The decoy effect in relative performance evaluation and the debiasing role of DEA

Heinz Ahn[*]

TU Braunschweig, Institute of Management Control & Business Accounting

Fallersleber-Tor-Wall 23, 38100 Braunschweig, Germany

Tel.: +49 531 391 3610

E-mail: hw.ahn@tu-bs.de

Nadia Vazquez Novoa

TU Braunschweig, Institute of Management Control & Business Accounting

[*] Corresponding author

The decoy effect in relative performance evaluation and the debiasing role of DEA

1 Introduction

Despite being described as a discipline aiming at facilitating decision making, operations research (OR) has largely omitted behavioral issues. Most recently, this deficit has been addressed by Hämäläinen et al. (2013). They comprehensively discuss the high potential of behavioral OR for advancing the practical benefit of normative OR methods and propose nine topics for a respective research agenda. The present paper refers particularly to the topic of behavioral/cognitive aspects, since it stresses performance evaluation biases related to the so-called decoy effect.

The decoy effect is a special kind of context effect, which refers to the influence that contextual variables have on decision making. Overwhelming evidence exists demonstrating that context is likely to influence performance evaluation (see, e.g., Damisch et al., 2006; Page & Page, 2010). For example, evaluators' preferences are biased not only by the past performance of the decision making unit (DMU), but also by the performance of other DMUs under analysis. With regard to the latter aspect, the decoy effect implies that the inclusion of a dominated alternative – the decoy – can influence the choice between the non-dominated alternatives. Concretely, the probability of preferring the target alternative, which is the non-dominated option that is most similar to the decoy, may increase. The existence of this effect has repeatedly been confirmed by research on consumer behavior, showing that customers tend to prefer the target alternative (Ariely, 2009; Huber et al., 1982; Wedell & Pettibone, 1996).

The research settings used for studying consumer behavior strongly resemble relative performance evaluation cases where alternatives are compared on a utility function level. Furthermore, these research settings are constructed in a way that they differentiate between non-dominated (i.e., efficient) and dominated (i.e., inefficient) units like in the context of data envelopment analysis (DEA). Against this background, two questions arise: (i) does the decoy effect also occur in cases where the relative performance of alternatives is evaluated, and (ii) to what extent can the application of DEA – namely the incorporation of respective efficiency scores and the mention of existing slacks – act as a debiasing tool?

To shed light on these questions, we conducted a vignette-based experiment with bachelor students taking management control and business accounting courses. For analyzing whether

relative performance evaluation may also be affected by a decoy effect, the well-known experimental setting from the area of consumer behavior was adapted to the performance context. The results show that adding a dominated DMU to the set of DMUs augments the attractiveness of the target DMU. Therefore, a decoy effect can be observed when DMUs are compared against each other.

In a second step, the role of reporting DEA results as a debiasing procedure for the identified decoy effect was considered. Participants in the corresponding treatments were provided with additional information about DEA efficiency scores and slacks. The results indicate that the scores discriminating between efficient and inefficient DMUs can significantly reduce the decoy effect in a relative performance evaluation context. The mention of the existence of slacks for distinguishing between strong and weak efficient DMUs also contributes to avoiding the decoy effect, but it is associated to unexpected effects, such as the increment of the proportion of participants choosing the dominated alternative as the best performing one.

An additional experiment was conducted to evaluate whether this unexpected effect was due to a negative formulation of the meaning of the slacks. The results show that a positive explanation of the slacks does not affect the percentage of participants considering the dominated DMU as the best performing one. This suggests that further research for understanding the interpretation and use of the slacks information should be conducted.

The rest of the paper is structured as follows: Section 2 depicts the theoretical background of our research, outlining the decoy effect in choice tasks as well as the basic aspects of relative performance evaluation with DEA. The hypotheses are derived in Section 3, and the method to investigate these hypotheses is described in Section 4. Section 5 presents the results of the study, addressing the decoy effect on the one side and the potential of DEA as a debiasing mechanism on the other side. Section 6 concludes the paper.

2 Motivation

2.1 Managerial issues related to performance evaluation and the role of DEA

Performance has been defined as a social construct that acquires its meaning within a decision-making context. It is a complex concept that refers not only to the actions performed, but also to their results in comparison to benchmarks. Without such a comparison, it is almost impossible to qualify performance as good or bad. The inherent complexity of the performance concept is augmented by the fact that each decision maker may interpret performance data

differently (Lebas & Euske, 2007). To this respect, research is limited and little is known about the information which managers use and the way they process it (Neely, 2007).

Traditional performance measurement systems (PMS) were based on financial measures and fixed benchmarks, but in the last decades, the important role of non-financial measures and flexible benchmarks has been recognized (Otley, 2007). Appropriate PMS are now seen as systems incorporating relatively few measures and non-financial leading indicators, applying the measures consistently, and compensating people according to those measures for which they are responsible (Meyer, 2007). Nevertheless, Meyer (2007) points out that currently used PMS differ from this ideal concept and are affected by a number of drawbacks. One of them is the difficulty of compensating people for their efforts, since aggregating multiple measures into an overall one may cause different problems. On the one side, a formula-based aggregation may produce game-playing behaviors; on the other side, a subjective evaluation may generate unfairness perception and impede the understanding of the results (Burney et al., 2009; Ittner & Larcker, 2001; Ittner et al., 2003; Meyer, 2007).

One performance measurement instrument especially dealing with this aggregation issue is DEA, a non-parametric approach to determine overall (relative) efficiency scores for DMUs based on financial and non-financial performance criteria. The idea of DEA is to calculate the relative efficiency of n DMUs on the basis of m weighted inputs and s weighted outputs, which serve as performance criteria. For each DMU_o of the set of DMUs, a mathematical programming model maximizes its efficiency score θ_o, ranging from 0% to 100%. Thereby, the weights for the inputs and outputs are endogenously determined for each DMU_o, with the consequence that weights are not constant over the DMUs in order to present each of them in its best possible light. Furthermore, the method allows the analyst to identify benchmarks for inefficient DMUs as well as slacks, defined as non-radial (additional) input excesses and/or output shortfalls (for an overview of DEA, see, e.g., Cook & Seiford, 2009; Cooper et al., 2007; Thanassoulis, 2001).

When applying DEA, the decision maker needs to build a performance evaluation framework, i.e., to determine the set of DMUs to be evaluated, the performance criteria to be used and the specific DEA model to be applied. Once these parameters have been specified and the mathematical calculation has been carried out, the DEA results should be interpreted and checked for plausibility. The framework building phase and the interpretation phase involve different cognitive tasks, such as choice and estimation.

Original DEA models were developed for measuring efficiency and took no major account of the preferences of decision makers in this process. This was criticized, since a DMU, e.g., can be classified as strong efficient even if it is just attaining a superior performance on only one criterion (Ahn et al., 2012; Wallenius et al., 2008). Different concepts have been proposed for including value judgments in DEA, but none of them has applied a behavioral approach. In fact, research in DEA has only concentrated on pitfalls from a normative and rational perspective (see, e.g., Dyson et al. 2001). The only exception so far is the paper by Zha et al. (2013), which proposes a halo heuristic to solve the problem of missing data.

