Optimism bias and cost overruns: experimenting on the internal and external views in resources and time estimation

Giuseppe Sassano *Bsc, MSc*

Supervised by:

Dr Antoine Vernet

Prof Grant Mills

Bartlett School of Sustainable Construction

University College London

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

02/2022

London, United Kingdom

Declaration

‘I, Giuseppe Sassano, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.'

Signature:

Word count (excluding tables and figures): 61,188

Date: 18/02/2022

Acknowledgments

I would like to express my deepest gratitude to all the people that, directly or indirectly, supported and followed me during the journey of this research and for the whole duration of my PhD.

However, I would have never completed this journey without the help and support of my supervisor, Dr. Antoine Vernet, that was always there when I had questions with prompt responses, valuable suggestions and a positive attitude that motivated me in doing the best I could throughout this process.

My gratitude also goes to the professional service staffs from the Bartlett School of Construction and Project Management, especially Satinder Kundhi, Anita Treso and Helen Pascoe, your support during these years has always gone the extra mile.

Moreover, I would like to thank all the people that, by dedicating their time to complete the experiments, allowed me to build this research from scratch.

To my PhD colleagues, I will miss sharing time with you inside and outside the office, you always made me feel part of a community and inspired me every day with your great minds.

To my friends and partner, I could have not been luckier to have you by my side during these years, you thought me that fragility and strength not only can coexist, but they can also elevate you.

Finally, I would like to thank my family, because with their everyday support, especially during the though times of the pandemic and confinement, donated me the wings to explore the World and strong roots to always come back to.

Abstract

Optimism bias affects most estimation based human decisions, from daily activities to the appraisal of big infrastructure projects. Building upon the underlying constructs of this behaviour through the lenses of support and prospect theories, operationalised in the internal and external view in the project management context, helped me to formulate a conceptual framework. This framework, called the “Holistic view” suggests ways to integrate the two perspectives with the aim to improve the quality of forecasts in infrastructure management. Integration can be achieved by including subjective probabilities and unpacking techniques into case-based reasoning methods. Based on this framework and on the analysis of the relevant policies, perspectives, definitions and techniques provided by the literature on optimism bias in project forecasting, I designed and administered four experiments, with a sample of 231 participants. The results of the first experiment show that there is a positive relationship between different levels of dispositional optimism and resource overruns. In the second experiment, I found that adding an optimism uplift to an estimate structurally increases forecast precision, however, this might lead to the use of more resources than when optimism uplift is not applied. In the third experiment, I found that unpacking, whilst making estimations slightly more precise, does not influence the estimations as in the previous case, but only the estimator. The results of the fourth experiment, combining unpacking with optimism uplift, indicated that forecast precision starkly increased, supporting the adoption of an “Holistic view”. This research shows the relevance of experimental methods in project management, unveils new relationships between different perspectives in the project forecasting field, and analyses, net of other factors, the impact of forecasting tools integrated in some of the policies addressing the effect of optimism bias in estimations, suggesting ways to improve those and the overall forecasting process of infrastructure projects.

Keywords: Optimism bias, Behavioural Science, Cost overruns, Holistic view, Project Forecasting

Impact statement

This research explores how different perspectives in project forecasting (the internal and external views) can be integrated to improve forecasts' precision in the front-end phase of projects and their overall performance.

The study looks at support and prospect theories - from behavioural economics - and how these theories can be integrated to create a new perspective on forecasting, introduced as a novel conceptual framework: the Holistic View.

The Holistic View is tested using the experimental method, scarcely used in project management research, opening to the importance of using this method to establish causal connections between phenomena that are currently left over in the field. In this study, experimental methods led to a better understanding of the relationship between different levels of dispositional optimism and estimation precision. Also, they helped in exploring the impact on forecast precision of adding an optimism uplift to estimates and unpacking a task/project in subcategories before performing the estimation activity.

The experiments, indeed, show the importance to consider the behavioural aspect (i.e. inaccuracy in estimations derives from biases not errors) when it comes to the estimation task, reporting that, whenever the level of optimism was artificially increased through the Best Possible Self manipulation both the estimations and the result was negatively impacted in terms of precision and success rate.

Furthermore, the experiments underline both strengths and weaknesses of using the tools of optimism uplift and unpacking during the estimation process, highlighting the two different natures of those: the first one having a structural impact on the estimate and the second one being a descriptive tool aiding the estimator.

Those characteristics, when put together in the last experiment, resulted in a stark improvement of forecast precision as well as an increase in the task success rate, opposed to a situation when no such tools are used.

The study, supporting the adoption of an Holistic View in forecasting, has several contributions: it expands academics' horizons beyond the two perspectives currently offered in the literature (often deemed to be antithetic); it highlights the benefits of experimental methods to advance our understanding of project management, it offers tools for policymakers to devise project estimation guidelines; finally, it suggests ways practitioners can develop and apply best practices tailored to their project needs.

Table of contents

Acknowledgments **3**

Abstract **4**

Impact statement **6**

List of figures **12**

List of tables **14**

List of Abbreviations **15**

CHAPTER 1. BACKGROUND PROBLEM DISCUSSION, AIM, OBJECTIVES AND RESEARCH QUESTION **16**

CHAPTER 2. THE "HOLISTIC VIEW" **23**

2.1. Introduction 23

2.2. Ecological rationality, heuristics and cognitive biases: the architecture of mind 26

2.2.1. Perfect rationality, Bounded rationality and Ecological rationality 27

2.2.2. Heuristics and Cognitive biases 29

2.3. The anatomy of cost underestimation: prospect theory and the planning fallacy 34

2.3.1. Expected Utility Theory and Prospect Theory: a normative versus descriptive model 35

2.3.2. The planning fallacy 40

2.4. The adoption of an outside view on project cost forecasting 44

2.5. Enhancing the positive impact of the outside view with attributes of the inside view: Support Theory and subjective probabilities 47

