the respondents. The map scale used on the paper for the five countries is $1{:}10^6$ which means 1 *cm* on the map equals 100 *km* on the ground. Simply normalising the actual scores gives us the normalised actual values of the alternatives w.r.t. the attribute size.

To check for anchoring bias, we start by converting the numbers in Table 1 into numbers that can be used in the two methods (SMART and Swing). This is done based on the actual values, not the respondents' estimated scores. That is to say, since 10 is used for the minimum score according to SMART, the converted score of each alternative is calculated as follows.

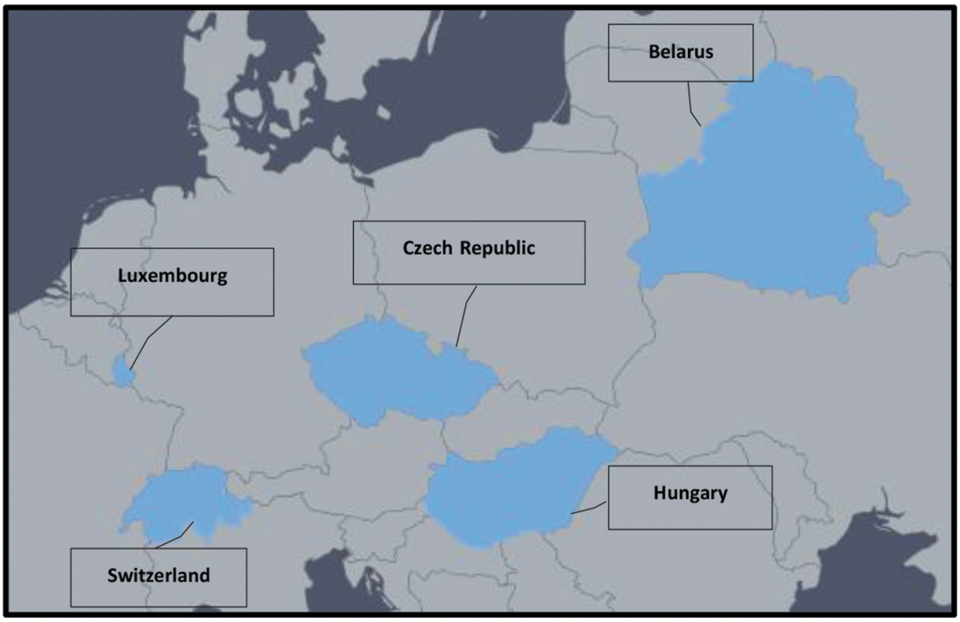

Figure 2. The map provided for the experiment.

$$Converted\ score_k(SMART) = \frac{Original\ score_k}{\min_i\{Original\ score_i\}} \times 10. \quad (5)$$

For Swing, where 100 is used for maximum score, the following formula is used:

$$Converted\ score_k(Swing) = \frac{Original\ score_k}{\max_i\{Original\ score_i\}} \times 100. \quad (6)$$

Table 2 shows the converted scores for all the alternatives of the experiment for the two methods.

Figure 3 shows the actual converted scores and the mean estimated scores (and 95% confidence interval) based on SMART data obtained from the sample.

In line with what we discussed earlier, the results show that the estimated scores are much lower than the actual scores. To test Hypothesis 1, comparing the scores obtained from the respondents to the converted actual ones using t-test we could find the potential anchoring bias. Table 3 shows the comparison results for SMART.

As we can see from Table 3, all the significant differences are negative, which means the respondents assign smaller scores to the alternatives compared to their actual

Table 1. The actual values of the set of alternatives.

Countries	Actual size (km^2)	Actual normalised value
Hungary	93,030	0.21974
Czech Republic	78,866	0.18629
Belarus	207,595	0.49035
Switzerland	41,285	0.09752
Luxembourg	2586	0.00611

Table 2. Converted scores for all the alternatives.

Countries	SMART	Swing
Hungary	360	44.8
Czech Republic	305	38
Belarus	803	100
Switzerland	160	19.9
Luxembourg	10	1.2

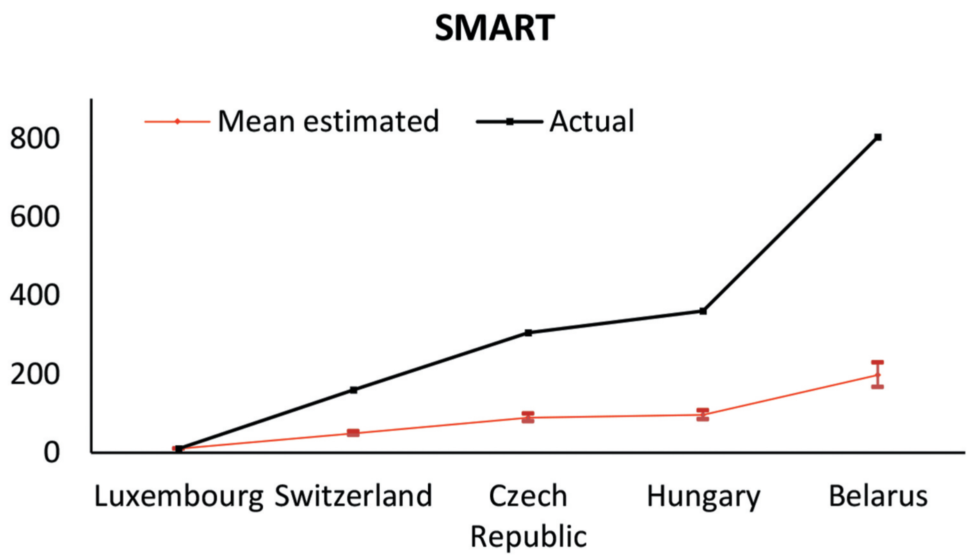

Figure 3. Actual vs mean estimated scores (and 95% confidence interval) of the countries (SMART).

