Erasmus University Rotterdam

Erasmus School of Economics

Master Thesis (Behavioural Economics – EMO Track)

Outcome Bias at Performance Evaluation: De-biasing through Formatted Information

Name Student: Peterson, J. A.

Student ID Number: 322738

Supervisor: Li, C.

First Reader: Xu, Y.

Date Final Version: March 7th, 2017

Abstract

The outcome bias is present when the outcome of a decision made by the decision maker extraproportionately influences the evaluator's evaluation regarding the decision maker's quality of decisions. In effort to de-bias the evaluator, formatted information is applied. This treatment does not prove to be significantly reducing the outcome bias. The results also do not provide clear effects from personal characteristics, except for the subject's risk aversion.

Table of Contents

1	Intro	oduction3
2	Lite	rature Review5
	2.1	Potential causes of outcome bias5
	2.2	Evaluator characteristics7
	2.3	De-biasing methods
3	Нур	otheses and Method10
	3.1	Hypotheses10
	3.2	Methods11
4	Resu	ılts13
	4.1	Data13
	4.2	Expectations15
	4.3	Interpretation17
5	Con	clusion and Discussion20
	5.1 Co	nclusion
	5.2	Future research
6	Refe	22
7	Арр	endix25
	Survey	T – Treatment group questionnaire set
	Survey	C – Control group questionnaire set
	Questi	onnaire Answers Sheet

1 Introduction

Employee performance evaluation is usually focused on the outcome of the decisions made by the employees. Easily observed and quantifiable performance such as taxi fares and factory production are rewarded according to performance pay, incentivizing the employees to exert maximum effort at relatively minimal risk. The same does not apply to, say, evaluating the performance of a CEO; where measuring the performance purely based on the company's stock return will unjustly reward or punish the agent (i.e., the CEO). For example, the share price of an oil company is for a significant part based on the world demand for crude oil. Therefore, the price of crude oil is beyond the control of the CEO. However, the decisions made by the CEO (e.g., company reorganizations, defining target areas, etcetera) do have a certain effect on the share price, albeit partly. In this example, evaluating the CEO's performance based on the stock return constitutes an outcome bias. Through recent decades and even the past century it has been the best practice to value a decision by observing the outcome of the decision. However, it is generally overlooked that the outcome of the decision is a mixture of different factors, flowing from several situations during and after the decision-making, but before the moment of evaluation. With this in mind, it is plausible that evaluators falsely reward wrong decisions, punish good intentions or elect lucky candidates through a natural fallacy called the *outcome* bias. Outcome bias is the false assumption in which principals are induced to think that the agents' decision is just as good as the outcome of their decision, thereby neglecting any factors that lie between the point of decision-making and the outcome of the decision itself. The term is best coined by Baron and Hershey (1988) and remain a most interesting subject in the field of (behavioral) economics. The article discusses the presence of the bias through a set of experiments conducted concerning either medical or monetary matters.

Baron and Hershey (1988) find the presence of outcome bias through an experiment in which the subjects are required to rate the decisions of the fictitious decision-makers. The experiment comprises of a series of cases where either the physician or the patient is the decision-maker, followed immediately by the outcome of the decision (success or failure). The experiment has a within-subject design. The undergraduate subjects on average rate the decisionmaker as competent if the outcome is favorable. At gambles they also rate that the thinking is better if the unchosen option turned out poorly.

The outcome bias is prevalent when people evaluate others' decisions. As argued before, it can lead to unjust evaluation. In a world where the performance of employees become less quantifiable and their tasks more diverse, the need of de-biasing methods at performance evaluations become crucial. Therefore, reducing the outcome bias is desirable for a better evaluation and reward system in organizations. This thesis investigates the following question: "Can we de-bias evaluators' decision evaluation through formatted (statistical) information?" Key elements of this research is to test whether providing formatted (statistical) information will alleviate the outcome bias, while also controlling for evaluators' personal characteristics and gender; key elements which have not been examined intensively in previous literature. I will start with a review of literature articles that contribute to the findings in this particular aspect in the field of behavioral economics, covering the background of the definitions I discuss in the thesis. Then follow the chapter where I discuss the hypotheses underlying the research, accompanied by the chosen methodological approach. The results chapter discusses the data and data analysis with regard to the experiment. Lastly, I will conclude with the findings that I acquired through the experiment. I will also discuss the shortcomings of the earlier and recent articles, yet also make a recommendation of potential future research in outcome bias.

2 Literature Review

2.1 Potential causes of outcome bias

While similar to *hindsight bias*, there should be remarked that there is a distinction between hindsight bias and outcome bias. The hindsight bias focuses on the memory distortion of the decision-maker, which can also be his/her own decision evaluator, to favor his/her own decision-making process after the outcome of the decision is known (Bukszar & Connolly, 1988; Hawkins & Hastie, 1990; Roese & Vohs, 2012), while outcome bias focuses on the evaluator weighting the outcome extra-proportionately in comparison with other relevant information. The main reason for outcome biases to be present is for the outcome itself to be presented to the decision evaluator. The sequence of information provided is important, as information which is available only after the decision is made will influence the evaluator's evaluation, while being irrelevant during the decision-making process (Berlin, 1984).

In addition to the hindsight bias, another explanation for the outcome bias to arise is the subjective need for the evaluator to rate the decision as either good or bad in a rather significant level. Somehow it is hard to evaluate a decision-making process if the outcome of the decision is accessible for the evaluator, even if there is no reward for good decisions or punishment for bad decisions (Emerson et al., 2010; Peecher and Piercey, 2008). Agents have been observed to change their decision in a similar sequence after an unfavorable outcome despite the fact that the probability of the unfavorable outcome being high (Lefgren, Platt and Price, 2015). If an agent encounters a similar prospect in the future (e.g., the probabilities and the payoffs are similar), but chooses a different decision, then the agent too is become affected extra-proportionately by the previous outcome and makes a biased decision.