2.2 Behavioral research on subjective performance evaluation: a brief review

Previous research dealing with subjective performance evaluation from a behavioral perspective has mainly concentrated on the use of unique and common performance measures in the context of balanced scorecards (BSC) as well as on the influence of strategy maps and the presentation format on performance judgments. A review of the three main accounting journals publishing articles on the topic of behavioral performance evaluation (Journal of Accounting Research, The Accounting Review and Accounting, Organizations and Society) permits us to identify ten highly influential articles in the period 1998–2012 dealing with management control topics. Seven out of the ten articles concentrate on performance evaluation, with a strong focus on BSC. The remaining three articles deal with the effect of control systems on trust, collaboration, norms and organizational change (Chenhall & Euske, 2007; Coletti et al., 2005; Tayler & Bloomfield, 2011). The contributions treating performance evaluation issues will be briefly described in the following.

The experiment conducted by Lipe & Salterio (2000) is the first to document cognitive limitations associated to the use of BSC as an instrument for performance evaluation. Their results suggest that (i) unit managers' decisions will be more influenced by common measures than by unique measures designed for a particular business unit and (ii) evaluators will underweight non-financial and leading measures, thus reducing the benefits offered by the implementation of a BSC. In the same research direction, Banker et al. (2004) evaluate the role played by information on strategic linkages. They confirm the common measures bias and also find that strategically linked measures dominate common measures only in the case that managers understand the unit's strategy. Libby et al. (2004) provide other debiasing alternatives for the common measures bias. Building on prior research in performance

evaluation and consumer behavior, the authors find that requiring managers to justify their performance evaluation (accountability) and augmenting the reliability of the measures increase the use of unique measures.

Contrary to the psychology-oriented research originated from the paper of Lipe & Salterio (2000), Banker et al. (2000) deal with performance evaluation from an economic perspective. They conclude that customer-related non-financial measures are associated with long-term financial performance. They also show that (financial and non-financial) performance improves after including non-financial measures in the incentive contract. Ittner et al. (2003) integrate the psychological and the economic perspectives in their study of subjectivity weighting of multiple performance measures. The authors conclude that their results are more consistent with psychology-based theories predicting an "outcome effect" – i.e., more weight will be placed on outcome measures (external measures, e.g. financial and customer measures) than on driver measures (internal measures, e.g., process improvement, employee satisfaction). Further, they find that people tend to assign more weight to "known" outcome measures than to "new" ones, confirming that other factors besides informativeness influence performance evaluation.

The remaining two articles deal with PMS characteristics and their effects on performance, thus extending their conclusions to other systems beyond the BSC. Hall (2008) shows that a comprehensive PMS has an indirect effect on managerial performance mediated by role clarity and psychological empowerment. Burney et al. (2009) analyze other mediating variables for the effect of PMS on performance. They find that the perception of two PMS characteristics – the PMS is technically valid and it reflects a strategic causal model – positively influence both distributive and procedural justice perceptions. The authors also show that procedural justice perception alters organizational citizenship behaviors, which in turns affects performance. As conclusion, they suggest that clearly communicating the characteristics of the PMS could avoid problems associated with formulaic incentive plans.

A thorough understanding of the possible biases and the corresponding debiasing mechanisms are of special importance for the design and implementation of adequate PMS (Cheng & Humphreys, 2012). The DEA approach, as a comprehensive measure of performance, becomes a more and more integrated part of such systems (Epstein & Henderson, 1989). Therefore, its cognitive effects should be considered. While there may be undesirable consequences of incorporating DEA results into performance reports, the paper concentrates on a desirable consequence, namely its role as a debiasing procedure for the decoy effect.

2.3 The decoy effect in choice tasks

Normative choice models are based on the assumption that the preference between two options is independent of the presence or absence of a third option, i.e., they satisfy the axiom of independence of irrelevant alternatives (Eisenführ et al., 2010; Tversky & Simonson, 1993). Therefore, the probability of choosing one of two efficient alternatives a and b should not increase when an inefficient – i.e. dominated – alternative c is added to the original data set (condition of regularity). Formally, given $\mathbf{A} = \{a, b\}$ and $\mathbf{B} = \{a, b, c\}$, the following inequalities for the probabilities should hold: P (a; $\mathbf{A}$) $\geq$ P (a; $\mathbf{B}$) and P (b; $\mathbf{A}$) $\geq$ P (b; $\mathbf{B}$) (Ariely & Wallsten, 1995; Huber et al., 1982). Despite appearing to be an intuitive principle, experimental research in judgment and decision making has shown that this axiom can be violated by adding a new option (decoy) to the set of alternatives. Such a decoy alters the preferences of the decision maker, increasing the attractiveness of the dominating or near-dominating alternative (Pettibone & Wedell, 2000; Wedell & Pettibone, 1996). This phenomenon, known as the decoy effect, is a case of local context-dependent preferences (Tversky & Simonson, 1993).

The first type of decoy to be studied was the asymmetrically dominated decoy, which refers to an alternative that is dominated by another one (target) but not by the rest. Huber et al. (1982) identify three different kinds of asymmetrically dominated decoys according to their position relative to the target: (i) range decoys (R), which increase the range of the weak dimension of the target, (ii) frequency decoys (F), which increase the number of alternatives on the strong dimension of the target, and (iii) range-frequency decoys (RF), a combination of both. All of them have been shown to affect choice, but the RF decoys evidence the weakest effect. A reason for this may be found in the difficulty of recognizing the dominance. In Fig. 1, there are two alternatives (a and b) with the same level of utility but different performance on the two dimensions. For a as the target option and b its competitor, a R_a decoy will increase the covered range on the weak dimension of a (for this dimension 1, the original range $d_{1b} - d_{1a}$ will be extended to $d_{1b} - d_{1Ra}$), thus decreasing the importance of the difference on this dimension. The F_a decoy, on the contrary, will augment the weight of the dimension on which a is superior (dimension 2) by reducing the variance on that dimension, thus increasing the perceived distance between the two extremes (a and b). Finally, the combined RF_a decoy will increase the range on dimension 1 and reduce the variance in dimension 2 (Huber et al., 1982; Wedell & Pettibone, 1996).
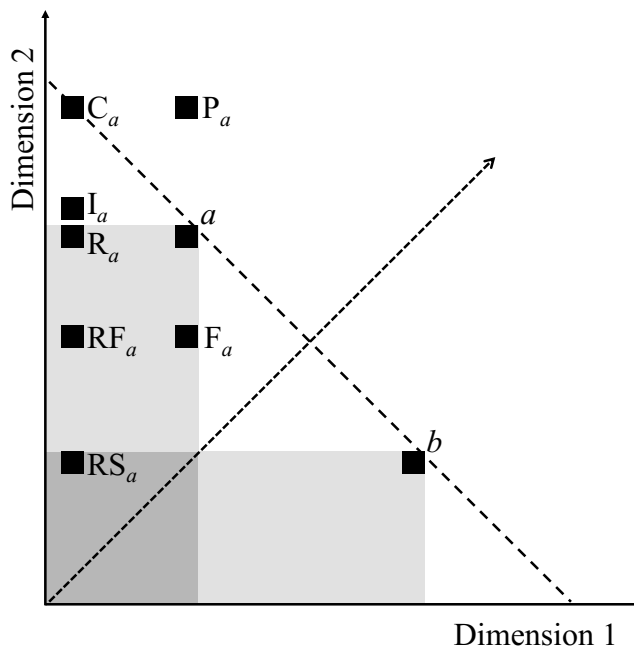
Fig. 1. Location of the different types of decoys (Pettibone & Wedell, 2000; Wedell & Pettibone, 1996)

Besides these decoys, not asymmetrically dominated decoys exist: the range symmetrical decoys (RS) and the non-dominated decoys. RS decoys extend the weak dimension of the target item but are symmetrically dominated by the target and the competitor. Non-dominated decoys include (i) the inferior decoy (I), which is similar to R but with a higher value than the target on one dimension, (ii) the compromise decoy (C), which is located close to the indifference curve defined by the dominating alternatives, and (iii) the phantom decoy (P), a decoy that dominates all alternatives but cannot be selected. In Fig. 1, the I_a decoy also extends the range of the dimension on which a is weak, but I_a is not strictly dominated by a as it has a higher value than the target on dimension 2. The C_a decoy has the same level of utility as options a and b, but it makes the target alternative a appear as a compromise between the two extreme items C_a and b. The P_a decoy differs from all others as it dominates the target option without being eligible itself; it corresponds to the "sold out" product or the candidate who has already signed a contract with another company (Pettibone & Wedell, 2000).

