

**LIKE STUDENT LIKE MANAGER?
USING STUDENT SUBJECTS IN
MANAGERIAL DEBIASING RESEARCH**

Lorenz Graf-Vlachy

University of Passau, Germany

+49-851-509-2515

lorenz.graf-vlachy@uni-passau.de

Managerial debiasing studies are rare because it is often challenging to obtain manager samples to perform the required experiments. Student subjects could mitigate this difficulty, but there is widespread uncertainty regarding their implications for a study's validity. In this paper, I first trace the debate, and structure the literature, on the use of student subjects in business research in general. Next, I propose a conceptual framework of criteria to identify under which circumstances student subjects can be valid surrogates in managerial debiasing research. Finally, I illustrate the use of the framework by repeating an extant debiasing study conducted with management practitioners with a large sample of business students ($N = 1,423$), showing that the student sample replicates the results from the manager sample to the expected degree. I close by discussing the study's implications, limitations, and opportunities for future research.

Keywords: Debiasing; behavioral strategy; surrogation; student subjects; experiments.

JEL classification: M10, M12, M53.

This article was published in the *Review of Managerial Science*.

The final publication is available at <http://link.springer.com>

<https://doi.org/10.1007/s11846-017-0250-3>

The author would like to acknowledge helpful comments from Editor-in-Chief Wolfgang Kürsten and two anonymous reviewers that shaped and improved the paper. Andreas König, Harald Hungenberg, Albrecht Enders, Jan Nopper, Sebastian Kreft, Andreas Fügener, Jan Krämer, Philip Meissner, as well as the participants of the 15th Annual Conference of the European Academy of Management in Warsaw and the 75th Annual Meeting of the Academy of Management in Vancouver also contributed valuable insights to this manuscript. Furthermore, Björn Baltzer, Christian Landau, and the local chapter chairpersons of Marketing zwischen Theorie und Praxis (MTP) e.V. greatly supported the data collection effort.

1 Introduction

Inspired by Kahneman and Tversky's (1973) seminal work, a great number of biases have been identified in psychological research over the last decades. Examples include anchoring, availability (Tversky and Kahneman 1974), framing, loss aversion (Kahneman and Tversky 1979), and escalation of commitment (Staw 1997).

Most, or possibly even all, of these biases also apply to managers (e.g., Bateman and Zeithaml 1989; Hodgkinson et al. 1999; Tversky and Kahneman 1986) and can have serious adverse consequences. Shefrin and Cervellati (2011), for example, show how the biases of excessive optimism, overconfidence, loss aversion, and the search for confirming evidence contributed to several catastrophic accidents at BP, including the 2010 explosion of the drilling rig Deepwater Horizon in the Gulf of Mexico, which caused extensive and sustained damage to marine and wildlife habitats and ended up costing the firm in excess of US\$ 40 billion.

Researchers urgently call for debiasing research to mitigate such negative effects and improve decision making both in psychology (Lilienfeld et al. 2009) as well as business research (Milkman et al. 2009). And in fact, psychological research has been attempting to answer this call for quite some time (see Arkes 1991; Fischhoff 1982; Larrick 2004 for initial reviews).

Management research has been rather silent on the issue, likely at least partially due to the challenges involved in recruiting suitable manager samples for debiasing studies. Debiasing research is usually best conducted in (survey) experiments so that causal relationships can be unambiguously identified (Schwenk 1982). Unfortunately, it is notoriously difficult and expensive to convince business professionals to participate in such experiments, as they usually have busy schedules and might have reservations about the utility of scientific

research (Liyanarachchi 2007; Liyanarachchi and Milne 2005). Abdellaoui et al. for example, explicitly report that it was “difficult to find professionals willing to participate” (2013: 416) in a study. Consequently, there exist only a small set of published managerial debiasing studies (e.g., Hodgkinson et al. 1999; Kaufmann et al. 2012; Kaustia and Perttula 2012; Meissner and Wulf 2016).

Some studies attempt to circumvent these problems by relying on student subjects (e.g., Schoemaker 1993; Meissner and Wulf 2013). However, most management scholars appear to exhibit a fairly strong tendency to reject such surrogation (Gordon et al. 1986; Walters-York and Curatola 2000). In fact, in many fields of business research, there appears to exist a widespread belief that experiments conducted with student subjects generalize poorly and are thus best avoided (Colquitt 2008; Croson 2010; Kohli 2011; Liyanarachchi 2007). This belief both reflects and likely also causes the somewhat negative attitude of professional gatekeepers such as journal editors and reviewers towards student samples (Trotman 1996). Gordon et al. (1986), for instance, provide several examples of statements made by editors indicating that student samples are seen as unfit to help answer questions about real-world issues and are considered a weakness that needs to be overcompensated for by especially interesting methodological or theoretical features in order for a containing article to get published. A current example is the editorial policy of the *Journal of Management Studies*, which states that “we do not publish empirical investigations based on student samples.”¹ This widespread skepticism is not surprising, given that a variety of researchers have repeatedly and strongly urged scholars to heavily discount results from student samples (e.g., Burnett and Dunne 1986; Gordon et al. 1986; Gordon et al. 1987). Such policies and appeals, explicit or implicit, clearly influence authors, which occasionally even feel the need to

¹ [http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1467-6486/homepage/ForAuthors.html](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1467-6486/homepage/ForAuthors.html). Accessed 1 October 2016

describe the measures they took “to avoid student sampling” (Wobker and Kenning 2013: 182).

Some scholars, however, disagree, leaving the question of student subject surrogation unresolved. They argue that surrogation can indeed be possible and that the value of student samples should not be underestimated (e.g., Bolton et al. 2012; Croson 2010; Greenberg 1987; Remus 1986). Further testimony to the lack of clarity around this issue is what must be one of the longer back-to-back exchanges of arguments about a single methodological topic in the literature (Dobbins et al. 1988a, b; Gordon et al. 1986; Gordon et al. 1987; Greenberg 1987; Slade and Gordon 1988) and the fact that new papers dealing with the issue of student subject surrogation continue to be published (e.g., Bolton et al. 2012; Croson 2010; Fréchette 2015).

In this manuscript, I attempt to shed some light on student subject surrogation for the specific field of managerial debiasing research and thereby make contributions to several literature streams. First, I trace the discussion on student subject surrogation. I find that it has been ongoing for a very long time but has provided little consistent guidance to practicing researchers. Additionally, it is apparent that despite, or possibly due to, its unsatisfactory state, the discussion has somewhat slowed in recent years. I thus contribute to the wider debate on student subject surrogation (Croson 2010) by surveying and structuring the extant literature to cut through the clutter and to provide a new impetus for the scholarly conversation.

Second, I propose and demonstrate the use of a conceptual framework to think about the suitability of student subject surrogates for use in managerial debiasing research. This framework can be useful to researchers when contemplating student subject surrogation, and to professional gatekeepers such as journal editors or reviewers when evaluating student

subject surrogation. It thus constitutes a methodological contribution to the debiasing literature (Larrick 2004; Meissner and Wulf 2016)

Third, I demonstrate the use of the framework by replicating an existing debiasing study on competitive irrationality that was conducted with managers. Using a large sample of student subjects, I find that the student sample replicates the findings not completely, but to the expected degree. This not only lends credibility to the framework, but it also corroborates the results of the reference study, contributing to research on competitive irrationality (Graf et al. 2012) and answering persistent calls for more replication research in management (Bettis et al. 2016; Evanschitzky et al. 2007).

The paper proceeds as follows. First, I trace the larger surrogation debate by surveying the literature on the topic in management research and other disciplines. Moving beyond prior literature, I then proceed to propose a framework to assess the permissibility of student surrogation in the field of managerial debiasing research. I subsequently demonstrate the framework's utility and tentatively empirically corroborate it. To do so, I develop nomological hypotheses on student subject surrogation with regard to five debiasing techniques. I then introduce the employed stimulus material as well as the subject sample. After presenting the empirical results, I close with a discussion that covers this study's implications, addresses its limitations, and highlights opportunities for further research.

2 The Surrogation Debate in Business Research and Beyond

To productively trace and structure the larger debate on student subject surrogation, I first present an overview of extant criticism of student subject surrogation and discuss the core argument that is fielded by researchers opposing surrogation. I then use selected literature to

discuss the key types of categorization schemes that attempt to give guidance under which circumstances the use of student subjects is deemed acceptable.

2.1 Criticism of Student Subject Surrogation

The question of whether using students, i.e., undergraduate subjects (or graduate subjects without extensive work experience), is permissible in research that should generalize to other populations has been debated for a long time in various disciplines and has not been fully settled. Quite to the contrary; it has produced a plethora of empirical evidence which sends “confusing signals to researchers in all functional areas of business” (Hughes and Gibson 1991: 158). Consequently, to this day, “opinions on the issue diverge dramatically” (Peterson 2001: 450).

Psychologists and researchers of judgment and decision making, as well as economists, have been criticized for their tradition of relying on student subjects as a surrogate for the general population (Ball and Cech 1996; Fréchette 2015; LaTour et al. 1990; Sears 1986). McNemar, for example, famously called psychology the “science of the behavior of sophomores” (1946: 333) and others provocatively questioned the generalizability of psychology’s findings by asking: “Are students real people?” (Cunningham et al. 1974: 399). Similar concerns have been voiced in economics (Ball and Cech 1996).

Business scholars have been debating the more specific question of whether students can take the role of managers, executives, accountants, or other business professionals as research subjects. Empirical findings are decidedly mixed, with some researchers concluding that surrogation will work (e.g., Remus 1986), but many arguing that it will produce misleading results (e.g., Moskowitz 1971). Some researchers gravitate towards one of the two positions,

but remain very careful in their wording because their evidence leaves some room for interpretation (e.g., Fréchette 2015).

Business researchers who are skeptical about surrogation argue that the use of student subjects often represents a substantial threat to external validity, whereas researchers taking the opposing view discount this argument. External validity refers to the degree to which the findings from a study can be generalized to a broader set of contexts, in particular to other environments and other populations (Campbell 1957; Campbell and Stanley 1963). Concerns about such generalizability are what leads surrogation skeptics to often reject student samples. Some other researchers, however, qualify the relevance of the skeptics' concerns by cautioning that external validity is not necessarily always what is desired or intended (e.g., Mook 1983). They argue that in cases where no predictions about the real world are to be made from experiment results, but predictions about what *should* happen under artificial conditions or with specific samples are to be tested, even "externally invalid" samples can be used.