Table 3. *T*-test results for the experiment considering the SMART scores.

					95% Confidence interval of the difference		
	t	df	Sig. (2-tailed)	Mean difference	Lower	Upper	Bias index
Hungary	−46.1	206	0.000	−263.8	−275.1	−252.5	−0.73
Czech	−43.7	206	0.000	−215.9	−225.6	−206.1	−0.71
Belarus	−38.1	206	0.000	−605.3	−636.3	−573.95	−0.75
Switzerland	−48.2	206	0.000	−110.8	−115.3	−106.2	−0.69
Luxembourg		206		0.000	0.000	0.000	

corresponding scores when using the SMART method, which means Hypothesis 1 is confirmed. The explanation is that, because the respondents start with a small alternative, assigning it the number 10, they tend to give smaller scores to the other alternatives in that category (compared to their corresponding actual values). Obviously, for the smallest alternative (the anchor, Luxembourg), the difference between the converted actual value and the obtained score is zero, which means *t* cannot be computed because the standard deviation is 0.

In this study, we also develop a *bias index* to check the potential bias, as follows:

$$Bias\ index_k(SMART) = \frac{mean\ difference_k}{Converted\ score_k(SMART)}. \tag{7}$$

For instance, for Hungary, the Bias index (using the data from Tables 2 and 3) is calculated as follows.

$$Bias\ index_{Hungary}(SMART) = \frac{mean\ difference_{Hungary}}{Converted\ score_{Hungary}(SMART)} = \frac{-263.8}{360} = -0.73$$

The bias index for Hungary ($-$ 0.73) means that, on average, respondents assign a score to Hungary that is 73% lower than its actual value. The Bias index for the other significant differences are calculated and presented in the last column of Table 3, which shows that the Bias index for the significant differences is all negative and high (all around 70%). Another interesting observation is that, as we move from the smaller countries to the larger countries, the absolute value of bias Index tends to increase, which confirms Hypothesis 1.

Figure 4 shows the actual converted scores and the mean estimated scores based on Swing data obtained from the sample.

It is interesting to see that, in line with our discussion above, the results show that the estimated scores are much higher than the actual values. To test Hypothesis 2, comparing the scores obtained from the respondents to the converted actual ones using *t*-test, we were able to identify the potential anchoring bias. Table 4 shows the comparison results for Swing.

Table 4 shows that all the significant differences are positive, which means the respondents assign higher scores to the alternatives compared to their actual values when they are applying the Swing method, which confirms Hypothesis 2. The explanation is that, when respondents start with a large alternative and assigning it the number 100, they tend to assign higher numbers (compared to their corresponding actual value) to the other alternatives. Obviously, the difference between the obtained scores and the actual

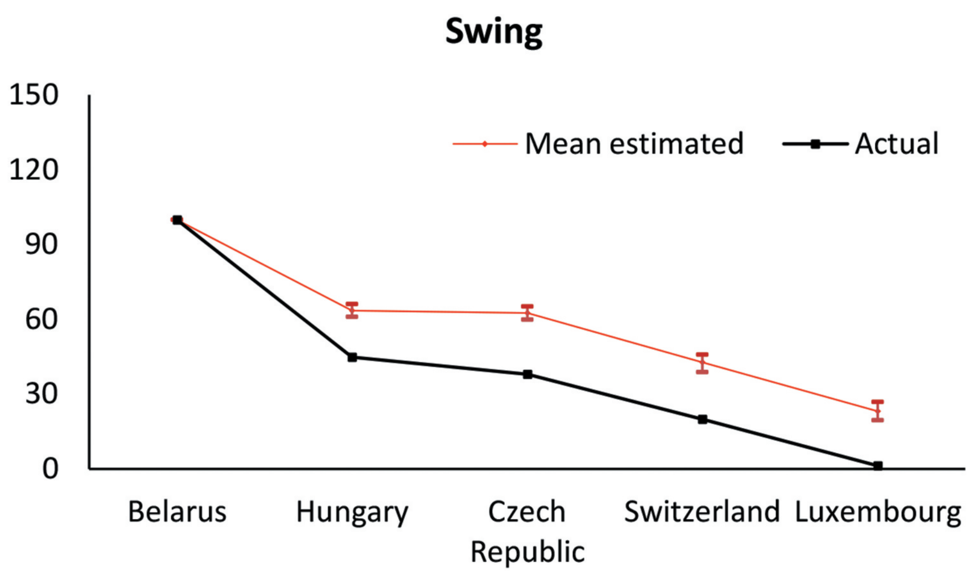

Figure 4. Actual vs mean estimated scores (and 95% confidence interval) of the countries (Swing).