Ariely (2009) suggests that the decoy effect occurs because the decision maker cannot easily evaluate the trade-offs necessary to choose between the two superior alternatives. The introduction of a dominated alternative permits the decision maker to reduce the cognitive load by creating a simple relative comparison between the decoy and the target. For further information about the different models proposed for explaining the decoy effect (weight change, value shift, and value added), see, e.g., Wedell & Pettibone (1996).

Research on the decoy effect for the case of choice tasks has mainly concentrated on the analysis of three alternatives (two options and a decoy) defined in only two dimensions. Huber et al. (1982) suggest that increasing the complexity of the task – by augmenting the number of alternatives or attributes – could limit the decoy effect as the dominated alternative might not easily been recognized. Simonson (1989) argues that increasing complexity and realism would reduce transparency, hence producing more decisional errors. An extension of the results to a multidimensional space is offered by Ariely & Wallsten (1995), who conclude that decision makers assign a higher weight to those dimensions that help discriminating among alternatives.

The decoy effect appears to be robust and has been identified in several choice domains, including consumer behavior and personnel selection. Different manipulations have been proposed to weaken this bias, but not all of them have satisfactorily reduced or eliminated it (Slaughter et al., 2011). One of the first studies dealing with this topic was the experiment conducted by Simonson (1989). He examines the influence of making participants accountable for their decisions on the decoy effect and finds that instead of reducing or eliminating the effect, accountability aggravates it. Recently, Connolly et al. (2013) test the impact of making participants think about the regret they might experience if they make a wrong decision (regret priming) on the decoy effect and conclude that it is a useful debiasing tool. Hamilton et al. (2007) find that incorporating fully dominated options to a set can cause the opposite reaction than the decoy effect does. The authors explain this effect as a consequence of grouping those alternatives with a repeated common attribute value and focusing on the only alternative with a unique value.

Previous research studying the decoy effect in performance judgments is very limited, with the consequence that issues such as the influence of decoys on performance ratings remain unexplained (Reb et al., 2014). In the area of individual performance assessment, Highhouse (1996) analyzes the effects of decoy alternatives in a simulated employee selection context and finds that rank-ordering of candidates depends on the presence of irrelevant alternatives. In a more recent study, Slaughter et al. (2006) confirm this conclusion in the context of a group-based decision and corroborate the strengthening effect that accountability has on the decoy effect. These common attempts to investigate the decoy effect strongly resemble the relative performance evaluation settings where DEA can be applied. Consequently, it seems worth investigating the decoy effect in this context.

3   Hypotheses and predictions

The present study investigates (i) the existence of the decoy effect in a hypothetical setting of relative performance evaluation and (ii) the potential of the DEA approach to reduce this effect. Thanks to the comment of a reviewer, it should be clarified here that both scenarios address utility determination, while scenario (ii) also deals with (DEA) efficiency determination. Utility determination and efficiency determination can be understood as two different evaluation levels which are hierarchically connected via the relevant inputs and outputs and the assumptions made in a specific evaluation case (Dyckhoff, 2006). In order to find attractive units, DEA forms a production frontier which represents the efficient input-output combinations based on a certain set of assumptions; with further preferences of the evaluator concerning the inputs and outputs, a utility function allows the decision maker to select among these efficient input-output combinations. From this hierarchical point of view, the evaluation of alternatives depends on the structure of the efficient frontier and further preferences which shape the utility function.

In our study, scenario (i) focuses on the utility evaluation level in order to investigate the decoy effect. By providing additional information about DEA results in scenario (ii), also the efficiency evaluation level is addressed. In both scenarios, the decision makers choose upon their individual set of (implicit) preferences needed to evaluate utility. Only in the second scenario, they additionally have access to (explicit) information about efficiency measures.

With respect to the first part of our study, we expect that decision makers evaluating the performance of different subsidiaries on a relative basis will behave in an analogous manner to consumers in a choice context. This means that they will favor the target subsidiary when a decoy is included in the set of options. Replicating previous experimental research, we present a case describing the performance of two superior subsidiaries and an asymmetrically dominated one by means of four attributes. Our first hypothesis can be summarized as follows:

> H1a: The preference for the target subsidiary as the best performing DMU will increase when an RF decoy is included in the choice set.

> H1b: The preference for the target subsidiary as the best performing DMU will increase when an R decoy is included in the choice set.

According to Ariely (2009), the presence of a decoy simplifies the comparison among alternatives by focusing attention on the decoy and the target options. In the second part of our study, we strive for avoiding such a misleading simplification on the utility evaluation level by making the evaluators aware of information on the efficiency evaluation level. In particular, we provide DEA efficiency scores and non-zero slacks which characterize the superior alternatives as (DEA-)efficient DMUs, while dominated decoys are characterized as weak efficient or inefficient DMUs. Thereby, we strive to prove that the undesired decoy effect can be reduced, since evaluators are expected to eliminate non-efficient DMUs from further analysis, thus limiting the set of relevant DMUs just to the superior alternatives. This can be summarized as:

H2a: The presence of DEA efficiency scores differentiating between efficient and inefficient DMUs will reduce the decoy effect.

H2b: The mention of non-zero slacks differentiating between strong efficient and weak efficient DMUs will reduce the decoy effect.

Hypothesis H2a will be evaluated for the case in which RF decoys are added to the DMU set, hypothesis H2b for the case containing R decoys. The decision to consider these decoys and not others originates in previous results suggesting that RF decoys have the weakest effect, while R decoys have the strongest one (Huber et al., 1982; Wedell & Pettibone, 1996).

4 Method

4.1 Participants and design

Bachelor students (N = 482 taking introductory management control and business accounting courses at a German university received a performance report during a lecture. Eight participants did not complete the task and their questionnaires were therefore eliminated from further analysis. The basic design variables were: (i) decoy type (R or RF), (ii) decoy target (*a* or *b*), and (iii) DEA analysis (with or without DEA results). Students were randomly assigned to the different treatments and control conditions, the latter were included for testing that the set of non-dominated alternatives were considered to have approximately the same utility level. Two different control conditions were considered, one containing DEA efficiency scores and one without any kind of performance aggregation measure. A total of ten different groups were analyzed (see Table 1).