In many cases however, generalization is desired, and at least for such cases, surrogation skeptics take a probabilistic approach towards inductive inference and impose the requirement that samples from which one desires to generalize should be similar to the target population (Greenberg 1987). Only in some cases, for example when testing job applicants' attitudes towards a potential employer's policies (Crant and Bateman 1990; Murphy et al. 1990), the use of student subjects is deemed acceptable, as they are highly similar to the actual population of interest (Gordon et al. 1986). This may also hold true for samples of employed students if the research question relates to workplace behavior (e.g., Fox et al. 2001).

In most cases in business research, then, student subjects represent a problem from the skeptics' point of view because scholars have identified a number of differences between

students and non-students (especially managers) which can be used to explain differing experiment results between the two groups. Differences in at least three categories are frequently mentioned in the literature. First, psychological characteristics and values often differ. Miller et al. (1983), for example, attributed differences in experimental results to differences in social values. Other researchers identified differences with regard to attitudes (Alpert 1967; Copeland et al. 1973; Remus 1986), personality traits (Oakes 1972), needs (South 1974; Stahl and Harrell 1982), and preferences (Fréchette 2015). Some scholars also highlight the fact that age has a substantial influence, as individuals' personality traits, preferences, and their susceptibility to biases evolve over time (Quoidbach et al. 2013; Roberts et al. 2006). Second, prior research used education, experience, and task familiarity to explain diverging outcomes (Gordon et al. 1986 and the sources reviewed therein; Schultz 1969) as experiment settings that contain a lot of context are potentially perceived differently by subjects with different backgrounds (Bolton et al. 2012; Cooper et al. 1999; Fréchette 2015; Gordon et al. 1986). Closely related are differences in knowledge about role expectations, for example about what is expected of a professional manager (Gordon et al. 1987). Finally, differing socioeconomic status has been used to explain differences between students and non-students (Latham and Dossett 1978). Taken together, all these differences between students and non-students can give rise to a great dose of skepticism against the use of student subjects for studies which attempt generalizations to other populations.

However, the skeptics' argument has limitations that make it difficult to accept without further qualifications. If one were to take the skeptics' logic to the extreme, this would preclude *any* generalization—a consequence that would likely be unacceptable to most management scholars. Strictly speaking, no experiment subject group whatsoever—be it a sample of students or of managers—will ever allow generalization to another group with complete certainty. As the name suggests, generalizations are always inferences from a

specific set of actual observations to a more general set of potential observations. Making such inferences requires the use of inductive logic which can never be fully justified and is hence problematic (Hume 1999). After all, as Oakes aptly put it: “*any* population one may sample is ‘atypical’” (1972: 962). Hence, it can be argued that even generalizing from one manager sample to another population of managers is impossible. Yet it appears reasonable and commonly accepted to do precisely that.

A theoretically and empirically grounded categorization scheme for deciding under which circumstances student subjects are similar enough to managers could help find a middle ground between naïvely accepting student subjects and skeptically rejecting all results from studies that do not exhaustively survey the entire population of interest. Such a categorization scheme could prevent cases in which researchers have to make surrogation decisions in an unsatisfactory fashion like based only on personal judgment or socially negotiated convention.

2.2 Extant Categorization Schemes

Several authors have suggested categorization schemes that go beyond the simplistic question of whether or not students are similar to the target population. Such categorization schemes provide criteria for when substituting students for other target populations would be permissible and when it would not be, depending on which property of decision makers is to be studied. *Table 1* provides an overview of such studies that focus on properties of the decision makers.

Ashton and Kramer (1980), for instance, took up Hawkins et al.’s (1977) notions of surface-level differences (attitudes) and underlying psychological processes (behavior) as well as Lamb and Stern’s (1980) distinction between state and process research, and proposed to

make a distinction between attitudes and decision making, more specifically the outcome of decisions as well as information processing. They suggested that students and managers might well diverge in attitudes, but might exhibit similar decision making behavior. The idea was later taken up again by Beltramini (1983) who distinguishes habits (behavioral level) and attitudes (attitudinal level).

In fact, with regard to attitudes, most researchers concur that the debate appears to be settled, with the outcome that students frequently have substantially different attitudes than managers (Alpert 1967; Copeland et al. 1973; Remus 1986).

The situation is much less clear with regard to decision outcomes and decision processes (Croson 2010). Some researchers find that students and managers reach similar conclusions when making decisions or estimating values (e.g., Dipboye et al. 1975; Hofstedt 1972; Liyanarachchi and Milne 2005; Remus 1986) and some have found the opposite (e.g., Abdelkhalik 1974; Fleming 1969; Khera and Benson 1970; Moskowitz 1971). Similarly, there is empirical evidence suggesting great similarities in the process of information processing and decision making (see Locke 1986 for an extensive review or Bolton et al. 2012 for a more recent example) and there is evidence which highlights possible differences (e.g., Barr and Hitt 1986; Harris and Sutton 1995; Hofstedt 1972; Hughes and Gibson 1991). In fact, even Ashton and Kramer themselves find evidence they consider “a matter of judgment” (1980: 11).

=====

Insert Table 1 about here

=====

Other researchers take a different approach to categorizing the situations in which students may or may not be used as proxies for managers or other target groups. Instead of focusing on properties of the decision makers, they focus on the purpose of a study. *Table 2* displays such studies and highlights the two key streams of research that follow this approach.

On the one hand, a fairly small group of scholars makes a distinction between phases of research. Miller (1966), for example, focuses on study design and takes a very restrictive stance. He limits the permissible use of students to trial runs of studies. Shuptrine (1975) is similarly careful and proposes that students may be used only for exploratory research, but not for confirmatory research.

The majority of scholars, on the other hand, takes a different approach and distinguishes between research studying psychological mechanisms and research studying real-world behavior. Calder et al. (1981), for example, proposed that students could be valid subjects when a study attempts to test theories, that is, the structure of cognitive processes (theory application research) as opposed to quantifying effect sizes (effects application research). This distinction is echoed in many subsequent works (e.g., Bolton et al. 2012; Dobbins et al. 1988a) and is quite similar to the earlier notions of conceptual-level and descriptive-level research put forward by Farber (1952) as well as Kruglanski's (1975) distinction between universalistic and particularistic research. In a similar vein, Enis et al. (1972) suggested that students can be used in studies which attempt to test internal validity rather than external validity. Others, such as Mook (1983), make a comparable point. Roth (1995) proposes another possible purpose of research, namely to explore anomalies observed in the field. For this purpose, the suitability of student subjects depends on whether the observed anomaly is caused by a universal bias or institutions that shape actors' behavior, or by specific features of the observed individuals (Croson 2010).

=====
Insert Table 2 about here
=====

3 Towards a Conceptual Framework to Assess Student Subject Surrogation in Managerial Debiasing Research

As outlined above, all categorization schemes on behavior research that do not restrict student samples to mere preparatory work permit the productive use of student subjects in specific cases, for example to build internally valid theories about psychological mechanisms. However, since the advice offered by the different schemes is also at least partially contradictory, scholars with an interest in debiasing managers are still faced with substantial uncertainty when contemplating or evaluating the use of student subjects.

In the following, I will—as a theoretical background—first briefly discuss the nature and aims of managerial debiasing research. Then, I will propose a practically useful framework of criteria to think about and assess whether student surrogation is permissible for a given debiasing research project.

3.1 Biases and Debiasing in Management Research

Cognitive biases are observed whenever normative decision rules are systematically violated (Babcock et al. 1997). Such violations may occur when people either do not know the appropriate normative decision rules or when they—for whatever reason—systematically do not follow them.

A prime reason for people not to follow cognitively taxing normative decision rules is that humans are “cognitive misers” (Taylor 1981: 190) and consciously or unconsciously attempt to reduce cognitive load (Simon 1955). To achieve this, people frequently resort to heuristics, which are “rules of thumb” that can be used as “shortcuts in the thinking process” (Neale and Northcraft 1990: 57).

However, the use of heuristics frequently gives rise to systematic cognitive biases (Arkes 1991; Gilovich and Griffin 2002; Tversky and Kahneman 1974). Consider, for instance, the availability heuristic (Tversky and Kahneman 1974). It helps people judge the frequency or probability of events by how easy it is for them to recall instances of such events. If, for example, it is easy for managers to recall many bankruptcies in their industry, they will attribute a high probability to this kind of event. It is easy to see that, if managers follow this heuristic instead of the normative rule of systematically studying the entire history of bankruptcies in their industry, a few recent bankruptcies in the industry may lead managers to overestimate the overall probability of bankruptcies. Similarly, a consistent string of low-profile bankruptcies may not be top of mind and may lead managers to underestimate overall bankruptcy risk.

Prior literature identified cognitive biases as a factor that can reduce managers’ effectiveness and have other detrimental effects for managers’ firms or firms’ stakeholders. Extant research on adverse consequences of biases highlights both flawed strategic decisions (Janis 1972) as well as flawed operational processes which can substantially lower day-to-day decision quality (Peón et al. 2016) or even bring about major crises (Shefrin and Cervellati 2011).

In recent years, scholars have therefore called for more “debiasing” research to develop “debiasing techniques” that allow to attenuate or fully remedy managers’ biases (Larrick 2004; Lilienfeld et al. 2009; Milkman et al. 2009). Such “debiasing” refers to the “deliberative and theory-driven correction” (Wilson et al. 2002: 189) of biases, and debiasing

techniques can be defined as deliberate “interventions intended to reduce the magnitude of [a] bias” (Babcock et al. 1997: 916).

Most, if not all, debiasing techniques rely on the basic notion of shifting people’s thinking from an automatic, heuristic-based thinking type to a more deliberate way of thinking which adheres to the respective normative decision rules (Lilienfeld et al. 2009). For example, making decision makers accountable for their decisions can under certain circumstances exert a positive influence by encouraging the use of deliberation and adherence to normative rules (Larrick 2004; Lerner and Tetlock 1999; Rausch and Brauneis 2015). Similarly, the debiasing strategy of “considering the opposite” (Lord et al. 1984) challenges decision makers to actively reconsider their decisions, thereby forcing them to resort to more deliberate thinking.

Various scholars have attempted to develop and test such debiasing techniques and thus answer the calls for increased debiasing research (Lilienfeld et al. 2009; Milkman et al. 2009), but the current methodological requirements of such research constrain progress. Debiasing research is frequently best conducted by way of experiments, because only this type of method allows an undisputable establishment of causality (Schwenk 1982). However, as managers are frequently very busy and thus highly reluctant to participate even in simple surveys (Cycyota and Harrison 2006), it is extremely challenging to recruit them to participate in experiments, especially if their physical presence in a laboratory is required. This fact has greatly hindered managerial debiasing research, resulting in a very limited body of knowledge in the field. Consequently, the use of student subjects could be highly beneficial to help increase research output, but clarity on when the results will generalize to managers is needed.