2.6. Integrating different perspectives on cost forecasting: the Holistic view 52

2.7. Theoretical implications 54

CHAPTER 3. COST AND SCHEDULE OVERRUNS: DEFINITIONS, PERSPECTIVES, TECHNIQUES AND POLICIES **58**

3.1. Schedule and cost overruns: different literature perspectives 58

3.1.1. The "Hiding Hand" perspective 59

3.1.2. The Evolutionist perspective 63

3.1.3 The Behavioural perspective 67

3.2. Application of the theories to the infrastructure construction industry: insurgence of cost overruns 73

3.2.1. From cost underestimation to cost overrun 74

3.2.2. Defining and measuring cost overrun 76

3.2.3. Magnitude and Frequency of cost overruns 80
3.2.4. Policy landscape 85
3.3. Mitigating cost overruns through the outside view on forecasting with Case Based Reasoning Methods: A focus on Reference Class Forecasting techniques 94
3.3.1. Mott MacDonald method 95
3.3.2. The Flyvbjerg method.. 98
3.3.3. The Salling Method... 102
3.3.3.1. Cost Benefit Analysis and Quantitative Risk Analysis 102
3.3.3.2. The CBA-DK Method 104
3.3.4. Further Considerations 107
3.4. Remarks and research gaps... 110
CHAPTER 4. RESEARCH METHOD 115
4.1. Research Strategy 115
4.2. Philosophical stance 116
4.2.1. Ontological and Epistemological assumptions 117
4.2.2. Selected Research Philosophy... 120
4.3. Why experiments? 124
4.4. Experiments Design... 130
4.4.1 The experimental platform and experimental task with instructions 130
4.4.2. LOT-R test... 135
4.4.3. Experiment 1 136
4.4.4. Experiment 2 137
4.4.5. Experiments 3 139
4.4.6. Experiment 4 140
4.4.7. Power Analyses of the experiments 140
4.4.7.1 Experiment 1 143
4.4.7.2 Experiment 2 and 3 143
4.4.4.3. Experiment 4 145
4.4.7.4. Conclusion 146
4.5. Further considerations on the experiments..................................... 146
4.6. Ethics in data collection 148
4.7. Research Hypotheses and Propositions.. 149
4.7.1. Literature review... 150
4.7.2. Experiments' Hypotheses.. 151

4.7.2.1. Experiment 1 151
4.7.2.3. Experiment 3 154
4.7.2.4. Experiment 4 154
4.7.4. Previous qualitative research 155
CHAPTER 5. EXPERIMENT 1 157
5.1. Introduction 157
5.2. Manipulation 157
5.3. Manipulation check 160
5.4. Analytical approach 162
5.5. Results.. 163
5.5.1. Manipulation check 163
5.5.2. Descriptive statistics 164
5.5.3. Hypothesis and sub hypothesis testing – preliminary analysis .. 166
5.5.4. Hypothesis and sub hypothesis testing – regression analyses.. 171
5.5.5. Hypothesis and sub hypothesis testing – robustness analysis .. 174
5.5.6. Theoretical implications 179
CHAPTER 6. EXPERIMENT 2 AND 3 RESULTS 182
6.1 Introduction 182
6.2. Experiment 2 183
6.2.1. Manipulation 183
6.2.2. Analytical approach 185
6.3. Results.. 186
6.3.1. Descriptive statistics 186
6.3.2. Hypothesis and sub hypothesis testing – preliminary analysis. . 190
6.3.3. Hypothesis and sub hypothesis testing – regression analyses.. 195
6.3.4. Hypothesis and sub hypothesis testing – robustness analysis .. 198
6.4. Experiment 3 203
6.4.1 Manipulation 203
6.4.1 Analytical approach 204
6.5 Results... 205
6.5.1 Descriptive statistics 205
6.5.2 Hypothesis and sub hypothesis testing – preliminary analysis. .. 208
6.5.3 Hypothesis and sub hypothesis testing – regression analyses... 212
6.5.4 Hypothesis and sub hypothesis testing – robustness analysis ... 215
6.6. Theoretical implications 220

Appendix A: Optimism uplift calculation 225
CHAPTER 7. EXPERIMENT 4 226
7.1. Introduction 226
7.2. Manipulation 227
7.3. Analytical approach 228
7.4. Results 229
7.4.1. Descriptive statistics 229
7.4.2 Hypothesis and sub hypothesis testing – preliminary analysis 233
7.4.3 Hypothesis and sub hypothesis testing – regression analyses 236
7.4.4 Hypothesis and sub hypothesis testing – robustness analysis 239
7.5. Theoretical implications 243
CHAPTER 8. DISCUSSION 250
8.1. Introduction 250
8.2. Results' discussion 251
8.2.1. Experiment 1 251
8.2.2. Experiment 2 253
8.2.3. Experiment 3 255
8.2.4. Experiment 4 256
8.3. Adjusted power analysis 258
8.3.1. Experiment 1 259
8.3.2. Experiment 2 260
8.3.2. Experiment 3 261
8.3.4. Experiment 4 262
8.3.5. Further considerations 263
8.4. Limitations 264
8.5. Future research 268
CHAPTER 9. CONCLUSIONS AND POLICY IMPLICATIONS 275
REFERENCES 279

List of figures

Figure 1 - Role of illusion in environment rationality (Gigerenzer and Selten, 2002)........ 30
Figure 2 - Value function: overweight losses and underweight gains (adapted from Baddeley, 2013)........ 39
Figure 3 - Integration of the External and Internal view to create a Holistic View on forecasting (by author) 53
Figure 4 - Causal chain leading to overruns according to Evolutionists (by author) 70
Figure 5 - Causal chain leading to overruns according to Behavioural perspective supporters (by author) 70
Figure 6 - Cumulative probability function of investment cost 90
Figure 7 - Mott MacDonald method (by author) 96
Figure 8 - Mott MacDonald and Flyvbjerg methods' calcutation procedure (Mott Macdonald, 2002) 98
Figure 9 - Flyvbjerg method 2004 (by author) 101
Figure 10 - CBA-DK method (by autor) 105
Figure 11 - Salling method calculation procedure (adapted from Salling and Banister, 2009)........ 107
Figure 12 - Experimental task’s structure with measures (by author) 132
Figure 13 - Screenshot of task instructions from the online experimental tool developed by the author 132
Figure 14 - Screenshot of back-end menu and results’ visualisation 134
Figure 15 - Experimental Flow 148
Figure 16 - Research Hypotheses and Propositions........ 149
Figure 17 - Manipulation administered per level of dispositional optimism 164
Figure 18 - Histogram difference between estimated and actual bricks by treatment........ 168
Figure 19 - Difference between estimated and actual brick by treatment .. 169
Figure 20 - Difference between estimated and “right number” of bricks by treatment........ 170
Figure 21 - Integration of the External and Internal view to create a Holistic View on forecasting (by author) 182
Figure 22 - Manipulation administered per actual number of bricks........ 190
Figure 23 - Difference between estimated and actual bricks by treatment. 193
Figure 24 - Difference between estimated and accrual time by treatment . 194
Figure 25 - Difference between estimated and “right” number of bricks by treatment........ 195
Figure 26 - Manipulation administered per actual number of bricks........ 208
Figure 27 - Difference between estimated and actual number of bricks by treatment........ 210
Figure 28 - Difference between estimated and actual time by treatment ... 211
Figure 29 - Difference between estimated and “right number of bricks by treatment........ 212