Table   **1**
Sample size of the different treatments

Conditions	Without DEA results		With DEA results	
	DMUs	N	DMUs	N
			Efficiency scores	
Control condition	a, b	57	a, b	53
RF-decoy	a, b, RF_a	53	a, b, RF_a	56
	a, b, RF_b	54	a, b, RF_b	51
			Efficiency scores and slacks	
R-decoy	a, b, R_a	38	a, b, R_a	34
	a, b, R_b	37	a, b, R_b	41

4.2 Case materials

A short vignette presenting the hypothetical case of a delivery chain with two/three subsidiaries was presented. The participants were required to assume the role of the central management and to decide which of these subsidiaries deserves a bonus based on the available performance criteria. All participants received a table containing the data of two criteria to be minimized and two criteria to be maximized for each of the subsidiaries. The DEA conditions also included a brief description of the method and the corresponding DEA efficiency results. For the treatments with RF decoys, only the DEA efficiency scores were included in the report as they are sufficient for discriminating between efficient and inefficient DMUs. For the treatments with R decoys, the existence of slacks was also mentioned since the efficiency scores could not be used as a discriminating cue. The vignette text presented in the appendix shows that the experiment was kept as simple as possible, following the premise that experimental research should reduce complexity in order to avoid that side effects or additional concerns influence the results. Especially, it was not discussed which DEA model should be chosen. Instead, the basic CRS model served as a first example to support our hypothesis that DEA efficiency scores can act as a debiasing procedure for the decoy effect. The VRS model could have also been applied in this case as it leads to the same classification of the DMUs into efficient, weak efficient and inefficient ones. Nevertheless, further research investigating the effects of using the scores of other DEA models should be conducted.

The attribute values of the superior DMUs were selected so that in the control case both alternatives are preferred by about half of the participants (Connolly et al., 2013). For evaluating whether the alternatives a and b were perceived as indifferent by the decision makers, a pretest was conducted (N = 78 high school students), with the result that 54% of the participants preferred alternative a over b. R decoys for a (R_a) and for b (R_b) were constructed based on the procedure presented by Wedell & Pettibone (1996), but adapting it for a task with four attributes instead of two. The decoys were obtained by separating the total space in two subspaces (one to be minimized and one to be maximized) and applying the construction rule in each of the two subspaces. For the construction of the RF decoys, a small modification was required since otherwise RF decoys for a (RF_a) and b (RF_b) would have had different DEA efficiency scores and a considerable improvement potential (RF_a = (52; 45; 155; 100), efficiency score = 67%; RF_b = (34; 65; 85; 168), efficiency score = 71%). Therefore, RF_a and RF_b were created by changing the original values in less than 10% in order to obtain almost efficient decoys with the same efficiency scores. Extremely inefficient DMUs may be immediately disregarded, thus transforming the decision task in an analogous case to the control condition. Table 2 presents the values corresponding to each of the performance criteria for the two superior DMUs and each of the decoys.

Table   **2**
Data for the experiments

| Performance criteria	a	b	R_a	R_b	RF_a	RF_b
To be minimized						
Number of call-center employees (monthly average)	43	25	52	25	45	27
Number of complains (daily average)	34	56	34	65	37	59
To be maximized						
Processed purchase orders of clothing articles (hourly average)	190	120	190	85	186	119
Processed purchase orders of household articles (hourly average)	134	202	100	202	127	198
Efficiency score	100%	100%	100%	100%	96%	96%
Slacks $\neq$ 0	no	no	yes	yes	n.a.	n.a.

4.3 Procedure

The participants were informed that the aim of the study was to capture the different responses to a performance assessment case from a descriptive point of view, implying that there was no 'right' or 'wrong' answer. Subjects received a paper-based version of the vignette corresponding to the treatment they had been randomly assigned to. They were asked to complete the questionnaire in approximately five minutes and under final exam conditions, with the difference that anonymity was guaranteed. No questions requiring them to justify their decision were included since augmenting accountability has been shown to intensify the decoy effect (Connolly et al., 2013; Simonson, 1989).

Participants in the control condition received a binary choice set represented as a 4 × 2 matrix. The choice set corresponding to the control condition including DEA efficiency scores was represented as a 5 × 2 matrix. Participants in the other treatments were provided with a trinary choice set. In the treatments without efficiency scores, the performance data was arranged in a 4 × 3 matrix. When RF decoys were added, the superior DMUs were strong efficient and the decoy was inefficient. In this case, only the efficiency scores were added to the original matrix (5 × 3 matrix). In the treatments including R decoys, all three DMUs were efficient, but the decoy was just weak efficient. To be able to discriminate between them, information regarding slacks was added (6 × 3 matrix).

Previous research has shown that a sequential presentation of alternatives affects the way they are evaluated (Damisch et al., 2006; Page & Page, 2010). Consequently, the data corresponding to each DMU was randomized to avoid undesired biases related to the presentation order.

5 Results

Consistent with previous research, the results show that the inclusion of a decoy in a performance evaluation context increments the proportion of participants preferring the target alternative. As shown in Fig. 2, approximately 60% of the participants in the two control conditions chose DMU *a* as the best performing DMU. This proportion increased when an inferior alternative similar to *a* was included in the report and decreased when the inferior option was similar to *b*. The decoy effect shrank when DEA results (efficiency score in the case of RF decoys; efficiency score and information about the slacks in the case of R decoys) were presented. The next sections report the statistical results for the two research hypothesis of these experiments.
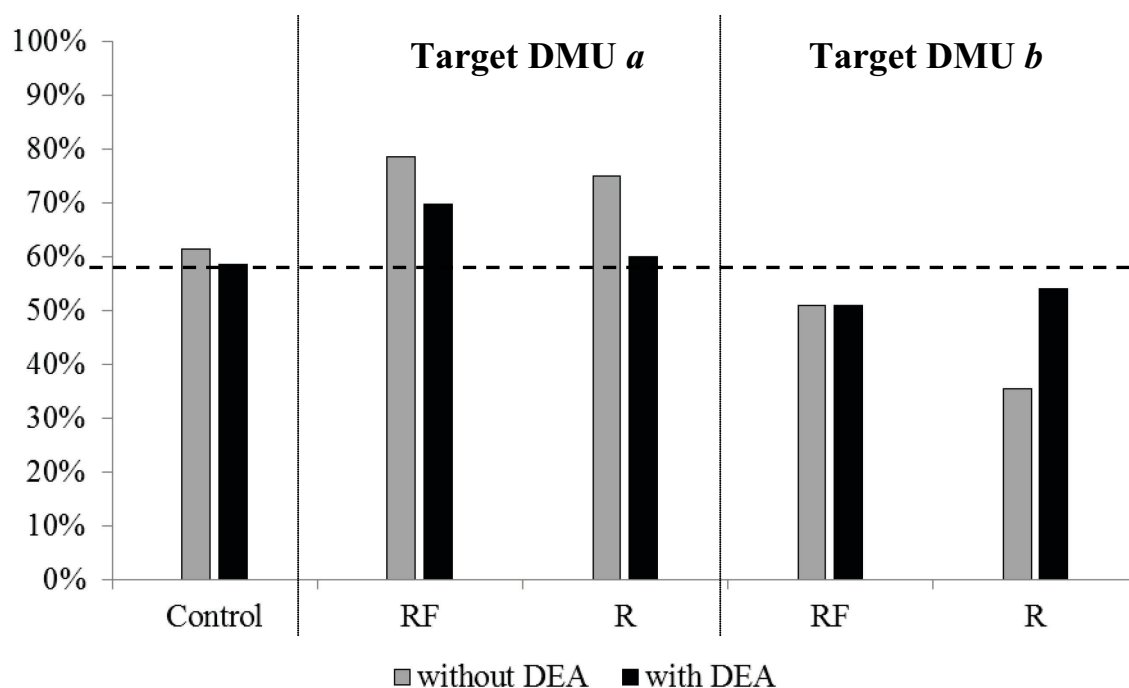
Fig. 2. Proportion of participants choosing DMU *a* as the best performing DMU in the different treatments