3.2 Development of a Surrogation Framework

As is evident from the above, debiasing research is not concerned with attitudes, but with psychological processes and behavior. Therefore, the confusion about student subject surrogation described in section 2 of this manuscript is relevant for debiasing research.

To remedy the confusion, I depart from extant categorizations and propose a conceptual framework to think about and assess whether or not student subject surrogation is permissible for a focal managerial debiasing research project. Notably, this objective is intentionally comparably narrow in scope to allow for the framework to contain criteria that are sufficiently concrete, rendering the framework useful to the practicing researcher.

To guide my reasoning in the development of the framework, I leverage the metaphor of decision making as a mental journey along a path toward a decision outcome (Cornelissen 2006). Metaphorically speaking, cognitive biases “lead people astray” from the normatively correct way of thinking, whereas debiasing techniques are designed to “walk people back to the right path” or make sure they stay there all along.

If we want to be confident that a debiasing technique, which was demonstrated to work with students, is also effective with managers, we must ensure that both populations “got lost” by taking the same mental “shortcut” (Neale and Northcraft 1990: 57), and ended up in a similar place. In other words, not only must both populations actually violate the appropriate normative decision rules, but they must do so in the same ways to yield a qualitatively similarly biased outcome. Additionally, the debiasing technique must successfully trigger the same mechanism in both populations to make them return to the normatively correct path. If students and managers, for instance, got lost because they did not know the way, then explaining the way back to the path to them can only have beneficial effects in both populations if both can understand the explanation. Alternatively, if students and managers

got lost because they were too lazy to check where they are headed, then a debiasing strategy can only create positive results in both samples if it manages to increase motivation in both. In other words, the debiasing technique must have similar immediate consequences on both populations.

This metaphor highlights three criteria which I discuss in greater detail in the following, turning them into three questions which need to be answered in the affirmative to allow for the use of students in managerial debiasing research (*Figure 1*). The first two questions pertain to the focal bias, while the last one pertains to the focal debiasing technique.

=====

Insert Figure 1 about here

=====

First, *do both managers and students exhibit the focal bias?* While there is no need for students and managers to display the same degree of bias (Walters-York and Curatola 2000), it is of crucial importance that both populations display at least some level of bias. This is because a debiasing technique that works for biased students can of course not have the same effect on a manager sample that is unbiased to begin with. Such a situation is conceivable, for example, if managers are not biased because they avoid taking in what Wilson et al. term “contaminating information” (2002: 191), or if managers are able to completely debias themselves without any outside intervention (Bhattacharjee and Moreno 2002). It is worth noting that not only may an unbiased managerial population prevent the researcher from finding the same effects in biased students that he or she would find in managers, but such a situation naturally has consequences for a study’s contribution in management research:

When managers are not biased to begin with, demonstrating that a debiasing technique successfully attenuates or removes a bias in a student sample may simply not constitute a contribution to the managerial literature. However, we can of course anticipate that most managerial debiasing studies are likely to fulfill this first criterion, as the key motivation for most managerial debiasing studies is precisely the observation that managers in fact do exhibit a certain bias.

Second, *do managers and students exhibit biased behavior due to the same underlying cause?* We can only expect a debiasing technique to yield comparable results in students and managers if the bias it addresses is caused by the same psychological mechanism (Arkes 1991). Several authors suggest that there might be different mechanisms giving rise to the same cognitive biases (e.g., Arnott 2006; Baron 2008). One can, for instance, differentiate between people's failures to recognize the need for and trigger an override of initial, intuitive impulses, and people's inability to perform the override of such impulses due to lack of the corresponding required "mindware" (Stanovich 2009). This could, for example, lead to a case in which managers are biased because they do not correctly apply a decision rule they know well, whereas students might be biased because they are not aware of the normative decision rule in the first place. Another example would be a case in which both managers and students may know the decision rule in the sense that they know the right decision criterion, but students do not have the necessary ability to obtain or process the data required for the decision. For student subject surrogation to be permissible, we therefore need to be able to expect that both managers and students are similar in their knowledge of the appropriate normative decision rule as well as their ability to apply it.

Third, *does the debiasing technique result in the same first-order consequences when used by/on managers and students?* Debiasing techniques are frequently designed to cause specific first-order consequences before they lead to the desired second-order consequence, i.e.,

debiased decision making. For instance, various types of bias trainings attempt to bring about the first-order consequence of making subjects aware of a bias or understand normative decision rules (Larrick 2004). The desired first-order consequence of “consider the opposite,” as the name suggests, is to make subjects assemble an array of arguments that run counter to their initial hunches (Lord et al. 1984). As a final example, the intended first-order consequence of providing incentives is to increase subjects’ motivation (Larrick 2004). Naturally, if we want to make valid inferences from the results obtained in a student sample to a managerial population, we must be able to expect that the first-order consequences of a given implementation of a debiasing technique will be similar in both students and managers.

This criterion is particularly relevant for debiasing techniques that require some form of competent active participation of the subjects. If student subjects, for example, lack the willingness to perform a debiasing technique properly, the technique will likely not achieve its first-order objectives and the results are likely to be not generalizable to a managerial population. Similarly, if subjects lack the ability to perform a technique—for example if they cannot comprehend debiasing instructions due to specific terminology that they are not familiar with—it must also be expected that the results will not generalize (Croson 2010; Walters-York and Curatola 2000). It should be noted that this criterion does not only require that students are willing and able to execute a technique, but it also conversely requires that managers would be willing and able to do the same in a similar context. Using students to demonstrate the effect of a debiasing technique that no manager would ever perform of course also precludes generalization to a manager sample.

It is worth noting that the criteria included in the proposed framework do not strictly constitute necessary conditions for student subject surrogation (Copi, Cohen and McMahon 2016). Not fulfilling one of the criteria does not necessarily imply that the same debiasing technique cannot possibly have a debiasing effect on both students and managers. If

managers, for example, are biased for one reason, and students for a different reason, it is—while very unlikely—theoretically conceivable that the same debiasing technique may lead managers to the normatively correct outcome via one mechanism, and students via another. Then, in a way, student subject surrogation would lead to correct results although one of the three criteria of the framework is not fulfilled.

4 Using the Surrogation Framework

In the following section, I demonstrate the application of the developed conceptual framework. First, I lay out the overall approach I will take when demonstrating the framework's application. Then, I select an extant debiasing study that was conducted with managers as a reference study. I proceed to use the framework to develop five testable hypotheses regarding the generalizability of findings from a new student sample to the reference study's manager sample. This is followed by a description of the stimulus material and the employed student sample. Finally, I discuss the results.

4.1 Approach

To demonstrate the utility of the surrogation framework, I will use it to derive hypotheses about whether student surrogation is possible for a set of debiasing techniques. I will then empirically test these hypotheses to tentatively assess the quality of the framework's recommendations.

To do this, I will compare findings regarding the set of debiasing techniques in a student sample to the corresponding results obtained in a manager sample. As mentioned in the introduction, accessing manager populations for the purpose of debiasing studies is challenging. To circumvent this problem, and to avoid “oversurveying” the population of

interest even further, I will select a suitable reference study that has already been conducted with management subjects and will complement it with a replication using student subjects. This allows me to compare the results obtained in the existing reference study with results from newly collected student data and assess whether they correspond to the degree the framework would suggest.

It is important to stress that since debiasing research is interested in a population's response to a debiasing technique, the only way to correctly assess if surrogation is accessible is to compare the target and the surrogate populations' *reactions* to a focal debiasing technique (i.e., *changes* in response to the baseline stimulus material; see *Figure 2*). It makes sense, for instance, to compare whether students and managers react similarly to the potentially debiasing intervention of increasing incentives. This contrasts with other methods that are frequently used to assess surrogation, such as a direct comparison of the populations' responses to the *baseline* stimulus material, i.e., the level of bias they exhibit. While the latter may indicate surrogation problems, only the former can reliably determine any such issues (Dobbins et al. 1988b; Walters-York and Curatola 2000).

=====

Insert Figure 2 about here

=====

4.2 Selection and Description of Reference Study

Four considerations led me to select a study by Graf et al. (2012) on possible antidotes to the bias of competitive irrationality as a baseline for the effects of debiasing techniques on actual managers. First, the study was conducted fairly recently, so the overall context of the

reference subjects and the student subjects is likely to be similar and not disturbed by long-term trends, for example overall changes in attitudes or societal norms. Second, exact descriptions of the stimulus material were readily available and the study was conducted online and is hence fully replicable with the exact same material as opposed to being tied to specific locations, technical devices, or experimenters. Third, aside from the recency and replicability criteria, the Graf et al. (2012) study also fulfills criteria of breadth and variance in findings. The study includes five different experiments on diverse debiasing techniques, allowing me to conduct multiple tests of surrogation. Finally, there is a certain degree of variance in the results on the individual experiments regarding the effectiveness of the tested debiasing strategies: Some, but not all, were found to be effective. This helps to prevent floor and ceiling effects, for example the possible erroneous conclusion that specifically tested debiasing strategies worked with students when in fact *any* intervention would have had a debiasing effect.

The purpose of the study by Graf et al. (2012) was to reduce managers' competitive irrationality. Competitive irrationality refers to managers' tendency to *not maximize absolute profits*, but instead secure or improve management's own and their company's profits *relative* to competitors. The authors demonstrate that competitive irrationality constitutes a bias caused by social comparison processes (Festinger 1954; Hill and Buss 2006) and can hence be addressed using techniques from the debiasing literature (Arkes 1991; Larrick 2004).

Graf and colleagues' (2012) empirical investigation was conducted using web-based survey experiments (Kraus et al. 2017). They surveyed 934 managers that were randomly assigned to either a control or one of seven treatment groups. The researchers captured subjects' degrees of competitive irrationality by asking their choice in a pricing decision scenario which gave them four possible options (high price, medium-high price, medium-low price, low price; indicating increasing competitive irrationality in the given order). Subjects'

choices between these options were interpreted to represent their competitive irrationality on an ordinal scale. For three groups, perceived time pressure was measured on a five-point scale using an instrument with four different items. The study's control group saw the text shown in *Figure 3*. The stimulus materials of the treatment groups were slightly different to prompt the performance of the respective debiasing techniques.

=====

Insert Figure 3 about here

=====

The five tested debiasing strategies yielded differentiated results. A simple display of instructions as an ad-hoc training in biases (Larrick 2004) reduced competitive irrationality. Similarly, the reduction of time pressure (Svenson and Maule 1993) attenuated competitive irrationality. The strategy of “considering the opposite” (Lord et al. 1984), in which subjects were asked to first make and then reconsider a decision, did not result in any reduction in biased behavior. Also, giving advice from the psychologically more distanced position of a consultant or friend rather than acting as a personally involved manager (Canback 1998) did not reduce competitive irrationality. Finally, managers were made publicly accountable (Lerner and Tetlock 1999; Rausch and Brauneis 2015) by asking them to imagine having to give a guest lecture to a university class about their decision process. This technique was found to not reduce but instead increase competitive irrationality.