The *hot-hand fallacy* – or the overestimation of autocorrelation in performance – is linked to the outcome bias, where an overreaction following a bad news may cause markets to collapse (Offerman & Sonnemans, 2004). The opposite is also true. Favorable outcomes are perceived to have sprung forth from good decisions (Lipshitz, 1989). The housing bubble in the U.S.A. building up until 2007 is a perfect example where banks and other agents perceive the ever increasing value of real estate stocks as the housing sector delivering good performance.

However, this increase was majorly a bandwagon effect¹ and had little to do with the performance in the particular sector.

Although it is often perceived obvious that good decisions are followed by a favorable outcome, care should be taken when evaluating the decision-making process as there are likely also exogenous factors that influence the outcome. Luck has been known to contribute to good outcomes, yet its influence is usually underestimated (Bertrand and Mullainathan, 2001). A regular complement to luck are cycles. In an economic downturn it is convenient to blame the policy makers, as it is evenly convenient to re-elect the incumbent after an overall good outcome (Wolfers, 2002). Furthermore, one explanation of an outcome to affect a principal's evaluation is the lack of interaction or limited access to decision-making processes between the evaluator as principal and the decision-maker as agent (Ghosh & Lusch, 2000).

Outcome bias during performance evaluation may occur due to agent's exposure to externalities, but there may also be several other explanations: the principal cannot (fully) observe the agent, the task of the agent contains multiple performance dimensions, and innovative agents may have different views on business compared to their principles, etcetera (Marshall & Mowen, 1993). All three explanations need to be considered to be integrated in the evaluation criteria, because lacking these explanations may expose the evaluator to outcome bias. Several studies therefore mention the importance of the *intention* of the agent in complement to the outcome of his/her decision (Krueger and Acevedo, 2007). However, the intention of the agent is commonly neglected and, while the intention might be different, mistaken for the cause of the outcome (Pizarro, Uhlmann & Bloom, 2003; Cushman, 2008). Therefore, the performance of the agent has to be analyzed for the factors that he/she is accountable for; both in the decisionmaking process, as well as in the outcome (Pollmann, Potters & Trautmann, 2014). Before making the decision, the agent should make his intentions clear to the evaluator, as well as his plan of approach. This transparent decision-making covers both the agent from risky decisions and the principal from rewarding the agent after an extra-proportionate outcome due to external factors. The decision-maker's decision therefore should be examined separately instead of jointly in order not to contaminate his/her decision with his/her colleagues. A joint evaluation can exacerbate the outcome bias when the intentions of the decision-maker is known to the evaluator

¹ Social desirability bias, which encourages an individual to choose the socially desired decision.

(Sezer et al, 2016). A de-biasing policy may consider to have multiple evaluators evaluating a decision.

Information provided for the evaluator in order to assess the decision of the decisionmaker needs to be concise and contain the relevant data. A non-concise information will leave the evaluator with a misplaced freedom of interpreting the situation during the decision-maker's decision-making, which may lead to exacerbation of information distortion (Russo, Meloy & Medvec, 1998). This distortion can even be intensified if a decision-maker needs to choose between two or more prospects visible to the evaluator and the evaluator has a clear preference among the prospects (DeKay, Patiño-Echeverri & Fischbeck, 2009). Certainly in many business or management cases where outcome is not easily quantified, the evaluating principal may automatically condemn any decision of the agent if it is not coherent to the evaluating principal's ideals.

2.2 Evaluator characteristics

A long list of articles discussed and analyzed the outcome bias but only a few articles considered the effect of gender and personal characteristics on the evaluators' tendency to be exposed to the outcome bias. Individuals with a higher SAT (Scholastic Aptitude Test) score are on average less prone to the effect of outcome bias when evaluating a decision (Stanovich & West, 1998). The SAT test consisting of among others critical reading and mathematics shows that a more rational mindset – which is commonly accepted to be attributional to higher educated individuals - can override the bias of outcome information (Gino, Moore and Bazerman, 2009), while information distortion is found greater among less educated individuals (Carlson & Russo, 2001). However, the test does not include the evaluator as a stakeholder in the outcome of the decision. Evaluators then tend to reciprocate on basis of the outcome, instead of the intention (Cushman et al, 2009). Given the situation where the evaluator receives a payoff dependent from the outcome of the decision and information being incomplete, reciprocal behavior and moral heuristics then play a larger role in evaluating the decision and the evaluator will still be exposed to the outcome bias (Sunstein, 2005; Murata et al, 2015). Logically, gender would not have a significant determinant effect on the influence of the outcome towards evaluating decisions. However, a few articles show that gender does have a correlation with certain personal characteristics (Deery & Fildes, 1999; Krahé & Fenske 2002; Laapotti & Keskinen, 2004),

especially the risk taking behavior of individuals, which also affect their assessments. Male subjects are on average more risk seeking when required to make a decision, compared to the female subjects. In addition to the risk-taking behavior, male subjects are also more tendentious to overestimate their ability when required to assess their own performance. One detail that needs to be pointed out about these articles is the fact that the surveys are self-assessments of the individuals' risk-taking behavior and perceived savviness of driving skills. This detail inclines the observations towards the hindsight bias as the evaluators are required to assess their own decision-making process. Nevertheless, the authors prove that there is a significant difference between males and females when it comes to decision evaluation.

2.3 De-biasing methods

Several literature discuss the possible methods to de-bias the outcome effect on decision evaluations. One common de-biasing method is to organize frequent feedback sessions between the decision-maker and the evaluator (Luckett & Eggleton, 1991), while also keeping in check the difference between the evaluation of the expected return and the evaluation of the actual return (Frederickson, Peffer & Pratt, 1999). A small difference between the two implies that the decision is considered good as it is close to the expectation. Still, there may be external factors that will affect the outcome, despite the actual decision. Evaluators should, in addition to basing their evaluation on the decision-maker's reports, also rely on expertise consultation from external consultants (Gold, Knechel & Wallage, 2012). This expertise consultation, preferably in collaboration with multiple external consultants in order to sketch a valid expectation, will form the basis of a benchmark to compare the decision-maker's decision with. This benchmark is called a *formatted information*, which in general takes the form of a statistical probability of failure and success. Recent papers suggest that future researches need to be conducted on the format of additional information provided at performance evaluations in order to reduce outcome bias (König-Kersting, Pollmann, Potters & Trautmann, 2016; Gurdal, Miller & Rustichini, 2013; Ratner & Herbst, 2005). This formatted information will function as a treatment.