Figure 30 - Integration of the External and Internal view to create a Holistic View on forecasting (by author) ... 226
Figure 31 - Manipulation administered per actual number of bricks ... 232
Figure 32 - Difference between estimated and actual number of bricks by treatment ... 234
Figure 33 - Difference between estimated and actual time by treatment ... 235
Figure 34 - Difference between estimated and "right" number of bricks by treatment ... 236
Figure 35 - Integration of the External and Internal view to create a Holistic View on forecasting (by author) ... 250

List of tables

Table 1 - Possible reasons for cost overruns (by author) 43
Table 2 - Optimism bias uplifts (from HM Treasury 2013) 93
Table 3 - Project risk areas of Mott MacDonald method (From Mott Macdonald, 2002) 97
Table 4 - Categorisation of projects, sources of optimism bias uplifts and optimism uplift per % percentile (Adapted from Flyvbjerg and COWI, 2004) 100
Table 5 - Main characteristics of the methods discussed (by author) 108
Table 6 - Benefits and limitations of the proposed framework 112
Table 7 - Assumptions of positivism and interpretivism (from Collis and Hussey, 2013) 119
Table 8 - Items composing LOT-R(Adapted from Scheier et al. 1994) 135
Table 9a and 9b- Power experiment analysis 1 143
Table 10a and 10b - Power analysis experiment 2 and 3 144
Table 11a and 11b - Power analysis experiment 3 145
Table 12 - Items composing LOT-R test (Scheier et al., 1994) 161
Table 13 - LOT-R scoring system (Scheier et al., 1994) 162
Table 14 - Experiment 1 descriptive statistics 166
Table 15 - Regression analysis and C.I. experiment 1 173
Table 16 - Mann-Whitney tests 175
Table 17 - Regression analyses and C.I. with Age as confounding variable 177
Table 18 - Collinearity 179
Table 19 - Descriptive statistics experiment 2 187
Table 20 - Regression analyses and C.I. experiment 2 197
Table 21 - Mann-Whitney tests 198
Table 22 - Regression models and C.I. with Age as a confounding variable 200
Table 23 - Collinearity 202
Table 24 - Descriptive statistics experiment 3 205
Table 25 - Regression models and C.I. Experiment 3 215
Table 26 - Mann-Whitney tests 215
Table 27 - Regression models and C.I. with Age as confounding variable 218
Table 28 - Collinearity 219
Table 29 - Descriptive statistics experiment 4 230
Table 30 - Regression models and C.I. experiment 4 239
Table 31 - Mann-Whitney tests 240
Table 32 - Regression models and C.I. with Age as a confounding variable 242
Table 33 - Collinearity 243
Table 34a and 34b - Power analysis experiment 1 259
Table 35a and 35b - Power analysis experiment 2 261
Table 36a and 36b - Power analysis experiment 3 262
Table 37a and 37b - Power analysis experiment 4 263

List of Abbreviations

BCR: Benefit-Cost Ratio

BPS: Best Possible Self

CBA: Cost-Benefit Analysis

CBR: Case-based reasoning

E1: Control condition Experiment 2

E2: Control condition Experiment 3

E3: Control condition Experiment 4

EUT: Expected Utility Theory

EVM: Earned Value ManagementIRR: Internal Rate of Return

LOT-R: Life Orientation Test – Revised

NPV: Net Present Value

OU: Optimism Uplift

OUP: Optimism uplift + Unpacking

QRA: Quantitative Risk Analysis

RCF: Reference Class forecasting

TD: Typical day

U: Unpacking

VIF: Variance Inflation Factor

CHAPTER 1. BACKGROUND PROBLEM DISCUSSION, AIM, OBJECTIVES AND RESEARCH QUESTION

Forecasting the cost and timescale of a project is an essential part of assessing its viability in the decision to build phase. Following what Flyvbjerg (2008) and the supplementary guidance for optimism bias of the Green Book (2013) suggest, there is a systematic propensity for project appraisers to be exceedingly optimistic in their forecasts. Usually, this behaviour, results in cost overruns for infrastructure projects for an average of 33% more than the initial predicted cost and in severe delays. Previous research showed, furthermore, that cost and schedule overruns affect a high number of infrastructure projects, reaching in some studies a striking percentage of 90% of the projects analysed (Flyvbjerg, 2008; Mott MacDonald, 2002).

Infrastructure projects represent the engine of economic growth, it is forecasted that between now and 2040 an average of $3.7 Trillion per year will be invested in these types of projects in the 50 countries with the highest GDP globally (Mills et al., 2011; Global Infrastructure Outlook, 2017). In the UK only, between 2020 and 2021 the Government Major Project Portfolio (GMPP) comprises 66 projects of Infrastructure and Construction with a Whole Life cost of £236 billion and an initial investment of over £100 billion (Annual Report on Major Projects 2020 to 2021; HM Treasury, 2016). Furthermore, as Hutton (2019) reports, by studying a sample of only 10 major public infrastructure projects ongoing in the country, the overruns of those projects reached the sum of £17.2 billion and delays equating to 32.7 years. In other words, those projects are costing £624 per household, showing that overruns are not only a problem for governments but also for taxpayers.

The phenomenon of cost and schedule overrun, for these reasons, has been extensively addressed in literature, whose primary concern was to understand the reasons behind it, its magnitude in the industry and, subsequently, to find some methods to mitigate it (Flyvbjerg et al. 2005). There are political and psychological reasons behind the propensity towards optimism; for the latter, we can trace its roots to the so called "planning fallacy" theorised, for the first time, by Kahneman and Tversky (1979) in the context of the behavioural economics model exemplified by prospect theory (Baddeley, 2013). The fallacy, highlights the fact that people, when asked to make a prediction about the resources needed to complete an activity, are likely to underestimate them and overestimate the potential benefits arising from the activity. The same thing can be said when people are asked to make a prediction about the time it will take them in order to complete a certain task.

Academics categorized this phenomenon naming it "internal view" (Buelher et al., 1994) and through extensive research and experimentations were able to establish that it represents a direct consequence of the planning fallacy, as decision makers, focusing on the uniqueness of the project at hand, tend to be overconfident towards the expected deliverables of the project in terms of schedule and resources used.

Several studies aimed to find a model able to account for the planning fallacy by considering a different perspective became known as the "outside view". A number of models taking the "outside view" as a theoretical perspective were developed. The first one was proposed by Mot Macdonald in 2002, which elaborated a method that accounts for a database of past projects, comprising risk mitigating factors (MotMacdonald 2002; Green book 2002). On the basis of this model, Flyvbjerg, suggested Reference Class Forecasting, part of the Case-Based Reasoning methods (Flyvbjerg, 2008; Ji et al. 2011). Finally, Salling (2009) proposed a model embedding deterministic, probabilistic and stochastic calculations; regrettably, this method has never been used in practice, if not in the context of the study itself.