Table 4. *T*-test results for the experiment considering the Swing scores.

	t	df	Sig. (2-tailed)	Mean difference	95% Confidence interval of the difference Lower	Upper	Bias index
Hungary	14.2	246	0.008	18.7	16.1	21.2	0.42
Czech	18.6	246	0.000	24.5	21.9	27.1	0.65
Belarus		246		0.000	0.000	0.000	
Switzerland	14.8	246	0.000	22.9	18.9	25.9	1.15
Luxembourg	11.9	246	0.000	21.9	18.3	25.5	18.25

scores is not significant for the largest alternative (anchor, Belarus). The bias index can be calculated using the following formula.

$$Bias\ index_k(Swing) = \frac{mean\ difference_k}{Converted\ measure_k(Swing)}. \tag{8}$$

For instance, for Hungary, the bias index (using the data from Tables 2 and 4) is calculated as follows.

$$Bias\ index_{Hungary}(Swing) = \frac{mean\ difference_{Hungary}}{Converted\ measure_{Hungary}(Swing)} = \frac{18.7}{44.8} = 0.42$$

The bias index for Hungary (0.42) means that, on average, respondents assign a score to Hungary that is 42% higher than its actual value. The Bias index for the other significant differences is calculated and presented in the last column of Table 4. The bias index for the significant differences is positive in all cases. However, unlike the case of SMART, here the Bias index varies a lot ranging from 0.42 to 18.25 (see Section 6.3 for other explanations). In this case, contrary to what we found in the case of SMART, as we move from the larger countries to the smaller countries the value of bias Index tends to increase, which confirms Hypothesis 2.

The statistical analysis shows the existence of anchoring bias in the two methods confirming both hypotheses. In the next section, we analyse the normalised values (which are obtained based on the collected scores).

6.2. Results considering the normalised values

In this section, we compare the normalised values resulting from the two methods, SMART and Swing, to the actual normalised values in the experiment. First, a normalisation is performed to determine the actual normalised values of the alternatives used in the experiment.

To identify the anchoring bias, we compare the normalised values obtained from the respondents based on the two methods (SMART and Swing) using Equation (4), to the actual normalised values (see Table 1) using *t*-test.

Figure 5 shows the actual normalised values and the mean estimated normalised values (and 95% confidence interval) based on SMART data obtained from the sample.

It is interesting to see that the mean estimated normalised values of the smaller alternatives are higher than their actual normalised values, while the mean estimated normalised values of the larger alternatives are lower than their actual normalised values. The results of *t*-test for SMART normalised values versus the actual normalised values for the experiment are presented in Table 5.

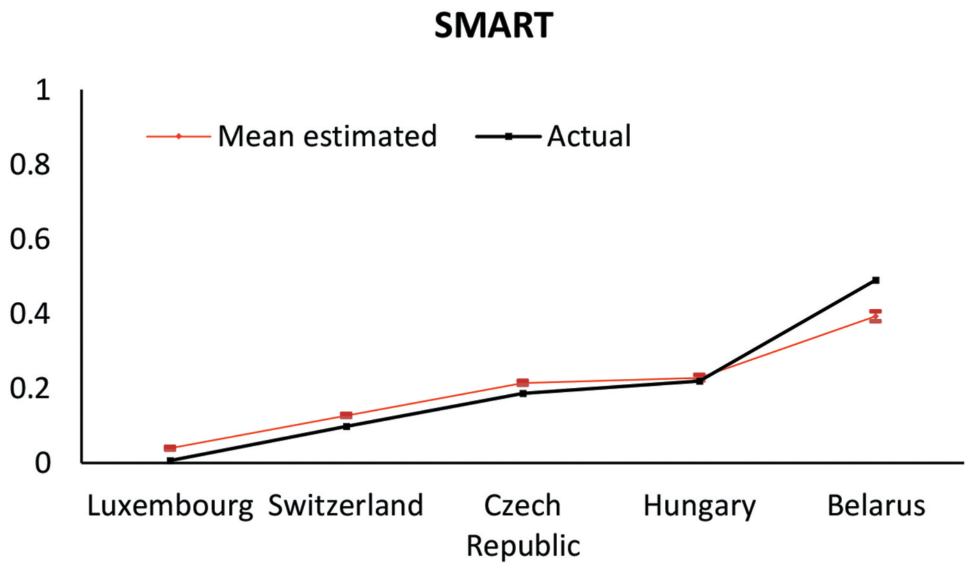

Figure 5. Actual vs mean estimated normalised values (and 95% confidence interval) of the countries (SMART).

Table 5. *T*-test results for the SMART experiment.

	t	df	Sig. (2-tailed)	Mean difference	95% Confidence interval of the difference	
					Lower	Upper
Hungary	2.10	206	0.037	0.0079	0.0005	0.0153
Czech	10.96	206	0.000	0.0278	0.0228	0.0328
Belarus	−14.55	206	0.000	−0.0975	−0.1107	−0.0843
Switzerland	14.77	206	0.000	0.0291	0.0252	0.0330
Luxembourg	18.05	206	0.000	0.0327	0.0291	0.0363

Table 5 indicates that all the differences are statistically significant, in that there is a significant positive difference between the normalised values estimated by the respondents and the actual normalised values for the four smaller countries (Luxembourg, Switzerland, Czech Republic, and Hungary), while for the largest country (Belarus), the significant difference is negative. For the case of Hungary, although the difference is significant, it has the smallest mean difference which is because this country is in a close neighbourhood of a turning point (see the meaning of d in Proposition 3).