5.1 The decoy effect in a performance evaluation context

The first purpose of our study was to prove the existence of a decoy effect in an assumed performance evaluation context. The hypothesis was that both, an RF decoy as well as an R decoy would influence the preferences of the performance evaluator. Table 3 summarizes the results for the treatments including both kinds of decoys and the control case when no DEA results were provided. In the control case, 61% of the participants chose subsidiary *a* over subsidiary *b*. This proportion is similar to the one suggested in the study of Wedell & Pettibone (1996). A test for binomial distribution can be used for determining whether the percentage of subjects choosing one of the options is significantly higher than 50%. In the control case, there is no evidence to conclude that any of the two alternatives is more preferred than the other one (*p*-value = 0.111).

As in other relevant previous studies (e.g., Huber et al., 1982; Pettibone & Wedell, 2000), a minor number of participants chose the inferior alternative. Even if this has been shown to be a common behavior (see, e.g., Russo et al., 2006, for some recent research on this topic), the following tests will adjust to the existing literature and therefore be based only on the cases in which the decoy was not chosen. In Table 3, the sample size excluding the choices for the decoy and the corresponding proportions for alternatives *a* and *b* are presented in parentheses.

Table **3**
Preferences in the treatments without DEA results

DMU set	N	*a*	*b*	Decoy
a, *b*	57	0.614	0.386	
a, *b*, RF$_a$	53	0.755	0.208	0.038
	(51)	(0.784)	(0.216)	
a, *b*, RF$_b$	54	0.481	0.463	0.056
	(51)	(0.510)	(0.490)	
a, *b*, R$_a$	38	0.711	0.237	0.053
	(36)	(0.750)	(0.250)	
a, *b*, R$_b$	37	0.324	0.595	0.081
	(34)	(0.353)	(0.647)	

Simonson (1989) notes that a decoy effect exists if the proportion of participants preferring DMU *a* in both treatments (one with target DMU *a* and the other with target DMU *b*) significantly differs. In our case, a corresponding effect can be observed for both kind of decoys by means of a χ^2 test for difference of proportions as well as a Fisher exact test for independence of categorical data (R$_a$ vs. R$_b$: H$_0$: P(a)$_{Ra}$ = P(a)$_{Rb}$; χ^2 = 11.173, *p*-value = 0.001, Fisher exact test *p*-value = 0.002; RF$_a$ vs. RF$_b$: H$_0$: P(a)$_{RFa}$ = P(a)$_{RFb}$; χ^2 = 8.414, *p*-value = 0.004, Fisher exact test *p*-value = 0.007). In a similar manner, if a decoy effect exists, the percentage of subjects choosing the target alternative is significantly larger than 50%. Connolly et al. (2013) note that this proportion could vary if one of the options is more popular than the other; nevertheless, they explicitly desist from modifying this parameter as it would not affect their conclusions but would add further complexity to the hypothesis testing. Accordingly, and since we could not find any evidence that the proportion of subjects choosing *a* in the control condition is significantly different from 0.50, we also applied an exact binomial test with P = 0.50. The results of this test confirm the existence of a decoy effect in the hypothetical performance evaluation context used in our experiment (R decoy: H$_0$: P(a)$_R$ = 0.50; *p*-value = 0.001; RF decoy: H$_0$: P(a)$_{RF}$ = 0.50; *p*-value = 0.007).

The addition of the R and RF decoys significantly altered the preferences of the evaluators. These experimental results suggest that performance evaluation is likely to be affected by the decoy effect.

5.2 Using DEA as a debiasing mechanism

The decoy effect in performance evaluation is problematic not only regarding the fairness of the appraisal, but also with respect to strategic consequences. When two DMUs are attaining a superior performance, but they are not easily comparable, the inclusion of a third DMU with a lower performance appears to be critical for the decision of which DMU should be presented as a benchmark for all others. The second purpose of our study is to evaluate whether this issue may be avoided by incorporating an overall performance measure that highlights the difference among strong efficient, weak efficient and inefficient DMUs. This debiasing mechanism is expected to focus decision maker's attention on the two strong efficient DMUs, thus avoiding the effect of the decoy alternative on the choice. The experimental results presented in Table 4 support this idea.

Table   **4**
Preferences in the treatments with DEA results

DMU set	N	a	b	Decoy
a, b	53	0.585	0.415	
a, b, RF_a	56 (53)	0.661 (0.698)	0.286 (0.302)	0.054
a, b, RF_b	51 (49)	0.490 (0.510)	0.471 (0.490)	0.039
a, b, R_a	34 (30)	0.529 (0.600)	0.353 (0.400)	0.118
a, b, R_b	41 (37)	0.488 (0.541)	0.415 (0.459)	0.098

The control condition including DEA efficiency scores can be assumed to follow a binomial distribution (H_0: P(a) = 0.50; exact binomial test p-value = 0.272). Further, a comparison between the results for both control conditions shows that reporting DEA efficiency scores and briefly explaining the DEA methodology does not affect preferences (H_0: P(a) $_{Control No DEA}$ = P(a)$_{Control DEA}$; χ^2 = 0.097, p-value = 0.755, Fisher exact test p-value = 0.846). Therefore, increasing of the amount of information provided to the respondents does not appear to be the reason for the preference changes described in the following paragraphs.

DEA efficiency scores differentiating between efficient and inefficient DMUs. This part of the analysis concentrates on the treatments including RF decoys, which correspond to inefficient DMUs. Reporting DEA efficiency scores that divide DMUs into efficient and

inefficient ones seems to act as a debiasing procedure for reducing the decoy effect. According to a χ^2 test for difference of proportions and a Fisher exact test, the proportion of participants choosing DMU a in both treatments did not significantly differ when efficiency scores were reported (H$_0$: P(a)$_{\text{RF}a\ \text{DEA}}$ = P(a)$_{\text{RF}b\ \text{DEA}}$, χ^2 = 3.772, p-value = 0.052, Fisher exact test p-value = 0.068; exact binomial test p-value = 0.059). This reveals the potential of efficiency scores to avoid the decoy effect. However, this result should be taken with caution and calls for re-examining experiments, as the p-values are close to 0.05.

DEA efficiency scores and information about slacks, differentiating between efficient and weak efficient DMUs. Additional information about the existence of non-zero slacks to differentiate among strong efficient and weak efficient DMUs also contributed to reducing the decoy effect (H$_0$: P(a)$_{\text{R}a\ \text{DEA}}$ = P(a)$_{\text{R}b\ \text{DEA}}$, χ^2 = 0.239, p-value = 0.625, Fisher exact test p-value = 0.804; exact binomial test p-value = 0.807).