The methods mentioned are able to provide the appraiser with comparative information extracted from a database and therefore have the benefit to reduce the deviation coming from cognitive bias. However, they do not consider the “internal view” (even if there is an attempt to do so in the Salling method) which, ideally, should be considered during the appraisal process, as Tetlock (2005) also suggested. Indeed, the “outside view”, considering past similar projects in order to build up cost/schedule estimations, may lead to underestimate the importance of the unique peculiarities the project at hand has and in doing so, it may create discrepancies between the estimated business case and the actual one (Love et al., 2012).

This research wants to give a closer look at the psychological process underlying decisions related to cost, schedule and resource forecasting in infrastructure projects. In so doing, the first objective is the one to establish and further understand the role of optimism bias during the forecasting process in order to detect it and minimise it through the implementation of a model that embeds features of the external and internal view.

Moreover, this study wants to explore how to establish a meaningful link between scholars and practitioners bridging theoretical models with suggestions on practical forecasting models, and how to blend different approaches to develop innovative techniques to collect and analyse data from different sources. The ultimate aim, therefore, is to investigate how to make budgeting process a more effective practice. Concurrently, by investigating the latest solutions implemented on the matter (e.g. Reference Class Forecasting), the research will be targeted at understanding how effective these approaches are in controlling cost and schedule overruns and whether the implementation has resulted in any unintended consequences as suggested by Siemiatycki (2009) and Jennings (2012).

Overall, the research question can be formalized as: "*To what extent the creation of a holistic model embedding the inside and outside views in forecasting can improve current policies and practices aimed at mitigating optimism bias in infrastructure projects?".*

Because of the Covid-19 pandemic, experiments that were to be run in a behavioural laboratory in the University of Florence had to be adapted for online delivery. Experiments were scheduled to start from mid-March 2020, however, the new regulations on social distancing and national lockdown prevented me to run them. The experiments had to be redesigned and adapted to be administered online. This included significant additional work, including the development of original software. The new environment presented many challenges, such as the one to include a construction task that subjects in the laboratory experiments would have done with real building blocks. For this reason, I had to develop an original online platform with the capability to run a

3D building game so that the experiments could be run in a setting reproducing the laboratory setting with fidelity. Creating the experimental platform allowed me to review and further analyse my experimental design: this process helped me in understanding that, in order to have a more complete view of how every manipulation administered in the context of my analysis worked, I needed to design another experiment. For this reason, the experiments I will present in this dissertation are not only the three experiments I initially planned to run in the laboratory but I added a fourth. The pandemic and its obstacles, therefore, gave me the opportunity to re-evaluate my initial design and to improve it further, allowing me, at the same time to create an experimental platform that is fully customisable for future studies.

To answer the research question, the dissertation will be structured in 8 Chapters after this one:

Chapter 2 sets out the basis for the theoretical discussion upon which the whole work is based, by introducing the architecture of mind human beings have according to behavioural economists. This will allow to start discussing and defining cognitive biases and some phenomena that are connected with those, such as the planning fallacy. Linking the planning fallacy with the concept of optimism bias will help setting out the foundations for the theoretical discussion looking at how to mitigate this bias, considering two different perspectives each linked to a theory the external view, based on prospect theory and the internal view based on the support theory. Finally, I will introduce the proposed conceptual framework for this research, named the "Holistic view".

Chapter 3 connects the theoretical discussion of the first part of the literature review with the context of this research: infrastructure project management. Indeed, this chapter discusses different perspectives, approaches, policies, techniques and definitions present in the literature when considering the issue of cost and schedule overruns and how optimism bias affects these, outlining some research gaps as well.

Chapter 4 outlines the research strategy I adopted in order to carry out this study, discusses the philosophical stance of the research and the research method I adopt: experiments. I analyse the design of the four experiments I carried out, with preliminary assumptions on sample size calculations, to conclude with the set of hypotheses and propositions I developed for this study.

Chapter 5 introduces the results of the first experiment, analysing what the impact of higher levels of dispositional optimism is in terms of forecast precision, so to justify and validate the behavioural perspective I adopted in the context of this study.

Chapter 6 looks at the comparative analysis of the results of experiments 2 and 3 each looking at the effect on forecast precision of one specific tool: one coming from the internal view and another one coming from the external one.

Chapter 7 introduces the result of experiment 4, which aims to understand the effect of using concurrently both tools explored in the previous two experiments and observe the effect this has on forecast precision, in line with what I express through the formalisation of the research question.

Chapter 8 discusses the links between the results presented in the previous three chapters, investigating the implications of those, both at a theoretical and practical level. I also provide adjusted power analyses for each experiment considering the means and standard deviations resulting from my experiments. Thereafter, I discuss some of the limitations of this research and future avenues of research based on the findings of this dissertation.

Chapter 9, finally, provides the conclusions of this dissertation, by answering the research question and providing recommendations for future research, practitioners, and governments on the usefulness of implementing a Holistic view during the appraisal process of a project.

CHAPTER 2. THE “HOLISTIC VIEW”

2.1. Introduction

Albert O. Hirshman ([1967], 2015), presented a principle known as the “Hiding Hand”, suggesting that in planning ignorance is a positive thing because, if decision makers would know the real costs and risks associated with a project, very few would decide to accept them and initiate the works. For this reason, a literature review that formalises the theoretical discussion to build a conceptual framework able to be operationalised and mitigate the possible errors arising from cognitive biases may seem something of little relevance for people that endorse the idea of the “Hiding Hand”.

However, as Flyvbjerg and Sunstein (2016) point out, ignorance does not benefit project success, rather it undermines it. In a scenario where real costs of a project are minimised, benefits coming from the project at hand will be overvalued, and therefore, the combined influences of the two phenomena will lead to a compounded negative effect, as a consequence of the initially high degree of uncertainty. Hence, the assumption that ignorance is bad for projects is made in the context of this research.

During this literature review, I will provide more details about why I chose to frame my investigation following the behavioural perspective, but first, in order to begin this analysis, I will introduce the concepts related to the decision-making process according to behavioural economics. Following that, I will analyse the idea of cognitive bias so to understand its potential impact on decision-making.

After having assessed the importance of biases, I will present a concept known as the “planning fallacy” (Flyvbjerg et al. 2010; Kahneman and Tversky, 1979) to explain the phenomenon of cost overruns that affect many (if not most) projects, by linking it to the concept of optimism bias. The planning fallacy draws from the idea that when something needs to be decided, an outside view on the issue should be adopted, thereby, comparing similar past endeavours and relative observable patterns to reach a more accurate estimate. However, as Tetlock (2005) reminds us, all estimations should balance an internal and external view in order to be a useful tool for the decision-making process: with this in mind, another theory, known as support theory (Tversky and Koheler, 1994), will be introduced to understand how the use of subjective probabilities and unpacking can help in the decision-making process during the initial phases of a project.