Figure 6 shows the actual normalised values and the mean estimated normalised values (and 95% confidence interval) based on the Swing data obtained from the sample.

Interestingly, in this case, we see the same trend as the one we identified in SMART. That is to say, the mean estimated normalised values of the smaller alternatives are higher than their actual normalised values, while the mean estimated normalised values of the larger alternatives are lower than their actual normalised values. The results of t-test for Swing normalised values versus the actual normalised values for the experiment are shown in Table 6.

Table 6 shows that all the differences (except for Hungary) are statistically significant: there is a significant positive difference between the normalised values estimated by the

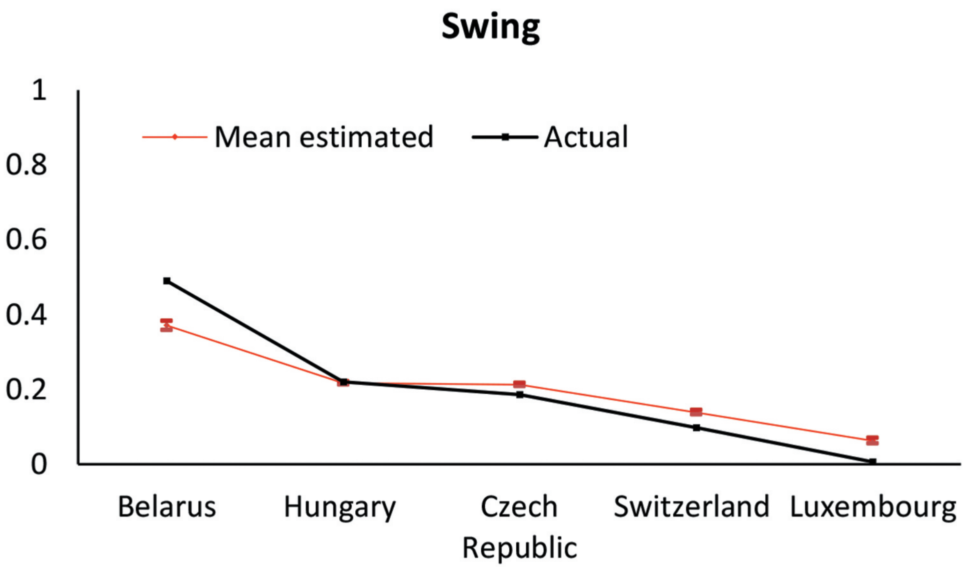

Figure 6. Actual vs mean estimated normalised values (and 95% confidence interval) of the countries (Swing).

Table 6. *T*-test results for the Swing experiment.

	t	df	Sig. (2-tailed)	Mean difference	95% Confidence interval of the difference	
					Lower	Upper
Hungary	−1.65	246	0.100	−0.0034	−0.0074	0.0007
Czech	13.33	246	0.000	0.0262	0.0223	0.0301
Belarus	−19.11	246	0.000	−0.1197	−0.1321	−0.1074
Switzerland	14.88	246	0.000	0.0403	0.0350	0.0465
Luxembourg	14.80	246	0.000	0.0566	0.0491	0.0641

respondents and the actual normalised values for the three smaller countries (Luxembourg, Switzerland and the Czech Republic), while the significant difference is positive for the largest country, Belarus. For the case of Hungary, we do not see a significant difference. This is because this country is in a very close neighbourhood of a turning point (see the meaning of *e* in Proposition 4).

Table 7 summarises the main findings of this study.

It is interesting to see that, while the findings of this study are in line with existing literature on anchoring bias as far as the *scores* are concerned, the findings with regard to

Table 7. A summary of the main findings.

	SMART	Swing
Scores	The estimated scores are *lower* than the actual scores.	The estimated scores are *higher* than the actual scores.
Normalised values	The estimated normalised values for the smaller alternatives are *higher* than their corresponding actual ones, while the estimated normalised values of the larger alternatives are *lower* than their actual corresponding ones.	The estimated normalised values for the smaller alternatives are *higher* than their corresponding actual ones, while the estimated normalised values of the larger alternatives are *lower* than their actual corresponding ones.

the normalised values open a new window into the decision-making field. It is interesting that, as can be seen from Table 7, while the direction of the anchoring bias for the two methods is opposite, the existence of the anchoring bias in the scores shows the same bias direction in the estimated normalised values with both methods. This is a major contribution to the field with significant scientific and practical implications.

6.3. Other explanations

If we look at the last columns of Tables 3 and 4, we can see that while the bias index for SMART method calculated for all the countries is very close to each other (ranging from 69% to 75%), the bias index for the Swing method shows some very large numbers for the smallest alternative (1825% for Luxembourg). We think that while, for SMART, associating the bias index values could relate to anchoring bias, the values for Swing could have partially another source too, when it comes to the smallest alternative (Luxembourg), which we would like to discuss here.