4.3 Hypotheses Development Using Surrogation Framework

In the following, I will use the surrogation framework proposed above to derive hypotheses on whether student subjects could have substituted managers in research testing the effectiveness of the five debiasing techniques against competitive irrationality. In my considerations, I will use the term “students” to signify students of business or related disciplines, as they are the kinds of students most managerial debiasing scholars will have access to.

Since two of the criteria in the surrogation framework pertain to the focal bias, I will assess these criteria without recourse to a particular debiasing strategy. I will then proceed to assess the third criterion separately for each debiasing technique, leading to hypotheses that are specific to each technique.

Do both managers and students exhibit the focal bias? The question of whether both samples exhibit competitive irrationality is naturally an empirical one. Graf et al.’s (2012) results indicate that 64% of managers in their control group did not respond with the normatively correct answer, and must thus be considered biased. While prior literature (e.g., Arnett and Hunt 2000) suggests that students are competitively irrational as well and the framework’s question is thus likely to be answered in the affirmative, this can only be known for sure after the replication experiment is performed and the results from its control group can be assessed. Any further hypothesis development below is thus contingent on the student control group exhibiting competitive irrationality.

Do managers and students exhibit biased behavior due to the same underlying cause? The consideration of three points supports the notion that this is indeed the case. First, prior researchers (e.g., Armstrong and Collopy 1996; Brouters et al. 2008) invoke the theoretical

construct of social comparison concerns (Festinger 1954; Garcia et al. 2013) in their studies to explain the bias of competitive irrationality in samples of experienced professionals. Likewise, extensive prior research in psychology explains the motivation of students to exhibit highly competitive behavior with social comparison concerns (Garcia et al. 2006; Garcia and Tor 2007). In both cases, the salience of a social referent triggers potentially painful comparison concerns (Brickman and Bulman 1977) and creates a motivation to best the social referent. Hence, the motivational processes underlying the bias in students and managers are the same. Second, the normative decision rule that is appropriate in the Graf et al. (2012) scenario (“Choose higher profits over lower profits”) is trivial and well-known not only to managers but also to business students, and the survey context is carefully constructed to not give the slightest hints that any other decision rule would be appropriate or any additional objectives would be relevant. The decision rule is also trivially easy to apply, since the outcomes of every option are clearly stated in the scenario description. Third, we can also be confident that there is no confusion caused by the terminology of the managerial task description. The task is simple enough so that only the most basic business terms (like “price” or “profit”) are being used, making it highly unlikely that a bias in the business student sample would be caused by a misunderstanding of the scenario description. It thus appears that managers and students are biased for the very same reasons, making the third criterion in the surrogation framework decisive for whether surrogation is permissible.

Does the debiasing technique result in the same first-order consequences when used by/on managers and students? Since the study to be replicated contains five different debiasing techniques, this question is to be answered separately for each technique, resulting in a corresponding hypothesis. It should be noted that the hypotheses always relate to the exact implementation of the debiasing techniques as performed by Graf et al. (2012) in order to allow for a valid comparison of the results.

I do not expect any difference between the samples regarding the first-order consequences of “training in biases.” The present implementation of training in biases simply requires the subjects to read a very short instruction text reminding them to be rational. Both students and managers are clearly competent to execute this task and are thus likely to similarly respond with heightened awareness of and attention to the idea of rationality, achieving the technique’s first-order objective. This consideration, in combination with the two general arguments made above, leads to the expectation that the debiasing technique has the same effect on students that it has on managers. I thus put forward the following hypothesis:

H1: Training in biases will have the same effect on student subjects and manager subjects.

Time pressure is manipulated by the presence or absence of a count-down clock which induces time pressure since the subjects can see how their time to make a decision is running out (Svenson and Maule 1993). Such time pressure was found to create stress both for students (Svenson and Maule 1993) and for managers (Hambrick, Finkelstein, and Mooney 2005). When the count-down clock is removed, this is likely to reduce the time pressure and thus achieve the first-order objective of reducing stress in both samples. Again, this effect is unlikely to be harmed by unmotivated or incapable subjects, as the debiasing technique does not require subjects’ active participation. I therefore hypothesize:

H2: Reducing time pressure will have the same effect on student subjects and manager subjects.

“Consider the opposite” explicitly instructs subjects to think of arguments against an initial impulse to act. This task requires the active participation of subjects, but is likely similarly easy or difficult for students as it is for managers. Since there is also no reason to assume different degrees of motivation to engage in this behavior in a completely voluntary

experimental setting, the first-order consequences of considering the opposite are likely to be comparable for students and managers. In combination with the two general arguments made above, I therefore hypothesize:

H3: “Considering the opposite” will have the same effect on student subjects and manager subjects.

Likewise, the situation in which one advises a third party is easy to imagine—virtually everybody has been in such a situation in their lives at some point. Consequently, students and managers are likely both competent to execute this debiasing technique. In turn, the first-order consequences of creating greater psychological distance between the subject and the decision are likely to be similar for students and managers. Accordingly, I posit:

H4: Giving advice rather than making a decision for oneself will have the same effect on student subjects and manager subjects.

Finally, the technique of creating accountability was implemented by asking subjects to imagine that they had to defend their decision process in a guest lecture at a university in front of a large audience.

Three reasons make it probable that while the present implementation of the technique is likely to bring about the first-order consequence of a sense of accountability in managers, it is unlikely to do so in a student sample. First, the focal scenario is likely to be perceived as somewhat contrived by students, as it is highly unlikely they would ever be invited as a guest speaker in their current position. This lack of realism is bound to lead students to not take the exercise seriously and thus diminishes the cognitive and affective consequences of the technique, likely altering subjects’ response to the scenario (Solomon and Samp 1998).

Second, the key notion of instilling accountability is to make people anticipate adverse consequences, including affective experiences like shame, in the case they should fail to

perform adequately (Lerner and Tetlock 1999). However, to assess the affective consequences of an event, people must mentally construct it, making use of prior experiences (Wilson and Gilbert 2003). As business students (in large German public universities) usually only have very limited experience in speaking to large groups, they will likely have difficulty adequately construing such an event and might therefore not appreciate the degree of pressure this situation puts on them. Managers, who are likely to have more experience in public speaking will have memories of such situations, which help them more adequately anticipate the level of stress this creates (Goldstein 2011).

Finally, students may view the audience as a group of peers. They might therefore expect the audience to “go easy” on them and consequently not prepare as thoroughly as the debiasing technique would require—and as actual managers would. Additionally, students may believe that they know the expectations and beliefs of their audience, and consequently fail to engage in the preemptive self-criticisms that accountability is intended to induce (Lerner and Tetlock 1999). For all these reasons, I expect the debiasing technique to not achieve its first-order objective in students and consequently expect the results from student and manager samples to be not compatible:

H5: Creating accountability will not have the same effect on student subjects and manager subjects.

4.4 Stimulus Material and Student Sample

To collect the necessary data on students to compare it to Graf et al.’s (2012) results, I followed very precisely their methodology. I obtained the exact stimulus material that was employed in the reference study. This includes not only the textual stimulus material, but also

the used software program, allowing not only for an identical introduction of the study and identical task framing, but even an identical screen layout.

In order to increase the response rate in the subject sample, the stimulus material was translated to German. Taking great care to consider psychological, linguistic, and cultural matters (Van de Vijver and Tanzer 2004), it was translated by a native German speaker who had lived and worked in the U.S. for extensive periods of time. The translation was then validated by a native English speaker with near-native command of the German language, completing the translation-backtranslation procedure (Werner and Campbell 1970). The currency used in the German version of the scenario was changed from Dollar to Euro since it is generally advisable to use the same units that would appear in real-life situations (Aiman-Smith et al. 2002).

In line with the objective of the study, the subjects used in this research project differ from the reference study's population in that the sample is not comprised of managers, but of business and economics students from three large German public universities as well as a large German marketing student club. I selected students of business and economics to ensure that they could comprehend the terminology of the stimulus material and were familiar with the general problem to be solved. I also deliberately selected a non-MBA sample, as it can be argued that MBA students frequently closely resemble actual managers with regard to a variety of characteristics, most notably work experience.

Approximately 7,100 students were addressed via e-mail and invited to participate in the experiments. I received a total of 1,423 usable responses, leading to an effective response rate of ~20%.² This response rate is comparable to the 25% (934 responses) obtained in the

² The number of students in the marketing club pool cannot be precisely determined because the invitation e-mail was manually forwarded to members by their respective local chapter chairpersons. The approximation builds on the club's member database and is hence likely to be conservative as the chairpersons partially only e-mailed members that frequently attended chapter meetings. Similarly, the number of students on the mailing list at one of the universities could not be determined exactly and was conservatively estimated by the mailing list

reference study. To test for non-response bias, I followed Armstrong and Overton's (1977) logic of considering late responders as a proxy for non-responders. I split responses of the largest participating university's sub-sample into two equally sized groups of early and late responders. A Mann-Whitney U-test indicated that the two populations did not differ with regard to their responses to the stimulus material ($p(\text{two tailed})=.916$) and hence suggested no problem of non-response bias.

The mean age of the students in the sample was 23.7 years. The mean number of semesters studied was 5.3 and the distribution shows the expected typical German pattern of a larger number of students beginning their studies in the fall semester. The sample exhibited an almost identical share of female and male students and 93% of the participants were German nationals.

4.5 Results

Figure 4 shows the decision making patterns of the student subjects, with darker areas of the bars indicating more competitively irrational answers. It is obvious that a majority of students made a non-rational, i.e., biased decision. This confirms the conjecture made in response to the first question of the surrogation framework and allows to proceed as it is now clear that both the manager and the student sample have the potential to display a positive reaction to debiasing interventions.

=====

Insert Figure 4 about here

=====

administrator, an employee of the university. Hence, the reported response rate represents a lower bound on the actual response rate.

To be able to compare my results about the effectiveness of the five debiasing strategies to the results obtained by Graf et al. (2012), I used the exact same testing logic that was employed in their study. For four debiasing techniques, the central tendency regarding subjects' competitive irrationality was compared between the control group and the respective treatment group(s). For the fifth debiasing technique of reducing time pressure, non-parametric statistical methods were used to identify a possible correlation between time pressure as perceived by the subjects (which was separately captured via the same scale that was used in the reference study) and their competitive irrationality. Non-parametric methods were employed since the dependent variable of competitive irrationality is not interval-scaled and hence not normally distributed (Gilpin 1993; Kendall 1938). As a robustness check, I also repeated all tests with their parametric counterparts and performed chi-squared tests (comparing dichotomized “more” and “less” rational responses) where possible. I obtained similar results.