Illustration 1 summarizes and schematically describes the influences that play a role in decision evaluation. As seen on the illustration there are mainly three factors that influence the decision evaluation: the decision outcome, the general information and the evaluator's own characteristics. General information may seem broad, but it generally comprises of information which is visible to the evaluator. Note that everything under the dotted line, which presents the factors visible to the evaluator, is unobserved by the evaluator. As such, information is strictly visible to the evaluator. Exogenous factors are assumed to be beyond the reach of the evaluator, meaning that these – along with the noises – are unobserved by the evaluator. Again, outcome bias arises when the exogenous factors influences the outcome significantly and induce the evaluator to rely extra proportionately on the outcome. Personal characteristics are somewhere between the observed and the unobserved, as individuals merely perceive their traits.

3 Hypotheses and Method

3.1 Hypotheses

I want to research whether formatted information can help evaluators get rid of outcome bias. Formatted information as a tool should align the evaluation of the subjects towards a more unbiased evaluation, on average, meaning:

1) H₀: Additional information does not reduce outcome bias.

H_a: Additional information reduces outcome bias.

Evaluation is simplified into a multiple level of either 'good' or 'bad'. The outcome of the decision is simplified into only two possibilities, which either takes the value positive if the outcome is perceived to be positive and negative if the outcome is perceived to be negative. The null hypothesis implies that additional formatted information does not have any effect on debiasing the evaluator, meaning that the evaluator is not more likely to evaluate unbiasedly. The alternative hypothesis (e.g., the formatted information works properly as a de-biasing tool) means that the subject, on average, will be more likely to evaluate the decision as 'good' given the additional information to suggest so, then the subject would if given the negative outcome of the decision as the only explaining variable.

As a sub part of the research, I am also interested to see whether gender and empathy play a role in evaluating the decisions and achievements of others. This implies that:

2) H₀: *Male and female subjects are equally outcome biased.*

H_a: Male and female subjects are not equally outcome biased.

Under the null hypothesis, gender should not play a role when evaluating a decision, given all other variables constant. This also implies that, if present, outcome bias should not have more effect on a specific gender. This null hypothesis indirectly goes against the listed articles from Deery & Fildes, Krahé & Fenske, and Laapotti & Keskinen, in which are found that gender has a significant effect on decision-making. However, decision-making is not necessarily the same as evaluating.

3.2 Methods

In order to de-bias performance measurement, a basic reference point is required for rating the components of the decision-making process. Even if there may be no one-size-fits-all method of de-biasing technique, ex-post data from previous and similar projects may give an objective prediction. In this way, probabilities of success given as formatted (statistical) information can help evaluators rationalize their evaluation when given a set of decisions and outcomes. However, the econometric technicalities involved in order to design the relevant formatted (statistical) information will not be covered in this thesis; subjects are required to assume that the given probabilities are reliable. This uncovered aspect of econometric modelling of the formatted (statistical) information should not be relevant to the results of the experiments, as this thesis is interested in the evaluator's decision evaluation through information processing – and not the technical design of the probabilities. I will design the experiment much like the one by Baron and Hershey (1988), but with a few minor differences: subjects are required to evaluate (fictional) business cases on the decision of the (fictional) characters. The questionnaire comprises descriptions of two business situations in which the fictional characters take the role of consultants and are required to make decisions on basis of some given information about the business climate. Taking the role of an evaluator, the subjects then evaluate the agents' decisions by rating the decisions with a value from a 7-points Likert scale, ranging from 'Worst decision ever' to 'Best decision ever'. This is the first round, comprised of 2 questions, in which the cases only describe the (1) situation and (2) decision of the consultant.

The second round will have the subjects evaluate the decisions of the consultants, much like the first round, but now the outcome of the decisions made by the consultants are disclosed. This implies that the 2 questions in this round contain (1) situation, (2) decision and (3) outcome of the decision made by the consultants. I expect to find a level of statistically significant correlation between the evaluation of the subject and the outcome of the decision; subject will be likely to state '*Good Decision*' if the outcome of the decision is positive and vice versa. If this is indeed the case, then there is a presence of outcome bias - as described by Baron and Hershey (1988).

Building forth on the research conducted by König-Kersting, Pollmann, Potters and Trautmann (2016), the subject in the treatment group will evaluate the decisions of the agents,

much like the subject in the control group. The treatment group, however, will have additional formatted (statistical) information regarding the outcome of the decision disclosed in both the first and the second round. The formatted information contains the probability of success of the decision, resembling an expertise consultation². The first round contains no outcome of the decision and any information in order to evaluate the decision has to be found in the situation description. I believe that outcomes, despite Hershey & Baron's (1992) recommendation to include the outcomes, will still affect the evaluator's judgment.

The 'Alice' case can be described as an agent that makes a suboptimal decision (according to the expected value) yet generates a positive outcome. As we are interested in the treatment group, we should first calculate the expected utilities from both 'Alice' and 'Charles' cases. At the 'Alice' case, opting for candidate A will yield an expected value (profit) of \$440k, while opting for candidate B will yield an expected value (profit) of \$450k. A risk-neutral agent will then opt for candidate B. Knowing these probabilities, I expect a risk-neutral evaluator to rate Alice's decision as rather bad. The 'Charles' case can be described as an agent who, despite the negative outcome, makes the optimal decision according to the expected value. Charles has chosen a prospect that yields the highest expected revenue, albeit a relatively risky one. His decision can seem quite risky, as the difference between success and failure of developing an application is large (\$1,700,000) compared to focusing on a new target group (\$560,000). I expect that on average subjects will rate Charles' decision positively. Intuitively, I also expect subjects who consider themselves risk-averse to not favor Charles current decision.

The data is obtained through an experiment where the subjects were required to fill in a questionnaire. There were 2 ways of approaching the subjects: either face to face or online. The online survey is done via Qualtrics³. Randomization is key of the experiment and is achieved through a randomizer command when filling in online and randomized by a coin flip when filling in face to face. Each subject has completed, or is currently following, a management study. Those that have completed a management study have no more than 2 years' work experience.