As we see in these two methods, for SMART one starts with 10 (for the least important attribute or the least important alternative w.r.t. an attribute) and has no upper limit. In our dataset, we could find, for instance, numbers as big as 1600 (which is used for Belarus). However, in Swing, one starts with 100 (which is used for Belarus), and the person has limitation in using a small number for the smallest country. In our dataset we could find, for instance, number 1 as the smallest which has been used for Luxembourg. So, while in the first case, here, the ratio is 160 (Belarus is 160 times larger than Luxembourg according to one respondent), in the second case, that ratio is 100. So, we think that the Bias indexes for the smallest alternative do not merely represent the anchoring bias as if it is subject to some limitation in the use of numbers in the Swing method. This scale limitation has also been previously discussed by Pöyhönen and Hämäläinen (2001).

7. Suggestions for debiasing

Earlier studies in behavioural psychology have shown that mitigation strategies are not effective enough to compensate for anchoring bias (Adame, 2016). It has been shown to be resistant to logic and decomposition (Montibeller & Von Winterfeldt, 2015). For instance, Chapman and Johnson (2002) have shown that we cannot reduce the impact of anchoring simply by alerting decision-makers to its potential existence. There is ample evidence to suggest that that in itself does not reduce the anchoring effect (Wilson et al., 1996). These studies show that the decision-makers should interpret the results more carefully. Although anchoring bias appears to be unavoidable, fortunately, it is possible to develop methods in such a way as to minimise the effects of anchoring bias. Montibeller and Von Winterfeldt (2015), conducting a comprehensive literature review, suggested three general strategies:

- Avoiding anchors;
- Providing multiple and counter-anchors;
- Using different experts who use different anchors.

Avoiding anchor might not be applicable here as a debiasing strategy for the methods discussed in this study, as it is actually part of the methods. However, we could think of other multi-attribute decision-making methods, such as DR (direct rating), where anchor is avoided. The second strategy, providing multiple and counter-anchors, which is also called consider-the-opposite strategy (Mussweiler & Strack, 2000; Mussweiler et al., 2000) or Rotating the reference point (Lahtinen et al., 2020), that has been proven to be effective in some other studies (Adame, 2016; Joslyn et al., 2011; Mussweiler, 2002; Mussweiler et al., 2000) could be considered here. Consider-the-opposite strategy means taking into account (multiple) opposite anchors. If we look at some other multi-attribute decision-making methods, we can see that, in some way or another, they actually include this consider-the-opposite strategy in their mechanism (although perhaps not for the purpose of minimising the anchoring effect). One attempt has been already done in literature (Mustajoki et al., 2005) by aggregating both SMART and Swing. The authors, however, did not check its results to those individual methods with respect to cognitive biases, which could be an interesting subject for future research. Based on our findings the two methods show significant bias, but as the direction of bias in the two is very similar, aggregating the two methods might not mitigate this type of bias. As another instance, if we look at the BWM (Rezaei, 2015), we see that this method is based on two evaluation vectors. The first vector is pairwise comparisons of the *best*, the most important, attribute (or alternative w.r.t. an attribute) to all the other attributes (or alternatives w.r.t. an attribute), while the second vector is the pairwise comparison of all the other attributes (or alternatives w.r.t. an attribute) to the *worst*, the least important, attribute (or alternative w.r.t. an attribute). The two vectors are used in one optimisation problem to find the weights. If a decision-maker is biased towards the best, because the opposite alternative is the worst, the other vector generates an opposite bias. Including these two opposite biases in a single model is expected to cancel out the anchoring impact in some way, resulting in less biased conclusions. We think the same applies to AHP (Saaty, 1977) and to other methods that are based on pairwise comparison matrix. In AHP, a decision-maker compares all the attributes (or alternatives w.r.t. an attribute) to each other, which means that, each time an attribute (or an alternatives w.r.t. an attribute) is compared to all the other attributes (or alternatives), the decision-maker is biased towards that attribute (or that alternative). However, because in the end, the weights are based on all the pairwise comparisons, it is to be expected that the anchoring bias is minimised, as all the opposite alternatives have been considered in the method. Finally, we think that the use of multiple experts, and maybe multiple methods could lead to less biased weights. All we suggest here as debiasing strategies are based on some features of the aforementioned methods and experimental studies need to be done to test the effectiveness of these proposals.

8. Conclusion and future research

This is one of a few experimental studies to examine the existence of a cognitive bias, i.e. anchoring bias, in multi-attribute decision-making methods. More specifically, we focused on two of the most basic and fundamental multi-attribute decision-making methods, SMART and Swing, for the purpose of this study, and we conducted an experiment for the purpose of analysis. Although the case used in this study is more complex than those usually encountered in psychological experiments, since, in our study, respondents were asked to estimate a property of more than one alternative compared to the anchor, the results involving the estimated scores are completely in line with earlier findings. That is to say, as far as the SMART method is concerned, where respondents start with a low anchor, their estimation of the other alternatives is lower than the corresponding actual values, while with regard to the Swing method, where respondents start with a higher anchor, their estimation of the other alternatives is higher than their corresponding actual values. What is very interesting, however, is that, despite the different direction of the bias in the estimations of the alternatives in the two methods, the impact of the anchoring bias on the final normalised attribute values obtained from the two methods is the same, in that *the normalised attribute values of the smaller alternatives are estimated to be higher than their actual normalised attribute values, while the normalised attribute values of the larger alternatives are lower than their actual normalised attribute values*.