As is indicated in *Table 3*, the two tests regarding training in biases (H1) and reducing time pressure (H2) yielded statistically significant results, while the other three debiasing techniques (H3 though H5) yielded null findings.

=====

Insert Table 3 about here

=====

These results largely concur with the findings by Graf et al. (2012) (see *Table 4*). With the exception of the “accountability” technique, the reference findings for all debiasing techniques were fully reproduced with the student sample. The detrimental effect of

accountability on decision quality that Graf et al. (2012) obtained in their study is not found in the student sample.

=====
Insert Table 4 about here
=====

5 Discussion

I will interpret and discuss the study's results in the following. Subsequently, I will explicate the study's contributions, as well as its key implication for research.

5.1 Interpretation of Results

The results of the debiasing experiments provide initial support for the surrogation framework proposed in this paper. This is the case for both the results that replicate the findings obtained by Graf et al. (2012) and those that do not.

Four out of five techniques replicated the results described in the initial study and hence directly support hypotheses H1 through H4. Two techniques led to significant results in both samples, indicating that debiasing was successfully performed and confirming H1 and H2. Two other techniques yielded null results in both samples, supporting H3 and H4.

The fifth debiasing technique, the creation of accountability, was found to have adverse effects on decision quality in the reference study. This finding was not replicated. This supports H5, which argued that at least the specific implementation of the debiasing technique chosen by Graf et al. (2012) would not lead to results that would be compatible

between student and manager samples. The stimulus material asked the experiment subjects to imagine giving a guest lecture to a group of business students. Subjects that are students themselves might have perceived this situation as more contrived and harder to imagine than managers. Hence, the stimulus material likely failed to manipulate accountability in the student subjects and thus did not achieve the debiasing technique's intended first-order consequences.

Overall, the empirical evidence produced by my study supports not only the individual hypotheses but it thereby also corroborates the proposed surrogation framework. It appears that business students can be a viable substitute for managers in debiasing research under the condition that they fulfill the criteria laid out in the surrogation framework.

5.2 Contributions

This study contributes to the literature in several ways. First, it contributes to the student surrogation literature by offering a structured overview of the debate in business research and beyond. It thus provides an account of the development of the debate and it integrates several hitherto highly stratified literature streams.

Second, this study makes a methodological contribution in the domain of managerial debiasing. While Libby et al. (2002) already advanced the notion that one should use subjects only as sophisticated as the goals of an experiment require, there is great uncertainty about how such requirements are to be assessed. This study consequently adds to the literature by offering and tentatively testing a conceptual framework that can be used to think about and assess student subject surrogation. Unlike prior categorization schemes, the proposed framework is specific to the question of surrogation in managerial debiasing. This makes the framework sufficiently concrete so that it can be applied in research practice.

Finally, this study makes an empirical contribution to the managerial debiasing literature by replicating the debiasing study of Graf et al. (2012) on competitive irrationality. First, it thereby strengthens the credibility of the findings of this reference study to the degree that they were replicated. This is critical for the integrity of the body of knowledge in the field of debiasing, as all scientific studies are merely speculative until replicated (Easley et al. 2000; Hubbard et al. 1998). Second, the replication performed in this paper constitutes, to the author's best knowledge, the very first time that results of identical debiasing efforts in managers and student subjects are reported in the literature.³

5.3 Implications for research

The study's key implications derive from the methodological and empirical contributions, namely the proposed surrogation framework and the empirical findings which corroborate it. These have implications for debiasing researchers and professional gatekeepers alike. Despite substantial prior doubts about the utility of student subjects, students appear to be acceptable surrogates for managers in management debiasing research—but only under the condition that certain specific criteria are fulfilled. At the beginning of research projects, researchers might wish to use the framework to perform a structured assessment of whether student subject surrogation is advisable. If the questions in the surrogation framework can be answered in the affirmative, researchers may feel reassured in using students as subjects in their focal study. If the framework's criteria are not fulfilled, however, it appears highly doubtful that student surrogation would be appropriate.

³ Hodgkinson et al. (1999) report findings for students and managers, but they use different stimulus materials for the two samples. Kennedy (1993, 1995) also reports debiasing results for managers and students. However, her student sample is largely comprised of managers with more than five years of experience that were enrolled in an executive program. The sample therefore represents actual managers more than student surrogate subjects.

Since my empirical results indicate that student subject surrogation can work under certain circumstances, professional gatekeepers such as editors and reviewers might—in precisely these circumstances—wish to become slightly more tolerant of business student samples in debiasing research than they have been in the past. Because gatekeepers can now refer to the proposed conceptual framework to structure their thinking about whether they believe student subject surrogation to be credible in a specific case, this does not require them to sacrifice the necessary scientific rigor in their assessments. In this way, gatekeepers could spur more research and increase the speed of advances in the field, thereby responding to the urgent calls for more debiasing research (Milkman et al. 2009).

6 Limitations and Future Research

As any empirical research endeavor, this study is subject to limitations. An empirical limitation, for instance, is that I did not attempt to provide specific insights about the strength of any debiasing effects, i.e., about effect sizes. Instead, I followed Locke’s approach in focusing “primarily on similarity in the direction of effects obtained” (1986: 6), largely because reliably replicating the effect sizes of the manager sample would have required impractically large samples (Maxwell et al. 2008; Simonsohn 2015). It appears quite conceivable that a debiasing technique might have different degrees of effectiveness in a student population as compared to a manager population (Ashton and Kramer 1980; Bolton et al. 2012; Calder et al. 1981). This is particularly the case when the bias has a different initial prevalence in the populations, for example when students are more systematically biased than managers to begin with, as is the case with competitive irrationality. The strength of debiasing effects might also differ because the same intensity of a debiasing measure may not always cause an identically large effect. For example, the same financial incentives to increase motivation (Larrick 2004) will in all likelihood yield a smaller effect on managers

than on students (Fréchette 2015). While this limitation has only minor consequences for the present study because the limitation only concerns the magnitude of a debiasing effect, not the question of whether an effect exists at all, it opens avenues for future research. Detailed studies of differences in subjects' degree of susceptibility to debiasing techniques, and the ensuing different magnitudes of debiasing effects, constitute great opportunities for scholars with access to sufficiently large subject pools.

Furthermore, a methodological limitation of my study is the use of hypothetical scenarios without financial incentives. Some authors assert that hypothetical questions in surveys may produce misleading results. Wallis and Friedman, for example, argued that it "is questionable whether a subject in so artificial an experimental situation could know what choices he would make in an economic situation" (1942: 179). However, newer research finds that there is "little, if any, evidence for this characterization. [...] Hypothetical questions appear to work well when subjects have access to their intuitions and have no particular incentive to lie" (Thaler 1987: 120). In fact, empirical evidence shows that subjects' stated intentions are a good predictor of actual behavior (Ajzen 1991) and there exist numerous studies that report parallel experiments with and without incentives for the subjects, finding hardly any influence of incentives (for a review, see Thaler 1987). It is thus unlikely that the absence of incentives constitutes a problem in this study. Nevertheless, future researchers may wish to conduct similar studies using material incentives.

My study opens several avenues for future research. In particular, the robustness of my proposed framework and empirical results should be rigorously tested and possibly enhanced by replicating the findings (Easley et al. 2000; Hubbard et al. 1998) and by attempting comparisons between students and managers with respect to other biases, for example the confirmation or anchoring biases, and other debiasing strategies, for example training in rules and representations, group decision making, or the use of decision support systems (Larrick

2004). In light of these opportunities for research, it is also worth highlighting that assessing the framework's second criterion requires a thorough understanding of the psychological processes that give rise to the biases to be remedied. To be able to provide sufficiently certain answers to this question, future research on the detailed nature of some biases may be needed (Fiedler 2016).

Furthermore, future research could attempt to go beyond the experimental set-up devised by Graf et al. (2012). In the spirit of "close" replication research (Easley et al. 2000: 85), I deliberately decided to not include any additional questions in my study to remain fully consistent with the reference study's stimulus material, as only such consistency enabled me to directly compare results from the Graf et al. (2012) sample with my newly obtained results. However, in particular with respect to H5, future researchers might wish to make this trade-off differently and include a manipulation check to assess whether a debiasing technique actually achieved its intended first-order consequences, for example by measuring mediating variables like perceived accountability in accountability creation (Fiedler 2016; Kidd 1976). The results of such a manipulation check would potentially allow a more nuanced interpretation of the subjects' behavior and increase researchers' confidence that the third question in the framework was answered adequately.

Finally, additional opportunities for research include, for instance, the question of whether survey experiments on debiasing actually generalize to real-world situations (Jung 1981). It could be addressed, for example, by conducting more complex business simulations as opposed to comparatively simplistic one-shot survey experiments. Also, it could be interesting to study if it is possible to conduct managerial debiasing research with populations that are possibly even less sophisticated than business students, such as, for example, workers on Amazon's Mechanical Turk platform (Paolacci and Chandler 2014).