Nevertheless, a closer consideration of the data reported in Table 4 posits one question regarding the relative high percentage of participants preferring the decoys as the best performing DMUs. Previous research has documented that some participants erroneously choose the decoy as the best alternative (Huber et al., 1982; Pettibone & Wedell, 2000). Furthermore, Russo et al. (2006) suggest that initially providing information supporting an inferior alternative leads people to consider this option as the leading one and to prefer it in a choice task.

This violation to rational decision making may be due to a misunderstanding of the meaning of the non-zero slacks that might be originated in the formulation used to explain the role of the slacks in a DEA-based performance evaluation. This definition was stated in negative terms indicating that a DMU with slacks is not completely efficient. The vignette included the following formulation: 'Additionally, DEA offers the possibility of identifying slacks. If a subsidiary has at least one slack $\neq$ 0, then this subsidiary is not completely efficient (even if it has a DEA efficiency score = 100%)'. Labeling the attributes in a positive or a negative manner is one way of modifying the presentation format, thus affecting the judgment of the attributes (Levin et al., 1998). Negative formulations are related to the word 'no', which acquires its significance associated to punishment, producing a negative affective value (Alia-Klein et al., 2007). Furthermore, negative sentences usually describe deviations from expectations (Kaup et al., 2006) and cognitive processing of negations is more complex, requires more time, and leads to more decision errors (Hasson & Glucksberg, 2006; Schul, 2011). Therefore, a DMU

with a 'no' in the slacks dimension could have been erroneously perceived as a low performer, while the inferior alternative could have been assumed to be the leading one.

5.3 Supplemental analysis

Based on the framing literature and previous research on negation, an additional experiment was conducted to evaluate whether the negative explanation of the slacks was the reason for the unexpected results obtained in the previous section. The role of the slacks was formulated in a positive manner not only in the verbal description but also in the table containing the performance criteria. The vignette included the following formulation: 'Additionally, DEA offers the possibility of identifying slacks. Only the subsidiaries that have all slacks = 0 are completely efficient, as no additional improvement has been identified.'

The results of this manipulation (see Table 5; N = 57 bachelor students) confirm that the decoy effect is eliminated by reporting the DEA efficiency scores and mentioning the presence of slacks (H_0: $P(a)_{Ra} = P(a)_{Rb}$; $\chi^2 = 0.827$, p-value = 0.363, Fisher exact test p-value = 0.542; exact binomial test p-value = 0.392). Nevertheless, it does not help to dissuade participants of choosing the inferior alternative.

Table 5

Preference in the treatments with a positive formulation concerning the slacks

DMU set	N	*a*	*b*	Decoy
a, *b*, R$_a$	29 (26)	0.655 (0.731)	0.241 (0.269)	0.103
a, *b*, R$_b$	28 (23)	0.500 (0.609)	0.321 (0.391)	0.179

A possible explanation for this behavior is based on the experimental results of Ha et al. (2009). They find that categorical attributes may mitigate the decoy effect by two different mechanisms: (i) editing and (ii) hierarchical elimination of options. Editing strategies permit the decision maker to simplify choice problems by means of dismissal of dominated options and cancellation of shared or redundant attributes. Hierarchical elimination is based on a previous categorization of alternatives based on salient features, allowing the elimination of alternatives belonging to a category without considering the other attributes. This process is more likely to be applied in cases that the decision maker has a strong preference for one category over the other.

We expected that the mention of existing slacks would make participants categorize the three alternatives into two groups: strong efficient and weak efficient DMUs. Since strong efficient should be preferred over weak efficient DMUs, participants should have dismissed the dominated alternative (weak efficient DMU), concentrating their attention on the two superior alternatives. This seems to be true for most of the participants, but not for all of them. Other participants may have conducted first a within-group comparison for the strong efficient DMUs and afterwards a comparison against the weak efficient DMU (Ha et al., 2009). If the comparison between a and b indicates that a is the best, participants may have eliminated DMU b for further analysis. In this case, R_b is not dominated anymore, therefore appearing as a rational solution for the choice between a and R_b. On the contrary, if the comparison between a and b finalizes with the elimination of a, the decision maker would need to choose between b and R_b. As R_b is dominated by b, participants may have chosen b as the superior alternative, consistent with the decoy effect.

6   Discussion

The present study offers an analysis of the decoy effect in a performance evaluation context, focusing on the debiasing characteristics of DEA results. The traditional experimental setting commonly used in consumer behavior was slightly modified with the aim of adapting it to a relative performance evaluation case. To the best of our knowledge, it constitutes the first attempt to link the DEA approach to the decoy effect. The results of the experiments can be summarized as follows:

- The first scenario (i) focused on the utility evaluation level. As has been shown, the utility comparison of the alternatives a and b was biased by the decoy effect.

- In the second scenario (ii), also the efficiency evaluation level was addressed by providing additional information about DEA results. As a consequence, the decision makers change their evaluation on the utility level towards a significant reduction of the decoy effect.

In detail, the first experiment considered the R and RF decoy, which significantly influenced the decision made by the evaluators. When the decoys were added to the DMU set, the proportion of participants considering the target DMU as the best performing DMU significantly increased. This can raise managerial issues not only regarding the fairness of the appraisal, but also about its strategic consequences, especially due to its influence on the

selection of benchmarks (Page & Page 2010). According to Simonson (1989), demands for justification may strengthen the decoy effect, thus not serving as a debiasing tool. Therefore, this effect is expected to be even stronger in real performance evaluations than in the experimental conditions, since the central management will usually provide feedback to the DMU managers to justify its judgments.

Our second main result indicates that reporting DEA efficiency scores that differentiate among efficient and inefficient DMUs seems to act as a debiasing procedure for the decoy effect. Adding an RF decoy accompanied by the corresponding scores contributed to reducing the proportion of participants declaring the target DMU as the best performing subsidiary. The results regarding the adequacy of mentioning the existence of non-zero slacks as a debiasing procedure were contradictory. Despite no evidence for a decoy effect is found, an unanticipated effect is observed since the proportion of participants choosing the dominated alternative as the superior one. This suggests that the meaning of non-zero slacks may have been misunderstood, probably as a consequence of the formulation used to explain the role of the slacks. To this respect, research on negation has shown that cognitive processing of negations is more complex, slower and causes more decision errors (Hasson & Glucksberg, 2006; Schul, 2011). An additional experiment was conducted to investigate whether the negative formulation of the slacks could have affected the interpretation of the performance criteria. The results indicate that a positive explanation of the slacks contribute to the elimination of the decoy effect, but not necessarily to the appropriate understanding of their meaning. As noted by one of the reviewers, these findings may have implications for the practical usage and the teaching of DEA. Similarly with the decoy effect, when DEA is illustrated by means of a simple 2-dimensional drawing, the localization of the inefficient observations in the illustration is likely to influence the viewer into favoring certain efficient DMUs. Therefore, further research tending to improve the correct utilization of slacks and other DEA results should be conducted.

Our findings need to be interpreted in light of various limitations. A main criticism of our design may be that the results were obtained from a sample of bachelor students taking management control and business accounting introductory courses. This recourse to students as surrogates for managers has been supported by many authors (e.g., Holm & Rikhardsson, 2008; Moore, 2005; Mortensen et al., 2012) and criticized by many others (e.g., Peterson, 2001). Fuchs & Sarstedt (2010) suggest that, despite being theoretically discussed, the use of students in empirical marketing and management research has been widely accepted.