² Gold, Knechel and Wallage (2012), yet here I 'summarize' the expertise consultation as the probabilities given in the formatted information.

³ <u>https://qtrial2016q3az1.qualtrics.com/SE/?SID=SV_5vxOhhEeEO2XGbr</u>

Although I have considered using a monetary incentive as a tool to attract subjects for the experiment, I decided not to opt for the use of monetary incentive. The first and foremost reason not to opt for the monetary incentive is the costs. Secondly, using money to attract subjects will undoubtedly shorten the time to gather the subjects, but judging by cases in the questionnaire the monetary incentive will unlikely result in more valid responses. The questionnaire does not contain right or wrong responses, but merely observes the behavior of the subjects. The incentive then will not reward the subject to for filling in the proper responses, but merely shorten the time to reach the goal of 60 participants. Lastly, the major part of the subjects fills in the questionnaire online. This would have required the subjects to fill in their bank account and the experimenter to transfer multiple small sums of money – which is a tedious work.

4 Results

4.1 Data

In accordance to the objective of the thesis, I want to examine whether the statistic information at least reduces the outcome bias and whether the personal characteristic plays a role in evaluating the decision.

Treat	A dummy variable which indicates 0 if the subject belongs in the control group
	and 1 if the subject belongs in the treatment group.
Ev1	The subject's evaluation of Alice's decision without the outcome.
Ev2	The subject's evaluation of Charles' decision without the outcome.
Ev3	The subject's evaluation of Alice's decision with the outcome.
Ev4	The subject's evaluation of Charles' decision with the outcome.
Team	The subject's sense of teamwork. Higher value means higher sense of teamwork.
Rewd	The subject's sense of positive reciprocal behavior. Higher value means higher likelihood to reward.
RisAv	The subject's self-reported level of risk aversion. Higher value means higher risk aversion.
Retl	The subject's sense of negative reciprocal behavior. Higher value means higher

	likelihood to retaliate.
Dconf	The subject's perception of the importance of educational degrees towards landing a good job
Koonf	The subject's perception of the importance of knowledge in general towards
Ktolii	landing a good job.
Pconf	The subject's perception of the importance of his/her personality towards landing a good job.
Fem	A dummy variable which indicates 0 if the subject is male and 1 if the subject is female.
Age	The subject's stated age.
OBAlice	<i>The value generated by</i> Ev3-Ev1 <i>. Will be used to identify</i> <u><i>outcome</i></u> <u><i>b</i></u> <i>ias.</i>
OBCharles	<i>The value generated by</i> Ev4-Ev2 <i>. Will be used to identify</i> <u>outcome</u> <u>b</u> <i>ias.</i>

Table 2 - Descriptive Statistics

Variable	Variable Obs		Std. Dev.	Min	Max
Treat	60	0.500	0.504	0	1
Ev1	60	3.950	1.281	1	6
Ev2	60	4.667	1.145	2	6
Ev3	60	4.600	1.317	1	7
Ev4	60	3.900	1.175	2	6
Team	60	3.617	1.091	1	5
Rewd	60	4.283	0.715	2	5
RisAv	60	3.500	1.017	2	5
Retl	60	3.183	1.033	1	5
Dconf	60	3.667	0.951	2	5
Kconf	60	4.050	0.790	3	5
Pconf	60	3.767	.647	2	5
Fem	60	0.500	.504	0	1
Age	60	25.217	2.877	20	37

Table 1 shows the description of the variables. The variables *Team* (teamwork), *Rewd* (reward behavior; positive reciprocal behavior), *RisAv* (level of risk aversion) and *Retl* (punish

behavior; negative reciprocal behavior) function to sketch the subject's attitude towards working in a team. The variables *Dconf* (confidence level on career success on basis of attained educational degrees), *Kconf* (confidence level on career success on basis of own general knowledge) and *Pconf* (confidence level on career success on basis of perceived personality) function to sketch the subject's confidence of own traits. Both the variables functioning to sketch the subject's attitude towards working in a team and the variables functioning to sketch the subject's confidence of own traits resemble represent a confirmation of previous findings from earlier research on the effects of personal characteristics on decision-making and evaluations. A summary of variables is shown in table 2. The median age of all subjects is 25 years, which is very close to the mean age in the population. There is an equal number of male and female subjects taking part in the experiment. In addition, there is an equal number of subjects taking part in the experiment in the control group and in the treatment group.

4.2 Expectations

If the de-biasing mechanism through providing formatted information works, I expect the treatment to result in the treatment group to rate Alice's decision lower and to rate Charles' decision higher, compared to the control group. Furthermore, I expect gender to have some influence in subjects rating the agents' decisions. As discussed in previous literature regarding decision-making in risky situations, male subjects tend to overestimate their capabilities and as such this bias will most likely influence their evaluation. If personal characteristics affect the evaluator's judgment when evaluating decisions, then it is presumable that gender will indirectly contribute to a difference in outcome bias levels between male and female subjects.

The questionnaire for both the control and treatment groups consists of 4 case assessments and a short personality questionnaire. The case assessments are 2 sessions of 2 questions. The cases were actually 2 fictitious business cases, where in the first session the 2 cases contain basic information, while in the second session the cases contain basic information and decision outcome. The difference between the control and the treatment group is that the treatment group – for both the first and the second session – contains the decisions' probability of success. The case assessments required to be rated with a 7-point Likert scale, whereas the short personality questions required to be rated with a 5-point Likert scale.

First to test whether there is an outcome bias at the evaluations, comparisons between EV1 and EV3, and EV2 and EV4 were performed in two Wilcoxon signed-rank tests (see Appendix A for detailed test statistics). Outcome bias exists in all cases (all p-values<0.05). Moreover, the finding of positive values at the signed rank test on the Alice cases show that the positive outcome of her decision influences the evaluators positively when evaluating her decision; similarly, for the evaluations of the Charles cases: the negative outcome of his decision influences the evaluating his decision; even when the statistical information provided in the treatment group shows that Charles' decision yields a higher expected value.