References

- Abdel-khalik AR (1974) On the efficiency of subject surrogation in accounting research. *Accounting Review* 49(4):743-750.
- Abdellaoui M, Bleichrodt H, Kammoun H (2013) Do financial professionals behave according to Prospect Theory? An experimental study. *Theory and Decision* 74(3):411-429. doi: 10.1007/s11238-011-9282-3
- Aiman-Smith L, Scullen SE, Barr SH (2002) Conducting studies of decision making in organizational contexts: A tutorial for policy-capturing and other regression-based techniques. *Organizational Research Methods* 5(4):388-414. doi: 10.1177/109442802237117
- Ajzen I (1991) The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 50(2):179–211. doi: 10.1016/0749-5978(91)90020-T
- Alpert B (1967) Non-businessmen as surrogates for businessmen in behavioral experiments. *Journal of Business* 40(2):203-207. doi: 10.1086/294956
- Arkes HR (1991) Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin* 110(3):486-498. doi: 10.1037/0033-2909.110.3.486
- Armstrong JS, Collopy F (1996) Competitor Orientation: effects of objectives and information on managerial decisions and profitability. *Journal of Marketing Research* 33(2):188-199. doi: 10.2307/3152146
- Armstrong JS, Overton TS (1977) Estimating nonresponse bias in mail surveys. *Journal of Marketing Research* 14(3):396-402. doi: 10.2307/3150783
- Arnett DB, Hunt SD (2002) Competitive irrationality: The influence of moral philosophy. *Business Ethics Quarterly* 12(3):279-303. doi: 10.2307/3858018
- Arnott D (2006) Cognitive biases and decision support systems development: A design science approach. *Information Systems Journal* 16(1):55-78. doi: 10.1111/j.1365-2575.2006.00208.x
- Ashton RH, Kramer SS (1980) Students as surrogates in behavioral accounting research: Some evidence. *Journal of Accounting Research* 18(1):1-15. doi: 10.2307/2490389
- Babcock L, Loewenstein G, Issacharoff S (1997) Creating Convergence: Debiasing Biased Litigants. *Law & Social Inquiry* 22(4):913–925. doi 10.1111/j.1747-4469.1997.tb01092.x
- Ball SB, Cech P-A (1996) Subject pool choice and treatment effects in economic laboratory research. In: Mark Isaac R (ed) *Research in experimental economics*. JAI Press, Greenwich, pp. 239-292.
- Baron J (2008) *Thinking and deciding*, 4th edn. Cambridge University Press, New York.
- Barr SH, Hitt MA (1986) A comparison of selection decision models in manager versus student samples. *Personnel Psychology* 39(3):599-617. doi: 10.1111/j.1744-6570.1986.tb00955.x
- Bateman TS, Zeithaml CP (1989) The psychological context of strategic decisions: A test of relevance to practitioners. *Strategic Management Journal* 10(6):587-592. doi: 10.1002/smj.4250100606
- Bhattacharjee S, Moreno K (2002) The impact of affective information on the professional judgments of more experienced and less experienced auditors. *Journal of Behavioral Decision Making* 15(4):361-377. doi: 10.1002/bdm.420
- Beltramini RF (1983) Student surrogates in consumer research. *Journal of the Academy of Marketing Science* 11(4):438-443. doi: 10.1177/009207038301100406
- Bettis RA, Helfat CE, Shaver JM (2016) The necessity, logic, and forms of replication. *Strategic Management Journal* 37:2193-2203. doi:10.1002/smj.2580
- Bolton GE, Ockenfels A, Thonemann UW (2012) Managers and students as newsvendors. *Management Science* 58(12):2225-2233. doi: 10.1287/mnsc.1120.1550

- Brickman P, Bulman RJ (1977) Pleasure and pain in social comparison. In: Suls JM, Miller RL (eds) *Social comparison processes: Theoretical and empirical perspectives*. Hemisphere, Washington, DC.
- Brouthers L, Lascu DN, Werner S (2008). Competitive Irrationality in Transitional Economies: Are Communist Managers Less Irrational? *Journal of Business Ethics* 83(3):397–408. doi: 10.1007/s10551-007-9627-6
- Burnett JJ, Dunne PM (1986) An appraisal of the use of student subjects in marketing research. *Journal of Business Research* 14(4):329-343. doi: 10.1016/0148-2963(86)90024-X
- Calder BJ, Phillips LW, Tybout AM (1981) Designing research for application. *Journal of Consumer Research* 8(2):197-207. doi: 10.1086/208856
- Campbell DT (1957) Factors relevant to the validity of experiments in social settings. *Psychological Bulletin* 54(2):297-312. doi: 10.1037/h0040950
- Campbell DT, Stanley JC (1963) *Experimental and quasi-experimental designs for research*. Houghton Mifflin Company, Boston.
- Canback S (1998) The logic of management consulting: Part 1. *Journal of Management Consulting* 10(2):3-11.
- Colquitt JA (2008) Publishing laboratory research in AMJ: A question of when, not if. *Academy of Management Journal* 51(4):616-620. doi: 10.5465/AMJ.2008.33664717
- Cooper DJ, Kagel JH, Lo W, Gu QL (1999) Gaming against managers in incentive systems: Experimental results with Chinese students and Chinese managers. *American Economic Review* 89(4):781-804. doi: 10.1257/aer.89.4.781
- Copeland RM, Francia AJ, Strawser RH (1973) Students as subjects in behavioral business research. *Accounting Review* 48(2):365-372.
- Copi IM, Cohen C, McMahon K (2016) *Introduction to Logic*, 14th edn. Routledge, London.
- Cornelissen JP (2006) Making Sense of Theory Construction: Metaphor and Disciplined Imagination. *Organization Studies* 27(11):1579-1597. doi: 10.1177/0170840606068333
- Crant JM, Bateman TS (1990) An experimental test of the impact of drug-testing programs on potential job applicants' attitudes and intentions. *Journal of Applied Psychology* 75(2):127-131. doi: 10.1037/0021-9010.75.2.127
- Crosen RTA (2010) The use of students as participants in experimental research. <http://experimental-instruments.com/BDOM/2010/students.pdf>. Accessed on 1 December 2015.
- Cunningham WH, Anderson Jr. WT, Murphy JH (1974) Are students real people? *Journal of Business* 47(3):399-409. doi: 10.1086/295654
- Cycyota CS, Harrison DA (2006) What (Not) to expect when surveying executives: A meta-analysis of top manager response rates and techniques over time. *Organizational Research Methods* 9(2):133-160. doi: 10.1177/1094428105280770
- Dipboye RL, Fromkin HL, Wiback K (1975) Relative importance of applicant sex, attractiveness, and scholastic standing in evaluation of job applicant resumes. *Journal of Applied Psychology* 60(1):39-43. doi: 10.1037/h0076352
- Dobbins GH, Lane IM, Steiner DD (1988a) A note on the role of laboratory methodologies in applied behavioural research: Don't throw out the baby with the bath water. *Journal of Organizational Behavior* 9(3):281-286. doi: 10.1002/job.4030090308
- Dobbins GH, Lane IM, Steiner DD (1988b) A further examination of student babies and student bath water: A response to Slade and Gordon. *Journal of Organizational Behavior* 9(4): 377-378. doi: 10.1002/job.4030090410
- Easley RW, Madden CS, Dunn MG (2000) Conducting marketing science: The role of replication in the research process. *Journal of Business Research* 48(1):83-92. doi: 10.1016/S0148-2963(98)00079-4

- Enis BM, Cox KK, Stafford JE (1972) Students as subjects in consumer behavior experiments. *Journal of Marketing Research* 9(1):71-74. doi: 10.2307/3149612
- Evanschitzky H, Baumgarth C, Hubbard R, Armstrong JS (2007) Replication research's disturbing trend. *Journal of Business Research* 60(4):411-415. doi: 10.1016/j.jbusres.2006.12.003
- Farber ML (1952) The college student as laboratory animal. *American Psychologist* 7(3):102-102. doi: 10.1037/h0059045
- Festinger L (1954) A theory of social comparison processes. *Human Relations* 7(2):117-140. doi: 10.1177/001872675400700202
- Fiedler K (2016) Functional research and cognitive-process research in behavioural science: An unequal but firmly connected pair. *International Journal of Psychology* 51:64-71. doi:10.1002/ijop.12163
- Fischhoff B (1982) Debiasing. In: Kahneman D, Slovic P, Tversky A (ed) *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press, Cambridge, UK, pp. 422-444.
- Fleming JE (1969) Managers as subjects in business decision research. *Academy of Management Journal* 12(1):59-66. doi: 10.2307/254672
- Fox S, Spector PE, Miles D (2001) Counterproductive work behavior (CWB) in response to job stressors and organizational justice: Some mediator and moderator tests for autonomy and emotions. *Journal of Vocational Behavior* 59(3):291-309. doi: 10.1006/jvbe.2001.1803
- Fréchette GR (2015) Laboratory experiments: Professionals versus students. In: Fréchette GR and Schotter A (ed) *Handbook of Experimental Economic Methodology*. Oxford University Press, Oxford, pp. 360-390.
- Garcia SM, Tor A (2007). Rankings, standards, and competition: Task vs. scale comparisons. *Organizational Behavior and Human Decision Processes* 102(1):95-108. doi: 10.1016/j.obhdp.2006.10.004
- Garcia SM, Tor A, Gonzalez RD (2006). Ranks and rivals: A theory of competition. *Personality and Social Psychology Bulletin* 32(7):970-982. doi: 10.1177/0146167206287640
- Garcia SM, Tor A, Schiff T (2013) The Psychology of Competition: A Social Comparison Perspective: *Perspectives on Psychological Science* 8(6):634-650. doi: 10.1177/1745691613504114
- Gilovich T, Griffin D (2002) Introduction - Heuristics and biases: Then and now. In: Gilovich T, Griffin D, Kahneman D (ed) *Heuristics and biases, the psychology of intuitive judgment*. Cambridge University Press, Cambridge, pp. 379-396.
- Gilpin AR (1993) Table for conversion of Kendall's Tau to Spearman's Rho within the context of measures of magnitude of effect for meta-analysis. *Educational and Psychological Measurement* 53(1):87-92. doi: 10.1177/0013164493053001007
- Goldstein EB (2011) *Cognitive Psychology: Connecting Mind, Research, and Everyday Experience*, 3rd edn. Cengage Learning, Wadsworth.
- Gordon ME, Slade LA, Schmitt N (1986) The science of the sophomore revisited: From conjecture to empiricism. *Academy of Management Review* 11(1):191-207. doi: 10.2307/258340
- Gordon ME, Slade LA, Schmitt N (1987) Student guinea pigs: Porcine predictors and particularistic phenomena. *Academy of Management Review* 12(1):160-163. doi: 10.2307/258002
- Graf L, König A, Enders A, Hungenberg H (2012) Debiasing competitive irrationality: How managers can be prevented from trading off absolute for relative profit. *European Management Journal* 30(4):386-403. doi: 10.1016/j.emj.2011.12.001