We think that as the main functions of the two methods studied in this paper are used in several other methods, especially methods that are based on the opinion of experts or decision-makers, future research could examine the existence of this bias and other cognitive biases in other multi-attribute decision-making methods. In our study, in order to control the range bias, we did not consider the more advanced versions of SMART and Swing which is an interesting direction for future research. Future research should also develop mechanisms designed to minimise the impact of anchoring bias on the conclusions of the methods. Our findings show that the two methods we considered here also provide different results which might be explained by the biases inherent in these methods. So, as a future suggestion, we think it would be interesting to investigate explaining the difference between the different methods by looking at their vulnerability to different cognitive biases. Finally, although the case we used in our experiment (size of the countries) has several advantages, it might be criticised as comparing surface areas might involve other types of biases (see, for instance, Krider et al. (2001)). As the theoretical part of our study is generic and supports different types of experiments, we recommend conducting more analyses (using one-dimensional cases for experimental purposes) as well as real-world case studies.

Disclosure statement

No potential conflict of interest was reported by the author.

ORCID

Jafar Rezaei http://orcid.org/0000-0002-7407-9255

References

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. https://doi.org/10.1016/j.lmot.2015.11.002

Arceneaux, K. (2012). Cognitive biases and the strength of political arguments. *American Journal of Political Science*, *56*(2), 271–285. https://doi.org/10.1111/j.1540-5907.2011.00573.x

Arkes, H.R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin*, *110*(3), 486. https://doi.org/10.1037/0033-2909.110.3.486

Belton, V. (1986). A comparison of the analytic hierarchy process and a simple multi-attribute value function. *European Journal of Operational Research*, *26*(1), 7–21. https://doi.org/10.1016/0377-2217(86)90155-4

Borcherding, K., Eppel, T., & Von Winterfeldt, D. (1991). Comparison of weighting judgments in multiattribute utility measurement. *Management Science*, *37*(12), 1603–1619. https://doi.org/10.1287/mnsc.37.12.1603

Bottomley, P.A., & Doyle, J.R. (2001). A comparison of three weight elicitation methods: Good, better, and best. *Omega*, *29*(6), 553–560. https://doi.org/10.1016/S0305-0483(01)00044-5

Buchanan, J.T., & Corner, J. (1997). The effects of anchoring in interactive MCDM solution methods. *Computers & Operations Research*, *24*(10), 907–918. https://doi.org/10.1016/S0305-0548(97)00014-2

Campbell, S.D., & Sharpe, S.A. (2009). Anchoring bias in consensus forecasts and its effect on market prices. *Journal of Financial and Quantitative Analysis*, *44*(2), 369–390. https://www.jstor.org/stable/40505928

Cen, L., Hilary, G., & Wei, K.J. (2013). The role of anchoring bias in the equity market: Evidence from analysts' earnings forecasts and stock returns. *Journal of Financial and Quantitative Analysis*, *48*(1), 47–76. https://doi.org/10.1017/S0022109012000609

Chapman, G.B., & Johnson, E.J. (2002). Incorporating the irrelevant: Anchors in judgments of belief and value. In D. Griffin, D. Kahneman, & T. Gilovich (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 120–138). Cambridge University Press.

Churchman, C.W., & Ackoff, R.L. (1954). An approximate measure of value. *Journal of the Operations Research Society of America*, *2*(2), 172–187. https://doi.org/10.1287/opre.2.2.172

Das, T., & Teng, B.S. (1999). Cognitive biases and strategic decision processes: An integrative perspective. *Journal of Management Studies*, *36*(6), 757–778. https://doi.org/10.1111/1467-6486.00157

Deniz, N. (2020). Cognitive biases in MCDM methods: An embedded filter proposal through sustainable supplier selection problem. *Journal of Enterprise Information Management*, *ahead-of-print*(ahead–of–print). https://doi.org/10.1108/JEIM-09-2019-0285

Edwards, W. (1977). How to use multiattribute utility measurement for social decisionmaking. *IEEE Transactions on Systems, Man, and Cybernetics*, *7*(5), 326–340. https://doi.org/10.1109/TSMC.1977.4309720

Edwards, W., & Barron, F.H. (1994). SMARTS and SMARTER: Improved simple methods for multiattribute utility measurement. *Organizational Behavior and Human Decision Processes*, *60*(3), 306–325. https://doi.org/10.1006/obhd.1994.1087

Ferretti, V. (2020). *Insights from an initial exploration of cognitive biases in spatial decisions. Behavioral operational research*. Springer.

Fisher, K.L., & Statman, M. (2000). Cognitive biases in market forecasts. *The Journal of Portfolio Management*, *27*(1), 72–81. https://doi.org/10.3905/jpm.2000.319785

French, S. (1989). *Readings in decision analysis*. CRC Press.