Functionally describing the problem is not the only objective of this work; indeed, I will investigate how to cross-fertilise the above-mentioned theories, belonging to the same theoretical background of behavioural economics, but putting emphasis on two different sides of the problem, the inside and the outside view on forecasting. Interestingly, in the academic literature (especially in the project management literature) these two sides are perceived to be incompatible (Flyvbjerg, 2018; Love and Dagbui, 2018) and the intense ongoing debate seems more oriented toward discrediting each other’s idea rather than finding a more effective solution to increase the accuracy of estimates and forecasting techniques.

With this in mind, the first part of this literature review (Chapter 2) will be devoted at proposing a new conceptual framework, to be used as a ground for further future analyses and improvement of current forecasting methods that have been put in place by organisations in the construction industry (Mott MacDonald, 2002), academia (Flyvbjerg, 2004(a); Salling and Banister, 2009), policy makers and multilateral organisations (Green Book, 2013) to mitigate the impact of optimism bias.

The second part of the literature review (Chapter 3), supported by the established conceptual framework, focuses on describing different perspectives, approaches, policies, techniques and definitions present in the literature when considering the issue of cost and schedule overruns. In order to create a direct link between the concepts captured in the theoretical framework and their application to the infrastructure construction industry I discuss the "Hiding Hand", evolutionist and behavioural perspectives. Following, I elaborate on the definitions considered in the context of this research of cost underestimation and cost overruns. These definitions shape the calculation procedure relative to the magnitude and frequency of cost overruns and the way those number are reported in different studies. Furthermore, I examine the most relevant studies on the topic in order to appreciate the importance of the subject in the academic literature but also to clarify the relevance of the theme for practitioners both at a project management and policy level.

It is with this background that I present the main policies addressing the issue of cost and schedule overruns and the most relevant techniques adopted to mitigate the insurgence of the phenomenon starting from the appraisal stage of the project. Those techniques, mainly deriving from the adoption of case-based reasoning (CBR) methods have been operationalised in different methods, some of which have been used in real life project and some of them only to evaluate case studies of already delivered projects (Flyvbjerg, 2008; Ji et al., 2011; Salling and Banister, 2009).

Finally, I provide remarks to summarise and further understand the main findings of this literature review on the topic, highlighting also a number of research gaps so to prepare the ground for a combined analysis with the results of the experiments.

2.2. Ecological rationality, heuristics and cognitive biases: the architecture of mind

In order to begin with the analysis of the problem related to cost underestimation in projects, the first step that needs to be made relates to the understanding of the problem under the perspective given by behavioural economics on cognitive biases. In so doing, I will analyse how the concept of rationality evolved over time, in order to frame the general context under which decisions are analysed in this research. After that, I will establish a link between the defined rationality and the concept of heuristics. Then, I show how the application of heuristics often leads to cognitive biases, diverting the judgement of human beings towards outcomes that are not optimal. Those three characteristics will allow in turn to define and build the concept of the architecture of mind I adopt in order to analyse the decision-making process in this study.

2.2.1. Perfect rationality, Bounded rationality and Ecological rationality

According to standard economic models, when an individual needs to make a decision, they rely on mathematical processes and calculations, generally defined as “perfect rationality” (Varian, 2010). Perfect rationality assumes that an economic actor, when prompted to make a decision, will have a preference based on the different sets of available alternative choices, allowing them to clearly state the option they prefer. Preferences, are directly linked with the concept of utility, first defined by Bentham (1823) as a property of any object of a decision that is able to produce an order of preferences based on different available options. In other words, whenever an individual states a preference, this will have a higher utility than the other available choices (Kapteyn, 1985). This model assumes the decision maker to have perfect (i.e. complete) information about the attributes, properties, constructs and outcomes of the decision to be made, hence the definition of it as “perfect rationality”. With this definition, people’s rationality is assumed not to have limits or constraints from a mathematical capacity and contextual points of view. This definition of rationality, over time, proved itself to be hardly applicable in reality and for this reason, many economists, among whom Keynes and Savage, started to study the limits of perfect rationality and perfect information during the decision-making process.

Herbert Simon (1955;1979), was the first to introduce the concept of bounded rationality, that, providing an alternative to the theory on perfect rationality of standard economics just described, has been able to clarify the reasons why the decision-making process of an economic actor may be influenced by boundaries on information, cognitive constraints (such as the partial inability of people to probabilistically assess events) and the complexity of situations, as it happens in the case of modern construction projects (Simon, 1955; Pryke and Smith, 2012).

Bounded rationality is a procedural rather than substantive type of rationality because does not rely on mathematical processes to make a decision and does not lead to optimal outcomes (Baddeley, 2013). In this sense, decision makers do not take an optimal decision aimed at utility maximisation (as predicted by the "perfect rationality" model), but a satisfactory one, which, as a consequence means that they do not possess perfect information on the decision to be made and they will try to act in the most reasonable way given the existing limits. Further to that, Simon (1979), specifies that the way in which economic actors will take decisions will be deeply affected by the above-mentioned limits, becoming a distinctive trait of the final outcome in taking the decision. Examples of those constraints may be represented by the limit to assess and access every single possible option or the emotional involvement there could be in deciding in a given situation.

According to Gigerenzer and Selten (2002), bounded rationality aims to explain the underlying reasons behind certain behaviours and the consequent decisions resulting from them. Bounded rationality, according to them, is focused towards the internal order of beliefs and preferences of the economic actor, failing to explore the structure of the outside environments in which people find themselves in. In response to this, the concept of ecological rationality has been suggested as a way to explain why and when bounded rationality works. Ecological rationality can be defined as a type of rationality that looks at the way in which economic actors think and act as a result of an adaptive process between their cognitive dimension and the ecological structures they find themselves in. Ecological structures refer not only to the structures present in the real world where people live but also to the human tasks and the relationships between the different actors. In other words, the environment in which decisions are taken is not only represented by the physical context but also by the social one.

Considering these assumptions behind ecological rationality, makes it easier to understand why individuals are not able, most of the times, to take optimal decisions, as is, instead, predicted by standard economics.

In this sense, the model of perfect rationality is not deemed to be the most accurate frame when it comes about correctly explaining the limits of decision-making process, since those limits are not considered by those set of theories. To frame better this discussion, therefore, the next step is to understand better what is the process underlying the formalization of human decisions. Most decisions we take are based on quick judgements that are the result of a process of intuition and reasoning: the next section will look at this process, highlighting also its points of failure, the so-called behavioural biases.