- Greenberg J (1987) The college sophomore as guinea pig: Setting the record straight. *Academy of Management Review* 12(1):157-159. doi: 10.2307/258001
- Hambrick DC, Finkelstein S, Mooney AC (2005) Executive job demands: New insights for explaining strategic decisions and leader behaviors. *Academy of Management Review* 30(3):472-491. doi: 10.5465/AMR.2005.17293355
- Harris JR, Sutton CD (1995) Unravelling the ethical decision-making process: Clues from an empirical study comparing Fortune 1000 executives and MBA students. *Journal of Business Ethics* 14(10):805-817. doi: 10.1007/BF00872347
- Hawkins DI, Albaum G, Best R (1977) An investigation of two issues in the use of students as surrogates for housewives in consumer behavior studies. *Journal of Business* 50(2):216-222. doi: 10.1086/295932
- Hill SE, Buss DM, (2006) Envy and positional bias in the evolutionary psychology of management. *Managerial and Decision Economics* 27(2-3):131-143. doi: 10.1002/mde.1288
- Hodgkinson GP, Bown NJ, Maule AJ, Glaister KW, Pearman AD (1999) Breaking the frame: An analysis of strategic cognition and decision making under uncertainty. *Strategic Management Journal* 20(10):977-985. doi: 10.1002/(SICI)1097-0266(199910)20:10<977::AID-SMJ58>3.0.CO;2-X
- Hofstede TR (1972) Some behavioral parameters of financial analysis. *Accounting Review* 47(4):679-692.
- Hubbard R, Vetter DE, Little EL (1998) Replication in strategic management: Scientific testing for validity, generalizability, and usefulness. *Strategic Management Journal* 19(3):243-254. doi: 10.1002/(SICI)1097-0266(199803)19:3<243::AID-SMJ951>3.0.CO;2-0
- Hughes CT, Gibson ML (1991) Students as surrogates for managers in a decision-making environment: An experimental study. *Journal of Management Information Systems* 8(2):153-166. doi: 10.1080/07421222.1991.11517925
- Hume D (1999) An enquiry concerning human understanding, Oxford philosophical texts. Oxford University Press, Oxford (Original work published 1739).
- Janis IL (1972) *Victims of Groupthink: A Psychological Study of Foreign-Policy Decisions and Fiascoes*. Houghton Mifflin, Boston.
- Jung J (1981) Is it possible to measure generalizability from laboratory to life, and is it really that important? In: Silverman I (ed) *Generalizing from laboratory to life*. Jossey-Bass, San Francisco, pp. 39-49.
- Kahneman D, Tversky A (1973) On the psychology of prediction. *Psychological Review* 80(4):237-251. doi: 10.1037/h0034747
- Kahneman D, Tversky A (1979) Prospect Theory: An analysis of decision under risk. *Econometrica* 47(2):263-292. doi: 10.2307/1914185
- Kaufmann L, Carter CR, Buhmann C (2012) The impact of individual debiasing efforts on financial decision effectiveness in the supplier selection process. *International Journal of Physical Distribution and Logistics Management* 42(5):411-433. doi: 10.1108/09600031211246492
- Kaustia M, Perttula M (2012) Overconfidence and debiasing in the financial industry. *Review of Behavioral Finance* 4(1):46-62. doi: 10.1108/19405971211261100
- Kendall MG (1938) A new measure of rank correlation. *Biometrika* 30(1/2):81-93. doi: 10.2307/2332226
- Kennedy J (1993) Debiasing audit judgment with accountability: A framework and experimental results. *Journal of Accounting Research* 31(2):231-245. doi: 10.2307/2491272

- Kennedy J (1995) Debiasing the curse of knowledge in audit judgment. *The Accounting Review* 70(2):249-273.
- Khera IP, Benson JD (1970) Are students really poor substitutes for businessmen in behavioral research? *Journal of Marketing Research* 7(4):529–532. doi: 10.2307/3149650
- Kidd RF (1976) Manipulation checks: Advantage or disadvantage? *Representative Research in Social Psychology* 7(2):160-165.
- Kohli AK (2011) From the editor: Reflections on the review process. *Journal of Marketing* 75(6):1-4. doi: 10.1509/jm.75.6.editorial
- Kraus S, Meier F, Niemand T, Bouncken RB, Ritala P (2017) In search for the ideal cooperation partner: an experimental study. *Review of Managerial Science*. doi: 10.1007/s11846-017-0237-0.
- Kruglanski AW (1975) The human subject in the psychology experiment: Fact and artifact. In: Berkowitz L (ed) *Advances in experimental social psychology*. Academic Press, New York, pp. 101-147.
- Lamb CW, Stern DE (1980) An evaluation of students as surrogates in marketing studies. *Advances in Consumer Research* 7(1):796-799.
- Larrick RP (2004) Debiasing. In: Koehler DJ, Harvey N (ed) *Blackwell handbook of judgment and decision making*. Blackwell Publishing, Malden, MA, pp. 316-337.
- Latham GP, Dossett DL (1978) Designing incentive plans for unionized employees: A comparison of continuous and variable ration reinforcement schedules. *Personnel Psychology* 31(1):47-61. doi: 10.1111/j.1744-6570.1978.tb02108.x
- LaTour M, Champagne PJ, Rhiel GS, Behling R (1990) Are students a viable source of data for conducting survey research on organizations and their environments? *Review of Business and Economic Research* 26(1):68-82.
- Lerner JS, Tetlock PE (1999) Accounting for the effects of accountability. *Psychological Bulletin* 125(2):255-275. doi: 10.1037/0033-2909.125.2.255
- Libby R, Bloomfield R, Nelson MW (2002) Experimental research in financial accounting. *Accounting, Organizations and Society* 27(8):775-810. doi: 10.1016/S0361-3682(01)00011-3
- Lilienfeld SO, Ammirati R, Landfield K (2009) Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare? *Perspectives on Psychological Science* 4(4):390-398. doi: 10.1111/j.1745-6924.2009.01144.x
- Liyanarachchi GA (2007) Feasibility of using student subjects in accounting experiments: A review. *Pacific Accounting Review* 19(1):47-67. doi: 10.1108/01140580710754647
- Liyanarachchi GA, Milne MJ (2005) Comparing the investment decisions of accounting practitioners and students: An empirical study on the adequacy of student surrogates. *Accounting Forum* 29(2):121-135. doi: 10.1016/j.accfor.2004.05.001
- Locke EA (ed) (1986) *Generalizing from laboratory to field settings*. Lexington Books, Lexington, MA.
- Lord CG, Lepper MR, Preston E (1984) Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology* 47(6):1231-1243. doi: 10.1037//0022-3514.47.6.1231
- Maxwell SE, Kelley K, Rausch JR (2008) Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology* 59(1):537-563. doi: 10.1146/annurev.psych.59.103006.093735
- McNemar Q (1946) Opinion-attitude methodology. *Psychological Bulletin* 43(4):289-374. doi: 10.1037/h0060985
- Meissner P, Wulf T (2013) Cognitive benefits of scenario planning: Its impact on biases and decision quality. *Technological Forecasting and Social Change* 80(4):801-814. doi: 10.1016/j.techfore.2012.09.011

- Meissner P, Wulf T (2016) Debiasing illusion of control in individual judgment: The role of internal and external advice seeking. *Review of Managerial Science* 10:245-263. doi: 10.1007/s11846-014-0144-6
- Milkman KL, Chugh D, Bazerman MH (2009) How can decision making be improved? *Perspectives on Psychological Science* 4(4):379-383. doi: 10.1111/j.1745-6924.2009.01142.x
- Miller HE (1966) Discussion of the accounting period concept and its effect on management decisions. *Empirical Research in Accounting: Selected Studies 1966, Supplement to Journal of Accounting Research* 4(1):15-17. doi: 10.2307/2490164
- Miller GR, Fontes NE, Boster FJ, Sunnafrank MJ (1983) Methodological issues in legal communication research: What can trial simulations tell us. *Communications Monographs* 50(1):33-46. doi: 10.1080/03637758309390152
- Mook DG (1983) In defense of external invalidity. *American Psychologist* 38(4):379-387. doi: 10.1037/0003-066X.38.4.379
- Moskowitz H (1971) Managers as partners in business decision research. *Academy of Management Journal* 14(3):317-325. doi: 10.2307/255076
- Murphy KR, Thornton GC, Reynolds DH (1990) College students' attitudes toward employee drug testing programs. *Personnel Psychology* 43(3):615-631. doi: 10.1111/j.1744-6570.1990.tb02399.x
- Neale MA, Northcraft GB (1990) Experience, expertise, and decision bias in negotiation: The role of strategic conceptualization. In: Sheppard BH, Bazerman MH, Lewicki RJ (eds) *Research on Negotiation in Organizations: A Biannual Research Series, Vol. 2*. AI Press, Greenwich CT: pp. 55-75
- Oakes W (1972) External validity and the use of real people as subjects. *American Psychologist* 27(10):959-962. doi: 10.1037/h0033454
- Paolacci G, Chandler J (2014) Inside the turk: Understanding mechanical turk as a participant pool. *Current Directions in Psychological Science* 23(3):184-188. doi: 10.1177/0963721414531598
- Peón D, Antelo M, Calvo A (2016) Overconfidence and risk seeking in credit markets: An experimental game. *Review of Managerial Science* 10(3):511-552. doi: 10.1007/s11846-015-0166-8
- Peterson RA (2001) On the use of college students in social science research: Insights from a second-order meta-analysis. *Journal of Consumer Research* 28(3):450-461. doi: 10.1086/323732
- Quoidbach J, Gilbert DT, Wilson TD (2013) The end of history illusion. *Science* 339 (6115): 96-98. doi: 10.1126/science.1229294
- Rausch A, Brauneis A (2015) The effect of accountability on management accountants' selection of information. *Review of Managerial Science* 9(3):487-521. doi: 10.1007/s11846-014-0126-8
- Remus W (1986) Graduate students as surrogates for managers in experiments on business decision making. *Journal of Business Research* 14(1):19-25. doi: 10.1016/0148-2963(86)90053-6
- Roberts BW, Walton KE, Viechtbauer W (2006) Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin* 132(1):1-25. doi: 10.1037/0033-2909.132.1.1
- Roth AE (1995) Introduction to experimental economics. In: Kagel JH, Roth AE (ed) *The handbook of experimental economics*. Princeton University Press, Princeton, pp. 3-109.
- Schultz DP (1969) The human subject in psychological research. *Psychological Bulletin* 72(3):214-228. doi: 10.1037/h0027880