4.3 Interpretation

As there is a significant presence of outcome bias⁴, I now run a test to see whether the treatment can decrease the outcome bias. Table 3 shows the output of the regression analysis with the treatment as an independent variable in order to explain the dependent variable of OBAlice and OBCharles for respectively the Alice and the Charles case. Although the results show that the treatment yields a negative sign at the Alice case and a positive sign at the Charles case, the lack of significance implies that the treatment's effect is insignificant. In this sample of subjects, the treatment has not shown a significant effect on the subject's evaluation on the cases. Therefore, the treatment has to be marked as an insufficient tool to eradicate, or at least decrease, the outcome bias.

	Alice	Charles
	b/se	b/se
Treat	-0.233	0.200
	(0.27)	(0.28)
Constant	0.767***	-0.867***
	(0.19)	(0.20)
R-sqr	0.012	0.009
dfres	58	58
BIC	183.1	186.1
* p<0.05,	**p<0.01,	***p<0.001

Table 3 - The treatment effect: Alice and Charles cases by statistical information, Regression Analysis.

⁴ See the appendix tables A1 and A2 for the Wilcoxon Signed rank-test.

	Alice C	Alice T	Charles C	Charles T
Team	-0.227	-0.072	0.009	-0.254
	(0.24)	(0.29)	(0.29)	(0.21)
Rewd	0.170	0.295	0.278	-0.369
	(0.45)	(0.34)	(0.53)	(0.24)
RisAv	-0.109	0.011	0.098	0.498*
	(0.23)	(0.31)	(0.27)	(0.22)
Retl	0.119	-0.176	-0.108	0.032
	(0.19)	(0.32)	(0.22)	(0.23)
Dconf	-0.212	0.141	-0.272	0.468
	(0.25)	(0.36)	(0.29)	(0.26)
Kconf	0.467	-0.490	-0.588	-0.068
	(0.35)	(0.42)	(0.42)	(0.30)
Pconf	-0.339	0.045	0.558	-0.014
	(0.35)	(0.48)	(0.42)	(0.34)
Fem	0.432	0.192	-0.674	-0.002
	(0.45)	(0.53)	(0.54)	(0.38)
Age	0.049	-0.033	0.024	-0.081
	(0.09)	(0.09)	(0.10)	(0.06)
constant	-0.420	2.077	-1.024	0.737
	(3.01)	(3.48)	(3.59)	(2.48)
R-sqr	0.147	0.169	0.237	0.359
dfres	20	20	20	20
BIC	111.6	120.3	122.1	99.9

Table 4 – The Personality effect: Alice and Charles cases by the personality, Regression Analysis.

. . .

* p<0.05, **p<0.01, ***p<0.001

Finally, I run a similar linear regression test to see whether the personal characteristics of the subjects play a role in evaluating the decisions. Table 4 shows that the personal characteristics have no significant influence in evaluating the decisions, both for the control and the treatment group, except for the independent variable risk aversion. It is found in this sample that, on average in the treatment group, an increase of 1 point in the variable risk aversion increases the evaluation of Charles' decision by 0.498 points. This effect is statistically significant at the 95% confidence level. However, whether this effect has any strong relevance may remain speculative. The finding would namely imply that, when given the related statistical

information around the Charles case, a relatively risk-averse individual on average will rate Charles' decision as 'good decision'. Risk-aversion implies that the individual will be more likely to choose a sure outcome than a risky one with the same expected value. The statistical information provided for the treatment group shows that Charles took the relatively riskier decision: the variance of the outcomes in his decision to recommend the development of a smart application is larger than the variance of the outcomes in the alternative decision. Logically, a risk averse individual with no outcome bias will prefer the alternative decision.

5 Conclusion and Discussion 5.1 Conclusion

Based on the findings in the Results chapter, I conclude that the de-biasing mechanism provided in this experiment does not significantly reduce the outcome bias. The insignificance of the effect of the de-biasing treatment simply does not provide sufficient support to the objective of implementing the formatted information. I also conclude that, based on the findings of this experiment, personal traits/characteristics do not play a significant role in subjects rating the agents' decisions – except for the level of risk aversion. This is also true for gender, as the results show no significant difference in performance evaluation between male and female subjects.

Although the current results of this experiment show that the de-biasing mechanism by providing an additional formatted information does not decrease the outcome bias significantly, it does not mean that the hypotheses and expectations stated in this thesis is pointing in the wrong direction. Firstly, there has been no interaction with the agents in the experiment as the cases are fictitious. As recommended by Luckett & Eggleton (1991), this interaction with the agent may reveal more about the nature of the decision and provide more useful information for the evaluator. Bertrand and Mullainathan (2001) also find that adding a large shareholder in the board will decrease the pay for external factors immensely. The large shareholder then acts as an evaluator with him/her in the board implying that there will most likely be an interaction with the CEO (i.e., the agent). Secondly, the size of the experiment population obviously plays a role in the results. As discussed in the Results chapter, the regression analysis of the outcomes on the treatment shows coefficients in the desired sign. A larger set of subjects may yield a more significant result. Lastly, the questionnaire regarding the subjects (perceived) personal characteristics may be more elaborate in order to retrieve a more accurate view on the subject's actual characteristics.

5.2 Future research

Future work should take the abovementioned discussion points into account. Generally, a research budget may help solve most of the mentioned challenges. Monetary incentives have been known to have the subjects to exert desired level of effort. This is an important requirement due to the lengthy questionnaire, as discussed in the concluding sub-chapter, regarding the personal characteristics of the subjects. The monetary incentive is then justified as there is more

effort required from the subjects to state their answers. The extended version of the questionnaire should contain several questions that will not only provide the subject's perception of his/her own characteristics, such as the questions provided in this experiment, but also check for any overconfidence or misperception of one's decision-making abilities. Gold, Knechel and Wallage (2012) stress the importance of expert consultation in complement of a performance evaluation, and this can be expressed in the form of an advice instead of a mere statistical information. In congruence with König-Kersting, Pollmann, Potters, and Trautmann's (2016) article, this formatted information may contain expert consultation which also takes into account the correction of the outcome with regard to exogenous factors. For example, such consult may well include the decision outcomes of peers (e.g., competitors) of the same industry and preferably those making a similar decision.