2.2.2. Heuristics and Cognitive biases

According to behavioural economics theories, our mind contains an “adaptive toolbox” that, thanks to the adoption of heuristics processes, a quick and instinctive decision-making technique that people use in situations of uncertainty, considers a relatively small amount of information when an individual has to choose on something and/or taking a decision (Gigerenzer and Selten, 2002; Baddeley, 2013). To clarify this concept, Gigerenzer and Selten (2002), illustrate two different scenarios as in fig. 1:

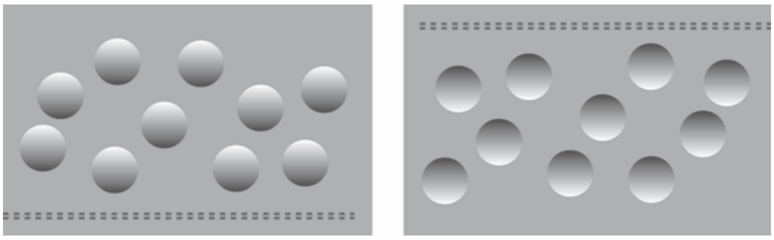


Figure 1 - Role of illusion in environment rationality (Gigerenzer and Selten, 2002)

In the left panel of figure 1, our cognitive system deduces that the spheres are concaves and are directed in an opposite manner in relation to the viewer whereas in the right panel spheres seem to be convex and oriented towards the observer. However, by turning the figure of 180 degrees, it can be noticed that the two figures are identical. This represents a useful analogy when compared to a decision-making context: perspectives are influenced by different environments or ecological structures, and a decision may be largely influenced by the context in which it is made.

The ecological structure in which a decision is taken is not the only element which may lead to a misguided interpretation and assessment of reality. In personal judgements, as a matter of fact, data that are taken into consideration are subject to limited validity (Tversky and Kahneman, 1974) given by the constraints in having access to full information that are subsequently processed following the principles of intuition and reasoning of heuristics.

In light of this, heuristics, can be defined as the set of systematic rules helping to take a decision where there is no complete information and therefore some level of uncertainty. This process might present some informational gaps, defined in the literature as cognitive biases. Many academics have studied the relationship between heuristics and biases, among which, the largest and more comprehensive studies have been made by Kahneman and Tversky (2000), identifying various types of heuristics with related biases.

On the other hand, Gigerenzer developed a comprehensive body of experimental work as well on heuristics, where the emphasis was to show how heuristics represents a “fast and frugal” decision-making tool that eventually leads people to take the right decision even if they are not perfectly informed. This perspective, in contrast with the one on biases resulting in fallacious decisions described by Kahneman, seems not to account for the fact that, in certain occasions, as the one I investigate in this dissertation of optimism bias, the resulting decision is negatively influenced by heuristics, leading people oftentimes not to take the right decision (Gigerener and Todd, 2007; Wójtowicz and Winkowski, 2018).

To have a more specific idea about what heuristics are and how they interact with the human information processing system, following, I will outline three heuristics principles employed to assess and elaborate judgment in different situations.

Representativeness heuristics, judges the probability that a specific individual or event falls in the general category it is supposed to fit in according to a subjective assignation process. In other words, it uses as a basis for the estimation the stereotype pertaining to the category the decision maker believes it belongs to (Kahneman and Tversky, 1972). In this sense, when judging the probability that someone is a doctor rather than an architect, in a situation of limited information availability on the person's personality, for example, the decision maker will be influenced mostly by the idea he or she has on how a doctor should look like rather than focusing on the attributes of the person subject to the judgement (Skitmore et al.,1989).

Availability is another type of heuristics; this process aids decision making in judging the likelihood of an event by connecting it with easily retrievable similar past events, based on the broader ecological context the decision makers is in (Baddeley, 2013). In other words, if someone is asked to estimate the risk linked with a death by natural disaster, the resulting estimate will most likely be highly dependent on personal experience and/or recalling of similar past cases from the person's social context. In a similar way, when a project appraiser is asked to conduct a risk assessment to estimate possible failure points of the project, the result of the analysis may be focused on events that are easier for the estimator to be recalled (Tversky and Kahneman, 1974).

The availability heuristics may be exposed to the emergence of some biases, such as the retrievability and familiarity biases, as exemplified in the cases discussed above of the estimation of death by a natural disaster and failure points estimates by the appraiser of a project. At the same time, it could drive someone's decision without considering other important factors that might have had a big impact on the decision itself if taken into consideration before. During this literature review, I will analyse existing techniques that may help in mitigating possible biases arising from availability heuristics, in order to understand how to apply them in the context chosen for this research.

Other heuristics principles often used when evaluating a circumstance, are anchoring and adjustment. In various occasions, as a matter of fact, people are asked to create an estimate that will be based, in most cases, either on a given initial value (called the anchor) or represented by the outcome of a partial computation. From the anchor, when more information about the situation is known, it usually happens a process of revising and adjustment, which most of times results to be non-sufficient to yield a realistic and precise final estimate. Along the same line of thought, Lichenstein et al. (1978), through a series of five experiments relating to the judged estimation of lethal events, were able to show that different starting points (or prospects) usually held to different results in terms of the estimate, which are usually biased towards the initial value.

In this sense, they were able to show that if people are prompted with an information under the form of an initial value, this may drive their estimates up or down, as the estimates will be highly dependent on the initial prospect given. It is straightforward to understand, from the considerations above, how estimates might be biased, as in most of the cases they are based on prospects. In fact, even if the prospect represents the initial core upon which the estimation is built, in many cases is based upon arbitrary information gathered from the context where the decision needs to be taken. The context, in most of the cases lacks pieces of information that may either be retrieved in time or in many cases, are never revealed and the final decision/estimate is therefore grounded on partial information (Tversky and Kahneman, 1974).

In summary, ecological rationality, heuristics and consequent biases that may arise from using this intuitive decision-making tool, represent the architecture of the human mind when a decision needs to be made. It is under these set of assumptions, that the issue of optimism bias and the consequent time and cost underestimations in projects will be addressed. The aim is to create a ground for comparison and cross-fertilisation of theories arising from this framework that over time grew into the formalisation of different (some would say opposite) perspectives in the literature. The first theory that will be examined, is directly related to the anchor heuristics and prospect above mentioned. The next section, will look at how the notion of prospect is linked with the issue of the planning fallacy, establishing the first founding blocks of this discussion on optimism bias.

2.3. The anatomy of cost underestimation: prospect theory and the planning fallacy

After having discussed the chosen framework relative to how rationality can be defined in terms of making decisions, by highlighting the fact that some biases may arise in a contextual decision-making event and that those can be exacerbated by uncertain information about the future and imprecise knowledge of the present, I will now examine how this "architecture" can influence people into taking different decisions. In particular, by looking at non-maximising behaviours, I introduce Kahneman and Tversky's (1979) prospect theory. I show how this theory is able to explain the phenomenon of the planning fallacy, which represents the basis to understand better what optimism bias really is and why there are different views on how to tackle this issue.