However, it is advisable to be cautious with generalizing our results, as the participants did not have previous experience with DEA and their knowledge of this approach was limited to the brief explanation included in the vignette. Nevertheless, our experience with implementing DEA in companies indicates that managers using DEA results for performance evaluations will also not have a much more comprehensive understanding of the approach. In this context, it seems interesting to investigate to what extent our results change if no explanation about DEA is provided.

Further research needs to be conducted towards achieving an exhaustive understanding of behavioral performance evaluation and the possible debiasing role of DEA. For example, the addition of other kinds of decoys – such as phantom decoys, which correspond to unobserved DMUs in the DEA literature – and the occurrence of the compromise and attribute-balance effect should be analyzed. To this respect, the unit of measurement of the attributes is worth being considered. Furthermore, it is interesting whether a perceptual focus effect occurs when augmenting the number of inefficient or weak efficient DMUs with the same efficiency score. Finally, research dedicated to analyze visualization effects on the use and interpretation of DEA can provide genuine benefits to the field (Mortenson et al., 2015).

References

Ahn, H., Neumann, L., & Vazquez Novoa, N. (2012). Measuring the relative balance of DMUs. *European Journal of Operational Research*, 221, 417–423. doi: 10.1016/j.ejor.2012.03.030

Alia-Klein, N., Goldstein, R.Z., Tomasi, D., Zhang, L., Fagin-Jones, S., Telang, F., Wang, G.-J., Fowler, J.S., & Volkow, N.D. (2007). What is in a word? No versus yes differentially engage the lateral orbitofrontal cortex. Emotion, 7, 649–659. doi: 10.1037/1528-3542.7.3.649

Almer, E.D., Hopper, J.R., Kaplan, S.E. (2003). A research tool to increase attention to experimental materials: manipulating presentation format. *Journal of Business and Psychology*, 17, 405–418. doi: 10.1023/A:1022868425897

Ariely, D. (2009). *Predictably irrational: the hidden forces that shape our decisions*. New York, NY: Harper Collins.

Ariely, D., & Wallsten, T.S. (1995). Seeking subjective dominance in multidimensional space: an explanation of the asymmetric dominance effect. *Organizational Behavior and Human Decision Processes*, 63, 223–232. doi: 10.1006/obhd.1995.1075

Banker, R.D., Potter, G., & Srinivasan, D. (2000). An empirical investigation of an incentive plan that includes nonfinancial performance measures. *The Accounting Review*, 75, 65–92. doi: 10.2308/accr.2000.75.1.65

Banker, R.D., Chang, H., & Pizzini, M.J. (2004). The balanced scorecard: judgmental effects of performance measures linked to strategy. *The Accounting Review*, 79, 1–23. doi: 10.2308/accr.2004.79.1.1

Burney, L.L., Henle, C.A., & Widener, S.K. (2009). A path model examining the relations among strategic performance measurement system characteristics, organizational justice, and extra- and in-role performance. *Accounting, Organizations and Society*, 34, 305–321. doi: 10.1016/j.aos.2008.11.002

Chenhall, R.H., & Euske, K.J. (2007). The role of management control systems in planned organizational change: an analysis of two organizations. *Accounting, Organizations and Society*, 32, 601–637. doi: 10.1016/j.aos.2006.09.007

Cheng, M.M., & Humphreys, K.A. (2012). The differential improvement effects of the strategy map on scorecard perspectives on managers' strategic judgments. *The Accounting Review*, 87, 899–924. doi: 10.2308/accr-10212

Coletti, A.L., Sedatole, K.L., & Towry, K.L. (2005). The effect of control systems on trust and cooperation in collaborative environments. *The Accounting Review*, 80, 477–500. doi: 10.2308/accr.2005.80.2.477

Connolly, T., Reb, J., & Kausel, E.E. (2013). Regret salience and accountability in the decoy effect. *Judgment and Decision Making*, 8, 136–149.

Cook, W.D., & Seiford, L.M. (2009). Data envelopment analysis (DEA): thirty years on. *European Journal of Operational Research*, 192, 1–17. doi: 10.1016/j.ejor.2008.01.032

Cooper, W.W., Seiford, L.M., Tone, K. (2007). *Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software* (2nd ed.). New York, NY: Springer.

Damisch, L., Mussweiler, T., & Plessner, H. (2006). Olympic medals as fruits of comparison? Assimilation and contrast in sequential performance judgments. *Journal of Experimental Psychology: Applied*, 12, 166–178. doi: 10.1037/1076-898X.12.3.166

Dyckhoff, H. (2006). *Produktionstheorie: Grundzüge industrieller Produktionswirtschaft* (Production theory: basics of industrial production management) (2nd ed.). Berlin et al.: Springer.

Dyson, R.G., Allen, R., Camanho, A.S., Podinovski, V.V., Sarrico, C.S., & Shale, E.A. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research*, 132, 245–259. doi: 10.1016/S0377-2217(00)00149-1

Eisenführ, F., Weber, M., & Langer, T. (2010). *Rational decision making*. Heidelberg et al.: Springer.

Epstein, M.K., & Henderson, J.C. (1989). Data envelopment analysis for managerial control and diagnosis. *Decision Sciences*, 20, 90–119. doi: 10.1111/j.1540-5915.1989.tb01399.x

Fuchs, S., & Sarstedt, M. (2010). Is there a tacit acceptance of student samples in marketing and management research? *International Journal of Data Analysis Techniques and Strategies*, 2, 62–72. doi: 10.1504/IJDATS.2010.030011

Ha, Y.-W., Park, S., & Ahn, H.-K. (2009). The influence of categorical attributes on choice context effects. *Journal of Consumer Research*, 36, 463–477. doi: 10.1086/599045

Hall, M. (2008). The effect of comprehensive performance measurement systems on role clarity, psychological empowerment and managerial performance. *Accounting, Organizations and Society*, 33, 141–163. doi: 10.1016/j.aos.2007.02.004

Hämäläinen, R.P., Luoma, J., & Saarinen, E. (2013). On the importance of behavioral operational research: the case of understanding and communicating about dynamic systems. *European Journal of Operational Research*, doi: 10.1016/j.ejor.2013.02.001

Hamilton, R., Hong, J., & Chernev, A. (2007). Perceptual focus effects in choice. *Journal of Consumer Research*, 34, 187–199. doi: 10.1086/519147

Hasson, U., & Glucksberg, S. (2006). Does understanding negation entail affirmation? An examination of negated metaphors. *Journal of Pragmatics*, 38, 1015–1032. doi: 10.1016/j.pragma.2005.12.005

Highhouse, S. (1996). Context-dependent selection: the effects of decoy and phantom job candidates. *Organizational Behavior and Human Decision Processes*, 65, 68–76. doi: 10.1006/obhd.1996.0006

Holm, C., & Rikhardsson, P. (2008). Experienced and novice investors: does environmental information influence investment allocation decisions? *European Accounting Review*, 17, 537–557. doi: 10.1080/09638180802016627

Huber, J., Payne, J.W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, 9, 90–98.