Furnham, A., & Boo, H.C. (2011). A literature review of the anchoring effect. *The Journal of Socio-economics*, *40*(1), 35–42. https://doi.org/10.1016/j.socec.2010.10.008

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. https://doi.org/10.1016/S0167-9236(00)00074-9

Ghaderi, M., & Kadziński, M. (2020). Incorporating uncovered structural patterns in value functions construction. *Omega*, 102203. https://doi.org/10.1016/j.omega.2020.102203

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases". *European Review of Social Psychology*, *2*(1), 83–115. https://doi.org/10.1080/14792779143000033

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*(1), 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge university press.

Griffiths, T.L., Lieder, F., & Goodman, N.D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, *7*(2), 217–229. https://doi.org/10.1111/tops.12142

Grossman, S.J., & Hart, O.D. (1992). *An analysis of the principal-agent problem*. Foundations of Insurance Economics, Springer.

Hämäläinen, R.P., & Alaja, S. (2008). The threat of weighting biases in environmental decision analysis. *Ecological Economics*, *68*(1–2), 556–569. https://doi.org/10.1016/j.ecolecon.2008.05.025

Hess, D., & Orbe, S. (2013). Irrationality or efficiency of macroeconomic survey forecasts? Implications from the anchoring bias test. *Review of Finance*, *17*(6), 2097–2131. https://doi.org/10.1093/rof/rfs037

Hilbert, M. (2012). Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making. *Psychological Bulletin*, *138*(2), 211. https://doi.org/10.1037/a0025940

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. https://doi.org/10.1287/deca.1070.0100

Joslyn, S., Savelli, S., & Nadav-Greenberg, L. (2011). Reducing probabilistic weather forecasts to the worst-case scenario: Anchoring effects. *Journal of Experimental Psychology. Applied*, *17*(4), 342. https://doi.org/10.1037/a0025901

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and Biases: The Psychology of Intuitive Judgment*, *49*, 81. https://doi.org/10.1017/CBO9780511808098.004

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.

Kakeneno, J.R., & Brugha, C.M. (2017). Usability of Nomology-based methodologies in supporting problem structuring across cultures: The case of participatory decision-making in Tanzania rural communities. *Central European Journal of Operations Research*, *25*(2), 393–415. https://doi.org/10.1007/s10100-016-0460-9

Keeney, R.L., & Raiffa, H. (1976). *Decisions with multiple objectives*. Wiley.

Kirkwood, C.W. (1997). *Strategic decision making*. Duxbury Press.

Köksalan, M., Wallenius, J., & Zionts, S. (2013). An early history of multiple criteria decision making. *Journal of Multi-Criteria Decision Analysis*, *20*(1–2), 87–94. https://doi.org/10.1002/mcda.1481

Krider, R.E., Raghubir, P., & Krishna, A. (2001). Pizzas: π or square? Psychophysical biases in area comparisons. *Marketing Science*, *20*(4), 405–425. https://doi.org/10.1287/mksc.20.4.405.9756

Lahtinen, T.J., & Hämäläinen, R.P. (2016). Path dependence and biases in the even swaps decision analysis method. *European Journal of Operational Research*, *249*(3), 890–898. https://doi.org/10.1016/j.ejor.2015.09.056

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. https://doi.org/10.1016/j.ejor.2019.09.004

LeBoeuf, R.A., & Shafir, E. (2006). The long and short of it: Physical anchoring effects. *Journal of Behavioral Decision Making*, *19*(4), 393–406. https://doi.org/10.1002/bdm.535

Lieder, F., Griffiths, T.L., Huys, Q.J., & Goodman, N.D. (2018). Empirical evidence for resource-rational anchoring and adjustment. *Psychonomic Bulletin & Review*, *25*(2), 775–784. https://doi.org/10.3758/s13423-017-1288-6

Montibeller, G., & Von Winterfeldt, D. (2015). Cognitive and motivational biases in decision and risk analysis. *Risk Analysis*, *35*(7), 1230–1251. https://doi.org/10.1111/risa.12360

Mumma, G.H., & Wilson, S.B. (1995). Procedural debiasing of primacy/anchoring effects in clinical-like judgments. *Journal of Clinical Psychology*, *51*(6), 841–853. https://doi.org/10.1002/1097-4679(199511)51:6<841::AID-JCLP2270510617>3.0.CO;2-K

Mussweiler, T. (2002). The malleability of anchoring effects. *Experimental Psychology*, *49*(1), 67. https://doi.org/10.1027//1618-3169.49.1.67

Mussweiler, T., & Strack, F. (2000). The use of category and exemplar knowledge in the solution of anchoring tasks. *Journal of Personality and Social Psychology*, *78*(6), 1038. https://doi.org/10.1037/0022-3514.78.6.1038

Mussweiler, T., Strack, F., & Pfeiffer, T. (2000). Overcoming the inevitable anchoring effect: Considering the opposite compensates for selective accessibility. *Personality & Social Psychology Bulletin*, *26*(9), 1142–1150. https://doi.org/10.1177/01461672002611010

Mustajoki, J., Hämäläinen, R.P., & Salo, A. (2005). Decision support by interval SMART/SWING—incorporating imprecision in the SMART and SWING methods. *Decision Sciences*, *36*(2), 317–339. https://doi.org/10.1111/j.1540-5414.2005.00075.x