6 Reference

- 1. Bailey, C. D., Brown, L. D., & Cocco, A. F. (1998). The effects of monetary incentives on worker learning and performance in an assembly task. *Journal of Management Accounting Research*.
- 2. Baker, G., Gibbons, R., & Murphy, K. J. (1994). Subjective performance measures in optimal incentive contract. *The Quarterly Journal of Economics*, *109*(4), 1125.
- 3. Banker, R. D., Potter, G., & Srinivasan, D. (2000). An empirical investigation of an incentive plan that includes nonfinancial performance measures. *The accounting review*, 75(1), 65-92.
- 4. Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of personality and social psychology*, 54(4), 569.
- Baron, J., Beattie, J., & Hershey, J. C. (1988). Heuristics and biases in diagnostic reasoning: II. Congruence, information, and certainty.*Organizational Behavior and Human Decision Processes*, 42(1), 88-110.
- 6. Berlin, E. (1984). Excess capacity, plant abandonments, and prudent management. *Public Util. Fortn.;(United States)*, *114*(11).
- 7. Bertrand, M., & Mullainathan, S. (2001). Are CEOs rewarded for luck? The ones without principals are. *Quarterly Journal of Economics*, 901-932.
- 8. Bol, J. C. (2011). The determinants and performance effects of managers' performance evaluation biases. *The Accounting Review*, *86*(5), 1549-1575.
- 9. Bukszar, E., & Connolly, T. (1988). Hindsight bias and strategic choice: Some problems in learning from experience. *Academy of Management Journal*, *31*(3), 628-641.
- 10. Carlson KA, & Russo JE. (2001). Biased interpretation of evidence by mock jurors. *Journal Of Experimental Psychology. Applied*, 7(2), 91-103.
- 11. Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*(2), 353-380.
- 12. Cushman, F., Dreber, A., Wang, Y., & Costa, J. (2009). Accidental outcomes guide punishment in a "trembling hand" game. *PloS one*, *4*(8), e6699.
- 13. Deery HA, & Fildes BN. (1999). Young novice driver subtypes: relationship to high-risk behavior, traffic accident record, and simulator driving performance. *Human Factors*, *41*(4), 628-43.
- 14. DeKay, M. L., Patiño-Echeverri, D., & Fischbeck, P. S. (2009). Distortion of probability and outcome information in risky decisions. *Organizational Behavior And Human Decision Processes*, 109(1), 79.
- 15. Emerson GB, Warme WJ, Wolf FM, Heckman JD, Brand RA, & Leopold SS. (2010). Testing for the presence of positive-outcome bias in peer review: a randomized controlled trial. *Archives Of Internal Medicine*, 170(21), 1934-9. doi:10.1001/archinternmed.2010.406.

- 16. Frederickson, J. R., Peffer, S. A., & Pratt, J. (1999). Performance evaluation judgments: Effects of prior experience under different performance evaluation schemes and feedback frequencies. *Journal Of Accounting Research*, *37*(1), 151-165.
- 17. Gino, F., Moore, D. A., & Bazerman, M. H. (2009). No harm, no foul: The outcome bias in ethical judgments. *Harvard Business School NOM Working Paper*, (08-080).
- 18. Gold, A., Knechel, W. R., & Wallage, P. (2012). The effect of the strictness of consultation requirements on fraud consultation. *The Accounting Review*,87(3), 925-949.
- 19. Ghosh, D., & Lusch, R. F. (2000). Outcome effect, controllability and performance evaluation of managers: some field evidence from multi-outlet businesses. *Accounting, Organizations and Society*, 25(4), 411-425.
- 20. Gurdal, M. Y., Miller, J. B., & Rustichini, A. (2013). Why Blame? *The Journal Of Political Economy*, *121*(6), 1205.
- 21. Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, *107*(3), 311.
- 22. Hecht, G., Tafkov, I., & Towry, K. L. (2012). Performance Spillover in a Multitask Environment*. *Contemporary Accounting Research*, 29(2), 563-589.
- 23. Heller, Y. (2014). Overconfidence and diversification. *American Economic Journal: Microeconomics*, 6(1), 134-153.
- 24. Hershey, J. C., & Baron, J. (1992). Judgment by Outcomes: When Is It Justified? *Organizational Behavior And Human Decision Processes*, 53(1), 89.
- 25. Holt, Charles A., and Susan K. Laury (2005), "Risk Aversion and Incentive Effects: New Data without Order Effects," *American Economic Review*, *95*, 902-904.
- 26. König-Kersting, C., Pollmann, M., Potters, J., & Trautmann, S. T. (2016). Good decision vs. good results: Outcome bias in the evaluation of financial agents.
- 27. Krahé, B., & Fenske, I. (2002). Predicting aggressive driving behavior: The role of macho personality, age, and power of car. *Aggressive Behavior*, 28(1), 21-29. doi:10.1002/ab.90003
- Krueger JI, & Acevedo M. (2007). Perceptions of self and other in the prisoner's dilemma: outcome bias and evidential reasoning. *The American Journal Of Psychology*, 120(4), 593-618.
- 29. Laapotti, S., & Keskinen, E. (2004). Has the difference in accident patterns between male and female drivers changed between 1984 and 2000? *Accident Analysis And Prevention*, *36*(4), 577-584. doi:10.1016/S0001-4575(03)00064-2
- Lefgren, L., Platt, B., & Price, J. (2015). Sticking with What (Barely) Worked: A Test of Outcome Bias. *Management Science*, 61(5), 1121-1136. doi:10.1287/mnsc.2014.1966
- 31. Lipshitz, R. (1989). "Either a Medal or a Corporal": The Effects of Success and Failure on the Evaluation of Decision Making and Decision Makers. *Organizational Behavior And Human Decision Processes*, 44(3), 380.
- 32. Luckett, P. F., & Eggleton, I. R. (1991). Feedback and management accounting: a review of research into behavioural consequences. *Accounting, Organizations and Society*, *16*(4), 371-394.