2.3.1. Expected Utility Theory and Prospect Theory: a normative versus descriptive model

As mentioned in the first paragraph, Expected Utility Theory (EUT) suggests that when an actor is prompted to take a decision, the alternative chosen will be the one with the highest expected utility, or in other words, the option that maximises its payoff (Kapteyn, 1985). In response to this, prospect theory has been postulated in order to consider and analyse behaviours relative to choices made which do not correspond to the best alternative and that, as a consequence do not validate the principle of EUT according to which every decision maker would choose the option maximising its own utility (Hansson, 1975). Prospect theory was elaborated as a result of empirical experimentations based on hypothetical decision-making scenarios that people had to face. Notwithstanding the fact that considerations arising from experimental evidence, especially if related to hypothetical decision-making scenarios, may have some limits in terms of validity and generalisability of results, the advantages the controlled environment of laboratory or online experiment provides, is the one of simplifying conditions relative to the measurement of utility levels, which in a field study would be harder to collect and therefore analysed (Kahneman and Tversky, 1979). In these experiments that took place in three different countries, the nature of choices was initially related only to the gambling field, soon enough, however, it was showed how the phenomenon was applicable to other fields such as social sciences, economics and international relations, becoming a prominent and widely recognised theory able to explain an abundance of phenomena (Tversky and Kahneman, 1992).

During these observations, some common anomalies that were not in line with the constructs of EUT were discovered, conceptualised under three main effects: the certainty, the isolation and the reflection effects. A detailed discussion of these three effects is beyond the scope of this dissertation, nonetheless, in order to clarify how prospect theory sets out the principle of decision-making in a clear contrast with the EUT, the certainty effect will be briefly introduced. When considering the following gamble given by Allais (1953), for example:

Problem 1: Choose between

A: *£4000 with probability .80* *or* ***B:*** *£3000 with certainty*

Results:

N=95 *A: 20%* *B:80%*

It is straightforward to understand that according to EUT, the choice most people would make is the option A, as it maximises one's utility, or the payoff coming from the decision. However, experimental findings showed that the opposite is true, and the striking majority of people (80%) would prefer the certain option to the other one, even if it would reach a lower level of satisfaction, being subject to the so-called certainty effect. In this sense, certainty effect directly violates the EUT, and as a matter of fact, undermines the descriptive nature of it, suggesting that it might be rewarded as a normative model expressing how people should behave in the decision-making context.

On the other hand, prospect theory, acknowledging the boundaries in rationality of people when taking a decision, may be regarded as a descriptive model of decision making (Kahneman and Tversky, 1979).

In order to understand better the descriptive nature that according to the two academics prospect theory has, as a next step, I analyse the decision model underlying prospect theory. According to prospect theory, decision-making is the result of two sequential phases: the editing phase and the evaluation phase. During the first phase, the individual simplifies the decision according to the context in order to create easier prospects as a matter of preliminary analysis. In the second phase, those prospects are evaluated and the one having the highest perceived value is chosen (Kahneman and Tversky, 1996; Baddeley, 2013). Throughout the editing phase, in order to simplify the decision under scrutiny, some actions are performed that can help in the process, such as combination, segregation and cancellation. These actions even though are very helpful in order to facilitate the decision-making process, are based, most of times, on the context the decision is taken and this is the reason why it is likely that some bias may arise from it, as explained in the previous section. In this sense, the initial prospects of the decision are edited according to the contextual framework of the decision itself and a value is given to each alternative.

The value will not only be a function of the probability attached to the final wealth of one option or another as predicted by EUT but will also be a function of the subjective value of each edited alternative in relation to the initial reference point. Therefore, values resulting from the editing phase, will be elaborated during the evaluation phase in order to isolate the alternative that has the highest value function according to the decision maker. In relation to this, Markowitz (1952) and Helson (1964), highlighted the fact that carriers of value, are not represented by the final states of well-being coming from the different alternatives but by the changes between the different states and the decision maker's adaptation level to the initial reference point, hence embedding the already discussed subjective nature into the process, in contrast to what EUT states.

The perception of changes, following Kahneman and Tversky (1979; 1992), can influence the final decision, but are highly dependent on the status quo (or reference point) of the decision-maker. According to prospect theory, the perceived value of an option is interpreted differently according to its position above or below the reference point. Consequently, the model predicts that decision-makers are likely to overestimate losses and underestimate gains, in a way that the value function resulting from the process will normally be concave above the reference point and convex below it (fig. 2, next page).

The point of prospect theory is that people are not always consistent in their choices and, even if it is able to capture the concept of value in a more subjective rather than standardised way presents some limitations as, for example it does not account for emotions towards a determined prospect or social norms that could influence it (Baddeley, 2013). However, more recent studies in the field of neuroeconomic analysis, have showed that subjects' responses to rewards and punishment in the context of gambling experiments are consistent with the S-shaped value function elaborated by Kahneman and Tversky, remarking the validity of it in the decision-making context (Windmann et al. 2006).

For the sake of this research, it is important to point out that the formalisation of prospect theory is directly linked with the concept of loss aversion which, in general, states that gains affect perceived value less than potential losses (intuitively depicted by the value function). This theory is however connected with other phenomena such as the planning fallacy: indeed, in certain circumstances, the prospect might not be represented by the initial status of the decision maker but by a misleading future image related to the decision at hand.

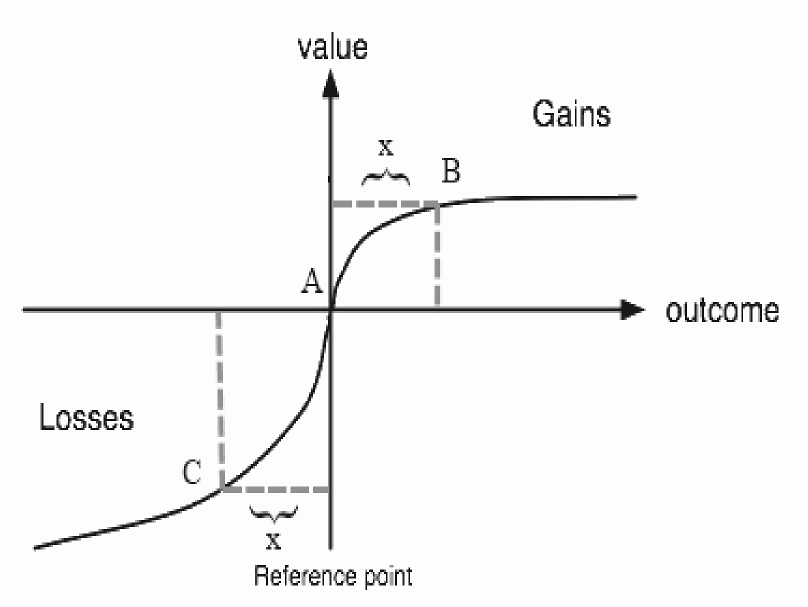


Figure 2 - Value function: overweight losses and underweight gains (adapted from Baddeley, 2013)

I have established the theoretical background to help us better understand the phenomenon of the planning fallacy, in the next section I explore what is it and why is an important founding block in the context of a discussion of optimism bias.