Ni, F., Arnott, D., & Gao, S. (2019). The anchoring effect in business intelligence supported decision-making. *Journal of Decision Systems*, *28*(2), 67–81. https://doi.org/10.1080/12460125.2019.1620573

O'Brien, D.B., & Brugha, C.M. (2010). Adapting and refining in multi-criteria decision-making. *Journal of the Operational Research Society*, *61*(5), 756–767. https://doi.org/10.1057/jors.2009.82

Olson, D.L., Moshkovich, H.M., Schellenberger, R., & Mechitov, A.I. (1995). Consistency and accuracy in decision aids: Experiments with four multiattribute systems. *Decision Sciences*, *26*(6), 723–747. https://doi.org/10.1111/j.1540-5915.1995.tb01573.x

Phillips-Wren, G., Jefferson, T., & McKniff, S. (2019). Cognitive bias and decision aid use under stressful conditions. *Journal of Decision Systems*, *28*(2), 162–184. https://doi.org/10.1080/12460125.2019.1643695

Pöyhönen, M., & Hämäläinen, R.P. (2001). On the convergence of multiattribute weighting methods. *European Journal of Operational Research*, *129*(3), 569–585. https://doi.org/10.1016/S0377-2217(99)00467-1

Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega*, *53*, 49–57. https://doi.org/10.1016/j.omega.2014.11.009

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. (2018). Piecewise linear value functions for multi-criteria decision-making. *Expert Systems with Applications*, *98*, 43–56. https://doi.org/10.1016/j.eswa.2018.01.004

Rezaei, J. (2020). A concentration ratio for nonlinear best-worst method. *International Journal of Information Technology & Decision Making*, *19*(3), 891–907. https://doi.org/10.1142/S0219622020500170

Riabacke, M., Danielson, M., & Ekenberg, L. (2012). State-of-the-art prescriptive criteria weight elicitation. *Advances in Decision Sciences*, 2012. https://doi.org/10.1155/2012/276584

Rouhana, N.N., O'Dwyer, A., & Morrison Vaso, S.K. (1997). Cognitive biases and political party affiliation in intergroup conflict. *Journal of Applied Social Psychology*, *27*(1), 37–57. https://doi.org/10.1111/j.1559-1816.1997.tb00622.x

Saaty, T.L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, *15*(3), 234–281. https://doi.org/10.1016/0022-2496(77)90033-5

Saaty, T.L. (1990). How to make a decision: The analytic hierarchy process. *European Journal of Operational Research*, *48*(1), 9–26. https://doi.org/10.1016/0377-2217(90)90057-I

Schoemaker, P.J., & Waid, C.C. (1982). An experimental comparison of different approaches to determining weights in additive utility models. *Management Science*, *28*(2), 182–196. https://doi.org/10.1287/mnsc.28.2.182

Schwenk, C.R. (1984). Cognitive simplification processes in strategic decision-making. *Strategic Management Journal*, *5*(2), 111–128. https://doi.org/10.1002/smj.4250050203

Stewart, T.J. (1992). A critical survey on the status of multiple criteria decision-making theory and practice. *Omega*, *20*(5–6), 569–586. https://doi.org/10.1016/0305-0483(92)90003-P

Tetlock, P.E. (2000). Cognitive biases and organizational correctives: Do both disease and cure depend on the politics of the beholder? *Administrative Science Quarterly*, *45*(2), 293–326. https://doi.org/10.2307/2667073

Thomas, D.E., Eden, L., Hitt, M.A., & Miller, S.R. (2007). Experience of emerging market firms: The role of cognitive bias in developed market entry and survival. *Management International Review*, *47*(6), 845–867. https://doi.org/10.1007/s11575-007-0055-8

Triantaphyllou, E. (2000). *Multi-criteria decision making methods. Multi-criteria decision making methods: A comparative study*. Springer.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, *185* (4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124
Tversky, A., & Kahneman, D. (1975). Judgment under uncertainty: Heuristics and biases. In D. Wendt and C. Vlek, (Eds), *Utility, probability, and human decision making: selected proceedings of an interdisciplinary research conference*, Rome, 3–6 September, 1973. Dordrecht, Springer Netherlands: 141–162.
von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research*. Cambridge University Press.
Weber, M., & Borcherding, K. (1993). Behavioral influences on weight judgments in multiattribute decision making. *European Journal of Operational Research*, *67*(1), 1–12. https://doi.org/10.1016/0377-2217(93)90318-H
West, R.F., Toplak, M.E., & Stanovich, K.E. (2008). Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions. *Journal of Educational Psychology*, *100*(4), 930. https://doi.org/10.1037/a0012842
Wilson, L. (2012). Anchoring bias in the TARP warrant negotiations. *Journal of Economics and Business*, *64*(1), 63–76. https://doi.org/10.1016/j.jeconbus.2011.03.002
Wilson, T.D., Houston, C.E., Etling, K.M., & Brekke, N. (1996). A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology. General*, *125*(4), 387. https://doi.org/10.1037/0096-3445.125.4.387
Yang, Z., Zhang, Z.-K., & Zhou, T. (2013). Anchoring bias in online voting. *EPL (Europhysics Letters)*, *100* (6), 68002. https://doi.org/10.1209/0295-5075/100/68002