- 33. Marshall, G. W., & Mowen, J. C. (1993). An experimental investigation of the outcome bias in salesperson performance evaluations. *Journal of Personal Selling & Sales Management*, 13(3), 31-47.
- Murata, A., Nakamura, T., Matsushita, Y., & Moriwaka, M. (2015). Outcome Bias in Decision Making on Punishment or Reward. *Procedia Manufacturing*, *3*, 3911-3916. doi:10.1016/j.promfg.2015.07.914
- 35. Offerman, T., & Sonnemans, J. (2004). What's Causing Overreaction? An Experimental Investigation of Recency and the Hot-hand Effect. *The Scandinavian Journal Of Economics*, 106(3), 533-554. doi:10.1111/j.0347-0520.2004.t01-1-00376.x
- 36. Peecher, M. E., & Piercey, M. D. (2008). Judging Audit Quality in Light of Adverse Outcomes: Evidence of Outcome Bias and Reverse Outcome Bias. *Contemporary Accounting Research*, 25(1), 7.
- Pizarro, D. A., Uhlmann, E., & Bloom, P. (2003). Causal deviance and the attribution of moral responsibility. *Journal Of Experimental Social Psychology*, 39(6), 653-660. doi:10.1016/S0022-1031(03)00041-6
- 38. Pollmann, M. M., Potters, J., & Trautmann, S. T. (2014). Risk taking by agents: The role of ex-ante and ex-post accountability. *Economics Letters*, *123*(3), 387.
- Ratner, R. K., & Herbst, K. C. (2005). When good decisions have bad outcomes: The impact of affect on switching behavior. *Organizational Behavior and Human Decision Processes*, 96(1), 23-37.
- 40. Roese NJ, & Vohs KD. (2012). Hindsight Bias. Perspectives On Psychological Science : A Journal Of The Association For Psychological Science, 7(5), 411-26. doi:10.1177/1745691612454303
- 41. Russo, J. E., Meloy, M. G., & Medvec, V. H. (1998). Predecisional Distortion of Product Information. *Journal Of Marketing Research*, *35*(4), 438-452.
- Sezer, O., Zhang, T., Gino, F., & Bazerman, M. H. (2016). Overcoming the outcome bias: Making intentions matter. *Organizational Behavior And Human Decision Processes*, 137, 13-26. doi:10.1016/j.obhdp.2016.07.001
- 43. Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal Of Experimental Psychology: General*,127(2), 161-188. doi:10.1037//0096-3445.127.2.161
- 44. Sunstein CR. (2005). Moral heuristics. *The Behavioral And Brain Sciences*, 28(4), 531-42; Discussion 542-73.
- 45. Wakker, Peter, and Daniel Deneffe (1996), "Eliciting von Neumann-Morgenstern Utilities when Probabilities are Distorted or Unknown," *Management Science*, 42, 1131-1150.
- 46. Wolfers, J. (2002). *Are voters rational?: Evidence from gubernatorial elections*. Graduate School of Business, Stanford University.

7 Appendix

Table A1 – Outcome bias at the Alice evaluation: Control and Treatment group, Wilcoxon Signed-Rank test.

sign	obs		sum ranks		expected	
	Control	Treatment	Control	Treatment	Control	Treatment
positive	14	15	329	315.5	164.5	205
negative	0	5	0	94.5	164.5	205
zero	16	10	136	55	136	55
all	30	30	465	465	465	465
			Control	Treatment		
unadjusted variance			2363.75	2363.75	-	
adjustment for						
ties			-9.63	-49.88		
adjustment for zeros			-374.00	-96.25		
adjusted variance			1980.13	2217.63		
Ho: Ev4 = Ev2						
z =			3.697	2.346		
Prob > z						
=			0.0002	0.0190		

Table A2 - Outcome bias at the Charles evaluation: Control and Treatment group, Wilcoxon Signed-Rank test.

sign	obs		sum	n ranks	expected	
	Control	Treatment	Control	Treatment	Control	Treatment
positive	2	1	35	20.5	187	180
negative	15	15	339	339.5	187	180
zero	13	14	91	105	91	105
all	30	30	465	465	465	465
			Control	Treatment		
unadjusted variance			2363.75	2363.75		
adjustment for						
ties			-15.00	-36.00		
adjustment for zeros			-204.75	-253.75		
adjusted variance			2144.00	2074.00		
Ho: Ev4 = Ev2						
z =			-3.283	-3.502		
Prob > z						
=			0.0010	0.0005		

Survey T – Treatment group questionnaire set.

The consulting Firm

Thank you in advance for taking part in the survey!

The questionnaires will have the topic of several fictional business cases; of which some are altered version based on actual business cases. You play the role of a manager of a management consulting firm, providing service to a client as described in the questionnaire.

As the role of the manager is demanding organizationally, you are not expected to take part in the operation itself. Instead, you are required to provide an evaluation of the consultant's decision. You must do so based on the information that is provided in the case.

This survey is expected to take less than 15 minutes, but practically most of the subjects of the experiment finish in 10 minutes.

Jason Peterson

First Round

Case 1

Alice is a hiring manager. She has tasked the HR to source and assess potential candidate for an analyst position in the Firm. This specific new-recruit will be immediately assigned to a client in the 'chemicals and natural resources' industry. After a thorough sourcing analysis and a recruitment campaign of three months, Alice manages to find two potential candidates: candidate A and candidate B.

According to analysis, candidate A's profile has 60% probability of generating \$800,000 profit and 40% probability of generating \$100,000 loss. While candidate B's profile has 75% probability of generating \$600,000 profit and 25% probability of generating nothing.

Alice decides to hire candidate A.

How do you evaluate Alice's decision?

Case 2

Charles is a senior manager. He is currently assigned to a client in the insurances industry. The client wishes to increase the sales of life-insurances, but seems to find difficulty in a market where the demand for these products are diminishing. Charles analyzes that, for most people, insurances are too difficult to be thoroughly understood. The target group is also selective; while the working class are usually automatically assigned to an insurance program, the youth nearly ever bothers about the importance of life-insurances.