- Schwenk CR (1982) Why sacrifice rigour for relevance? A proposal for combining laboratory and field research in strategic management. *Strategic Management Journal* 3(3):213-225. doi: 10.1002/smj.4250030304
- Schoemaker PJH (1993) Multiple scenario development: Its conceptual and behavioral foundation. *Strategic Management Journal* 14(3):193-213. doi: 10.1002/smj.4250140304
- Sears DO (1986) College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology* 51(3):515-530. doi: 10.1037/0022-3514.51.3.515
- Shefrin H, Cervellati EM (2011) BP's failure to debias: Underscoring the importance of behavioral corporate finance. *Quarterly Journal of Finance* 1(1):127-168. doi: 10.1142/S2010139211000043
- Shuptrine FK (1975) On the validity of using students as subjects in consumer behavior investigations. *Journal of Business* 48(3):383-390. doi: 10.1086/295763
- Simon HA (1955) A behavioral model of rational choice. *Quarterly Journal of Economics* 69(1):99-118. doi: 10.2307/1884852
- Simonsohn U (2015) Small telescopes detectability and the evaluation of replication results. *Psychological Science* 26(5):559-569. doi: 10.1177/0956797614567341
- Slade LA, Gordon ME (1988) On the virtues of laboratory babies and student bath water: A reply to Dobbins, Lane, and Steiner. *Journal of Organizational Behavior* 9(4):373-376. doi: 10.1002/job.4030090409
- Solomon DH, Samp JA (1998) Power and problem appraisal: Perceptual foundations of the chilling effect in dating relationships. *Journal of Social and Personal Relationships* 15(2): 191-209. doi: 10.1177/0265407598152004
- South JC (1974) Achievement motivation among managers of small businesses, corporation managers, and business students. *Journal of Applied Psychology* 59(4):509-510. doi: 10.1037/h0037342
- Stahl MJ, Harrell AM (1982) Evolution and validation of a behavioral decision theory measurement approach to achievement, power, and affiliation. *Journal of Applied Psychology* 67(6):744-751. doi: 10.1037//0021-9010.67.6.744
- Stanovich KE (2009) Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory? In: Evans J, Frankish K (ed) *In two minds: Dual processes and beyond*. Oxford University Press, Oxford, pp. 55-88.
- Staw BM (1997) The escalation of commitment: An update and appraisal. In: Shapira Z (ed) *Organizational Decision Making*. Cambridge University Press, New York, NY, pp. 191-215.
- Svenson O, Maule AJ (ed) (1993) *Time pressure and stress in human judgment and decision making*. Plenum Press, New York.
- Taylor SE (1981) The interface of cognitive and social psychology. In: Harvey JH (ed) *Cognition, social behavior, and the environment*. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 189-211.
- Thaler RH (1987) The psychology of choice and the assumptions of economics. In: Roth AE (ed) *Laboratory experimentation in economics. Six points of view*. Cambridge University Press, Cambridge, UK, pp. 99-130.
- Trotman KT (1996) *Research methods for judgment and decision making studies in auditing*. Coopers and Lybrand and Accounting Association of Australia and New Zealand, Melbourne.
- Tversky A, Kahneman D (1974) Judgment under uncertainty: Heuristics and biases. *Science* 185(4157):1124-1131. doi: 10.1126/science.185.4157.1124
- Tversky A, Kahneman D (1986) Rational choice and the framing of decisions. *Journal of Business* 59(4):251-278. doi: 10.1007/978-3-642-74919-3_4

- Van de Vijver F, Tanzer NK (2004) Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology* 54(2):119-135. doi: 10.1016/j.erap.2003.12.004
- Wallis WA, Friedman M (1942) The empirical derivation of indifference functions. In: Lange O, McIntyre F, Yntema TO (ed) *Studies in mathematical economics and econometrics*. In memory of Henry Schultz. University of Chicago Press, Chicago, IL, pp. 175-189.
- Walters-York M, Curatola AP (2000) Theoretical reflections on the use of students as surrogate subjects in behavioral experimentation. *Advances in Accounting Behavioral Research* 3:243-263.
- Werner O, Campbell DT (1970) Translation, wording through interpreters, and the problem of decentering. In: Naroll R, Cohen R (ed) *A handbook of method in cultural anthropology*. Natural History Press, Garden City, NJ, pp. 398-420.
- Wilson TD, Centerbar DB, Brekke N (2002) Mental contamination and the debiasing problem. In: Gilovich T, Griffin D, Kahneman D (ed) *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press, Cambridge, pp. 185-200.
- Wilson TD, Gilbert DT (2003) Affective Forecasting. *Advances in Experimental Social Psychology* 35:345–411. doi:10.1016/S0065-2601(03)01006-2.
- Wobker I, Kenning P (2013) Drivers and outcome of destructive envy behavior in an economic game setting. *Schmalenbach Business Review* 65(2):173-194.

Table 1 Selected Evidence on Surrogation by Focal Property of Decision Makers

Categories:	Selected publications highlighting similarities between students and managers	Selected publications highlighting differences between students and managers	Debate resolved?
Attitudes	-	e.g., Alpert (1967), Copeland et al. (1973), Remus (1986)	Yes
Behaviors	Decision processes	e.g., the literature reviewed in Locke (1986)	No
	Decision outcomes	e.g., Dipboye et al. (1975), Hofstedt (1972), Liyanarachchi and Milne (2005), Remus (1986)	
		e.g., Barr and Hitt (1986), Harris and Sutton (1995), Hofstedt (1972), Hughes and Gibson (1991)	
		e.g., Abdel-khalik (1974), Fleming (1969), Khera and Benson (1970), Moskowitz (1971)	

^aCategories suggested, for instance, by Ashton and Kramer (1980) and Beltramini (1983)

Table 2 Selected Categorization Schemes by Purpose of Research

Fundamental logic	Research purpose that permits student subjects	Research purpose that does not permit student subjects	Exemplary key publication
Study design/ preparation vs. actual research	Trial runs	Actual research projects	Miller (1966)
	Exploratory research	All other (i.e., confirmatory) research	Shuptrine (1975)
Studying psychological mechanisms vs. real-world behavior	Conceptual-level research	Descriptive-level research	Farber (1952)
	Universalistic research (theory-centered)	Particularistic research (context-centered)	Kruglanski (1975)
	Theory application research	Effects application research	Calder et al. (1981)
	Internal validity	External validity	Enis et al. (1972)
	Anomaly exploration when anomaly caused by universal bias or institutions	Anomaly exploration when anomaly caused by specific features of observed individuals	Roth (1995)

Table 3 Test Statistics

Hypothesis	Debiasing technique	Test type	Test statistic	p-value (two-tailed)	Comment	
H1	Training in biases	Mann-Whitney U-test	U	16872.5	.038 *	
H2	Reducing time pressure	Nonparametric correlation	Kendall's tau-b	-.077	.025 *	
H3	“Consider the opposite”	Mann-Whitney U-tests	U	19580.5	.847	
H4	Relying on external advice			16646.5	.917	Control group vs. Consultant
				17347.5	.323	Control group vs. Friend
H5	Accountability			17649.0	.378	

* Significant at $p < .05$

Table 4 Comparison of Findings

Hypothesis	Debiasing technique	Hypothesis as formulated by Graf et al. (2012)	Effect found by Graf et al. (2012)	Effect found in student sample
H1	Training in biases	If the bias of competitive irrationality is made salient to the decision maker through specific training, competitive irrationality decreases	Competitive irrationality <i>decreased</i>	Competitive irrationality <i>decreased</i>
H2	Reducing time pressure	Reducing time pressure reduces competitive irrationality	Competitive irrationality <i>decreased</i>	Competitive irrationality <i>decreased</i>
H3	“Consider the opposite”	An explicit appeal to “consider the opposite” decreases competitive irrationality	<i>Null</i>	<i>Null</i>
H4	Relying on external advice	Reliance on external advice reduces competitive irrationality	<i>Null</i>	<i>Null</i>
H5	Accountability	The creation of pre-decision process accountability to a reasonably well-informed, legitimate audience with unknown views and an interest in accuracy decreases competitive irrationality	Competitive irrationality <i>increased</i>	<i>Null</i>

Fig. 1 Surrogation Framework

	Question	Assessment possibilities
1 Existence of bias in both populations	<i>Do both managers and students exhibit the focal bias?</i>	<ul style="list-style-type: none"> • Empirical tests for bias in both populations
2 Identical cause for bias in both populations	<i>Do managers and students exhibit biased behavior due to the same underlying cause?</i>	<ul style="list-style-type: none"> • Theoretical argumentation • Prior research on the process of bias in both populations
3 Identical first-order consequences of debiasing technique in both populations	<i>Does the debiasing technique result in the same first-order consequences when used by/on managers and students?</i>	<ul style="list-style-type: none"> • Theoretical argumentation • Prior research on the debiasing technique • Manipulation checks

Fig. 2 Possible Comparisons when Assessing Surrogation in Debiasing Research

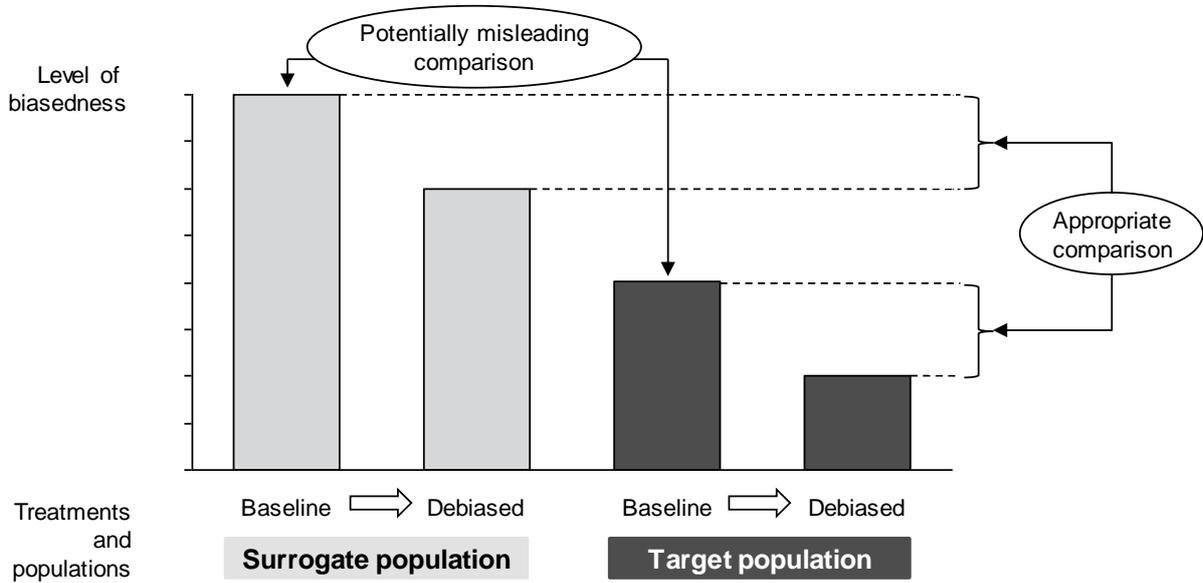


Fig. 3 Control Group Stimulus Material from Graf et al. (2012)

You are the owner and general manager of a manufacturing firm known as **YouCo**.

Recently your company introduced a new product for industrial customers and **you must decide the pricing strategy** for this product.

You are aware that **your main competitor, CompetitorCo**, is producing a **product that is very similar** to the one that your firm has just introduced. It is **as good as yours** and the **market is the same for both products**.

Unfortunately, new legal regulation will ban both products in one year. Therefore, the **products will only be on sale for one year**. The product's technology cannot be transferred into another product and any **experience with the old product will not help in developing a new product**.

After calculating the **total profits** expected for your firm and the competitor until the end of the year, you know that **four pricing strategies** will yield the following results:

	Low price	Medium-low price	Medium-high price	High price
YouCo	€ 25 million	€ 35 million	€ 45 million	€ 55 million
CompetitorCo	€ 10 million	€ 30 million	€ 50 million	€ 70 million

(Followed by four-point Likert-type scale to select desired pricing strategy)

Fig. 4 Experiment Results