Ittner, C.D., & Larcker, D.F. (2001). Assessing empirical research in managerial accounting: a value-based management perspective. *Journal of Accounting and Economics*, 32, 349–410. doi: 10.1016/S0165-4101(01)00026-X

Ittner, C.D., Larcker, D.F., & Meyer, M.W. (2003). Subjectivity and the weighting of performance measures: evidence from a balanced scorecard. *The Accounting Review*, 78, 725–758. doi: 10.2308/accr.2003.78.3.725

Kaup, B., Lüdtke, J., & Zwaan, R.A. (2006). Processing negated sentences with contradictory predicates: is a door that is not open mentally closed? *Journal of Pragmatics*, 38, 1033–1050. doi: 10.1016/j.pragma.2005.09.012

Lebas, M., & Euske, K. (2007). A conceptual and operational delineation of performance. In A. Neely (Ed.), *Business performance measurement: unifying theories and integrating practice* (2nd ed., pp. 125–139). Cambridge et al.: Cambridge University Press.

Levin, I.P., Schneider, S.L., & Gaeth, G.J. (1998). All frames are not created equal: a typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, 76, 149–188. doi: 10.1006/obhd.1998.2804

Libby, T., Salterio, S.E., & Webb, A. (2004). The balanced scorecard: the effects of assurance and process accountability on managerial judgment. *The Accounting Review*, 79, 1075–1094. doi: 10.2308/accr.2004.79.4.1075

Lipe, M.G., & Salterio, S.E. (2000). The balanced scorecard: judgmental effects of common and unique performance measures. *The Accounting Review*, 75, 283–298. doi: 10.2308/accr.2000.75.3.283

Meyer, M. (2007). Finding performance: the new discipline in management. In A. Neely (Ed.), *Business performance measurement: unifying theories and integrating practice* (2nd ed., pp. 113–124). Cambridge et al.: Cambridge University Press.

Moore, D.A. (2005). Commentary: conflicts of interest in accounting. In D.A. Moore, D.M. Cain, G. Loewenstein, & M.H. Bazerman (Eds.), *Conflicts of interest: challenges and solutions in business, law, medicine, and public policy* (pp. 70–73). Cambridge: Cambridge University Press.

Mortensen, T., Fisher, R., & Wines, G. (2012). Students as surrogates for practicing accountants: further evidence. Accounting Forum, 36, 251–265. doi: 10.1016/j.accfor.2012.06.003

Mortenson, M.J., Doherty, N.F., & Robinson, S. (2015). Operational research from Taylorism to Terabytes: a research agenda for the analytics age. *European Journal of Operational Research*, 241, 583–595. doi: 10.1016/j.ejor.2014.08.029

Neely, A. (2007). Measuring performance: the operations management perspective. In A. Neely (Ed.), *Business performance measurement: unifying theories and integrating practice* (2nd ed., pp. 64–81). Cambridge et al.: Cambridge University Press.

Otley, D. (2007). Accounting performance measurement: a review of its purposes and practices. In A. Neely (Ed.), *Business performance measurement: unifying theories and integrating practice* (2nd ed., pp. 11–35). Cambridge et al.: Cambridge University Press.

Page, L., & Page, K. (2010). Last shall be first: a field study of biases in sequential performance evaluation on the Idol series. *Journal of Economic Behavior & Organization*, 73, 186–198. doi: 10.1016/j.jebo.2009.08.012

Peterson, R.A. (2001). On the use of college students in social science research: insights from a second-order meta-analysis. *Journal of Consumer Research*, 28, 450–461. doi: 10.1086/323732

Pettibone, J.C., & Wedell, D.H. (2000). Examining models of nondominated decoy effects across judgment and choice. *Organizational Behavior and Human Decision Processes*, 81, 300–328. doi: 10.1006/obhd.1999.2880

Reb, J., Greguras, G.J., Luan, S., & Daniels, M.A. (2014). Performance appraisals as heuristic judgments under uncertainty. In S. Highhouse, R.S. Dalal, & E. Salas (Eds.), *Judgment and decision making at work* (pp. 13–36). New York, NY, & East Sussex: Routledge.

Russo, J.E., Carlson, K.A., & Meloy, M.G. (2006). Choosing an inferior alternative. *Psychological Science*, 17, 899–904. doi: 10.1111/j.1467-9280.2006.01800.x

Schul, Y. (2011). Alive or not dead: implications for framing from research on negations. In G. Keren (Ed.), *Perspectives on framing* (pp. 157–176). New York, NY: Psychology Press, Taylor & Francis Group.

Simonson, I. (1989). Choice based on reasons: the case of attraction and compromise effects. *Journal of Consumer Research*, 16, 158–174. Retrieved from: http://www.jstor.org/stable/2489315

Slaughter, J.E., Bagger, J., & Li, A. (2006). Context effects on group-based employee selection decisions. *Organizational Behavior and Human Decision Processes*, 100, 47–59. doi: 10.1016/j.obhdp.2006.01.003.

Slaughter, J.E., Kausel, E.E., & Quiñones, M.A. (2011). The decoy effect as a covert influence tactic. *Journal of Behavioral Decision Making*, 24, 249–266. doi: 10.1002/bdm.687

Tayler, W.B., & Bloomfield, R.J. (2011). Norms, conformity, and controls. *Journal of Accounting Research*, 3, 753–790. doi: 10.1111/j.1475-679X.2011.00398.x

Thanassoulis, E. (2001). *Introduction to the theory and application of data envelopment analysis: a foundation text with integrated software*. New York, NY: Springer.

Tversky, A., & Simonson, I. (1993). Context-dependent preferences. *Management Science*, 39, 1179–1189. doi: http://www.jstor.org/stable/2632953

Wallenius, J., Dyer, J.S., Fishburn, P.C., Steuer, R.E., Zionts, S., & Deb, K. (2008). Multiple criteria decision making, multiattribute utility theory: recent accomplishments and what lies ahead. *Management Science*, 54, 1336–1349. doi: 10.1287/mnsc.1070.0838

Wedell, D.H., & Pettibone, J.C. (1996). Using judgments to understand decoy effects in choice. *Organizational Behavior and Human Decision Processes*, 67, 326–344. doi: 10.1006/obhd.1996.0083

Zha, Y., Song, A., Xu, C., & Yang, H. (2013). Dealing with missing data based on data envelopment analysis and hallo effect. *Applied Mathematical Modelling*, 37, 6135–6145. doi: 10.1016/j.apm.2012.11.015

Appendix

You are the manager of a delivery chain with three subsidiaries. The year is finishing and you have to decide which of these three subsidiaries deserves a bonus. Following information is available:

Criteria to be minimized	Criteria to be maximized
(monthly average) **Number of call-center employees**	(hourly average) **Processed purchase orders of articles of clothing**
(daily average) **Number of complains**	(hourly average) **Processed purchase orders of household articles**

Additionally, you got from the management control department an evaluation that aggregates these four performance criteria into an overall performance score via data envelopment analysis (DEA). This instrument is briefly described below:

Brief explanation of DEA

Decide now (from your own point of view and based on the following data) which subsidiary deserves a bonus.

Performance criteria		Subsidiary A	Subsidiary B	Subsidiary C
To be minimized	Number of call-center employees	43	25	Decoy
	Number of complains	34	56	
To be maximized	Processed purchase orders of articles of clothing	190	120	
	Processed purchase orders of household articles	134	202	
DEA-efficiency score		*100%*	*100%*	
Slacks ≠ 0[1]		*no*	*no*	

Response: Subsidiary _____ [2]

---

1   This information as well as the last paragraph of the brief explanation of DEA regarding slacks were only included in the treatments corresponding to R_a and R_b with a DEA report.

2   The text written in italics was included only in the treatments with DEA scores. The control case included only the data corresponding to DMUs *a* and *b*. The vignette was originally written in German.