According to analysis, developing a smart application has 55% probability of generating \$1,500,000 <u>profit</u> and 45% probability of generating \$200,000 <u>loss</u>. While focusing on a new target group has 70% probability of generating \$800,000 <u>profit</u> and 30% probability of generating nothing.

Charles decides to recommend the smart application.

How do you evaluate Charles' decision?

Second Round

Case 3

Alice is a hiring manager. She has tasked the HR to source and assess potential candidate for an analyst position in the Firm. This specific new-recruit will be immediately assigned to a client in the 'chemicals and natural resources' industry. After a thorough sourcing analysis and a recruitment campaign of three months, Alice manages to find two potential candidates: candidate A and candidate B.

According to analysis, candidate A's profile has 60% probability of generating \$800,000 profit and 40% probability of generating \$100,000 loss. While candidate B's profile has 75% probability of generating \$600,000 profit and 25% probability of generating nothing.

Alice decides to hire candidate A.

It turns out that candidate A has managed to generate \$800,000. The client is very satisfied.

How do you evaluate Alice's decision?

Case 4

Charles is a senior manager. He is currently assigned to a client in the insurances industry. The client wishes to increase the sales of life-insurances, but seems to find difficulty in a market where the demand for these products are diminishing. Charles analyzes that, for most people, insurances are too difficult to be thoroughly understood. The target group is also selective; while the working class are usually automatically assigned to an insurance program, the youth nearly ever bothers about the importance of life-insurances.

According to analysis, developing a smart application has 55% probability of generating \$1,500,000 <u>profit</u> and 45% probability of generating \$200,000 <u>loss</u>. While focusing on a new target group has 70% probability of generating \$800,000 <u>profit</u> and 30% probability of generating nothing.

Charles decides to recommend the smart application.

It turns out that developing the application leads to a loss of \$200,000. The client is very dissatisfied.

How do you evaluate Charles' decision?

Survey C – Control group questionnaire set.

The consulting Firm

Thank you in advance for taking part in the survey!

The questionnaires will have the topic of several fictional business cases; of which some are altered version based on actual business cases. You play the role of a manager of a management consulting firm, providing service to a client as described in the questionnaire.

As the role of the manager is demanding organizationally, you are not expected to take part in the operation itself. Instead, you are required to provide an evaluation of the consultant's decision. You must do so based on the information that is provided in the case.

This survey is expected to take less than 15 minutes, but practically most of the subjects of the experiment finish in 10 minutes.

Jason Peterson

First Round

Case 1

Alice is a hiring manager. She has tasked the HR to source and assess potential candidate for an analyst position in the Firm. This specific new-recruit will be immediately assigned to a client in the 'chemicals and natural resources' industry. After a thorough sourcing analysis and a recruitment campaign of three months, Alice manages to find two potential candidates: candidate A and candidate B.

According to estimations, candidate A's profile can either generate \$800,000 profit or \$100,000 loss. While candidate B's profile can either generate \$600,000 profit or nothing (e.g., profit of \$0).

Alice decides to hire candidate A.

How do you evaluate Alice's decision?

Case 2

Charles is a senior manager. He is currently assigned to a client in the insurances industry. The client wishes to increase the sales of life-insurances, but seems to find difficulty in a market where the demand for these products are diminishing. Charles analyzes that, for most people, insurances are too difficult to be thoroughly understood. The target group is also selective; while the working class are usually automatically assigned to an insurance program, the youth nearly ever bothers about the importance of life-insurances.

According to analysis, developing a smart application can either generate \$1,500,000 profit or \$200,000 loss. While focusing on a new target group can either generate \$800,000 profit or nothing (e.g., profit of \$0).

Charles decides to recommend the smart application.

How do you evaluate Charles' decision?

Second Round

Case 3

Alice is a hiring manager. She has tasked the HR to source and assess potential candidate for an analyst position in the Firm. This specific new-recruit will be immediately assigned to a client in the 'chemicals and natural resources' industry. After a thorough sourcing analysis and a recruitment campaign of three months, Alice manages to find two potential candidates: candidate A and candidate B.

According to estimations, candidate A's profile can either generate \$800,000 profit or \$100,000 loss. While candidate B's profile can either generate \$600,000 profit or nothing (e.g., profit of \$0).

Alice decides to hire candidate A.

It turns out that candidate A has managed to generate \$800,000. The client is very satisfied.

How do you evaluate Alice's decision?

Case 4

Charles is a senior manager. He is currently assigned to a client in the insurances industry. The client wishes to increase the sales of life-insurances, but seems to find difficulty in a market where the demand for these products are diminishing. Charles analyzes that, for most people, insurances are too difficult to be thoroughly understood. The target group is also selective; while the working class are usually automatically assigned to an insurance program, the youth nearly ever bothers about the importance of life-insurances.

According to analysis, developing a smart application can either generate \$1,500,000 profit or \$200,000 loss. While focusing on a new target group can either generate \$800,000 profit or nothing (e.g., profit of \$0).

Charles decides to recommend the smart application.

It turns out that developing the application leads to a loss of \$200,000. The client is very dissatisfied.

How do you evaluate Charles' decision?

Questionnaire Answers Sheet

Case Evaluation

Case 1

Worst ever!	Very Bad	Quite Bad	Neutral	Quite Good	Very Good	Genius!

Case 2

Worst ever!	Very Bad	Quite Bad	Neutral	Quite Good	Very Good	Genius!

Case 3

Worst ever!	Very Bad	Quite Bad	Neutral	Quite Good	Very Good	Genius!

Case 4

Worst ever!	Very Bad	Quite Bad	Neutral	Quite Good	Very Good	Genius!

Short questionnaire

Please state your opinion about the following statements.

	Strongly	Somewhat	Neutral	Somewhat	Strongly
	Disagree	Disagree		Agree	Agree
I find the 'group' more					
important than the 'individual'					
I always reward helpful people					
I only speak out when I am					
certain of my case					
I always retaliate to saboteurs					
I am sure my degrees will help					
me in finding good jobs					
I am sure my knowledge will					
help me in finding good jobs					
I am sure people find me a					
pleasant person					

Gender Age

Female / Male