2.3.2. The planning fallacy

The first definition of the planning fallacy was provided by Kahneman and Tversky (1979), to describe the phenomenon according to which people underestimate the time necessary to complete a future task, disregarding past undertakings of the same or similar tasks. With this definition, the two scholars did not want to invalidate the principle of realism in relation to similar past undertakings completed, but they wanted to stress the role that optimism can play in task estimation time (Gilovich et al., 2002).

In order to assess whether an estimation has been influenced by the planning fallacy two issues are likely to occur: first, the estimated time to complete a task will result to be more optimistic than the average of the distribution of past completion time for similar tasks. Moreover, the estimated time to complete the task will be lower than the actual time to complete it (Buehler et al. 2010).

Kahneman and Tversky, did not provide empirical studies of the planning fallacy, however they reported a case study in which a group of academics they were part of were asked to initiate and conduct a project from the Ministry of Education in Israel. When estimations on the time in order to complete this project were requested, the whole team communicated that it would have taken around two years in order to complete it, even though, in their previous experiences, similar project were usually delivered in around seven years. In the end, the project was terminated after eight years and the Ministry of Education was no longer interested in it because of internal priorities changing. This case shows how planning fallacy can be detrimental not only at an individual level, but also at a collective level, with the risk to undermine the success of the task that needs to be completed (Kahneman, 2011).

Following this case study, a large number of experimental studies have been devoted to find out the impact of planning fallacy in personal time predictions to complete a task. Keeping in mind the two main issues that describe planning fallacy as mentioned above, the studies conducted by Buehler and other scholars in the realm of individuals' task estimation times, were able to show how across different categories of tasks these two characteristics were consistently observed (Buehler et al. 1994; Buehler et al. 1997; Buehler et al. 2010). Further to this, the tendency to underestimate tasks completion time has not only been studied for personal time predictions, but also in work and academic related undertakings, touching the fields of large-scale projects, software development, information communication technology projects, public administration and entrepreneurship (Bruzelius et al. 2002; Cassar, 2010; Lefley, 2013; Liu et al., 2017; Min and Arkes, 2012; Pezzo et al., 2006; Phychyl et al. 2000; Shmueli et al. 2016).

The bond between prospect theory, planning fallacy and optimism bias is now clear. The definition of the planning fallacy highlights that, most of the time, the reference point does not refer to the status quo or a previous situation but to a distorted image of the future. Transposed to a project and/or task management context, the planning fallacy is epitomised by the optimism bias, a cognitive bias arising from the deceptive formulation of project and/or task initial appraisals.

Specifically, when it comes to the project management realm, the deceptive formulation of appraisals is given by the delusional optimism regarding the attributes of the iron triangle (cost, quality and time) from one side and an excessive optimism in terms of stakeholders' capabilities during the project life cycle, from the other. Planners and project promoters, indeed, tend to overvalue positive outcomes coming from the envisioned success of the project, to oversimplify project activities and not to focus on potential risks. Therefore, promoters will tend to undertake projects that are unlikely to have the benefits planned at the appraisal stage that in many cases will lead to a situation of cost overruns and/or delays (Flyvbjerg et al. 2010; Meyer, 2014; Weyer, 2011; Denicol et al., 2020).

Cost overruns as pointed out by Flyvbjerg et al. (2002; 2003; 2005) and Kahneman and Tversky (1996), may happen for different reasons. The one just explained refers to the psychological explanation. The psychological explanation, together with the technical one, which takes into account the inadequateness of tools and forecasting systems in general, as well as "honest" mistakes by cost and risk specialists, represent the internal explanations for cost overruns.

In the literature, moreover, also external explanations have been detected, namely, those associated with economic and political motives, such as strategical misrepresentation and deliberate underestimation of costs and risks. These kind of distortions of reality, not only are deliberate, but they also entail a fully rational process and as Flyvbjerg (2008) suggests, a variety of measures can be enforced to mitigate those intentional misrepresentations. Examples of those mitigating actions are represented by the implementation of systems of rigorous accountability for projects' stakeholders and the elaboration of a set of procedures that provide incentives to get more precise estimates, for example. However, given the deliberate nature of those situations, external explanations are out of the scope of this dissertation. In Table 1, the explanations for cost overruns just discussed are summarised.

As mentioned earlier for external explanations, also internal ones cannot be completely eliminated but can be mitigated. For this reason, practitioners and academics are working towards the establishment of new forecasting techniques, not only by looking at the internal mechanisms of the project but also outside of it. In the next section, I will introduce the differentiation between the inside and outside view on forecasting. From this discussion, I will formalise a conceptual framework able to gather relevant attributes from both perspectives in order to provide a more effective basis to analyse and subsequently implement more precise forecasting techniques in project management. With the help of the conceptual framework, moreover, I will provide a more detailed discussion on cost overruns and on the different perspectives on optimism bias in the literature which will be the main focus of the second part of the literature review.

Internal		External	
Technical	**Psychological**	**Economical**	**Political**
• *Mistakes on forecasts* • *Honest errors* • *Inadequacy of business case* • *Inadequacy of project schedule*	• *Planning fallacy* • *Excessive risk-taking conducts* • *Delusional optimism vs uncertainty*	• *Intentional underestimations on accountability of resources and misrepresentations activities*	• *Use of strategies in order to underestimate costs and make a project happen (e.g. political pressure)*

Table 1 - Possible reasons for cost overruns (by author)

2.4. The adoption of an outside view on project cost forecasting

In general, when a person needs to make a prediction related to time or cost decisions, there is a natural tendency to concentrate on the given project by gathering information, developing different scenarios and ground those forecasts on single or few analogies (Lovallo et al., 2012). Therefore, people are inclined to examine the uniqueness of the endeavour or task they have to perform, rather than look for distributional similarities of precedent projects (Lovallo and Kahneman, 2003). This phenomenon was labelled by psychologists as the "inside view". According to many academics, among which Buehler et al. (1994), the inside view, triggers events that often make optimism bias arise, as decision-makers, by focusing on the uniqueness of the project, tend to be overconfident and more optimists than they should be.

The internal view, therefore, is strongly linked with the phenomenon of planning fallacy described in the last section and, for this reason, many pieces of research have been devoted to find a way to mitigate cognitive biases by adopting a different perspective that, through the use of distributional similarities, both from an historical and statistical point of view, can make forecasts and appraisal of projects more precise (Flyvbjerg, 2006).

This perspective came to be known as the "outside view" on forecasting and when linked to a project management context, it refers to the adoption of specific classes of past projects used as a base for the appraisal. Therefore, unlike the inside view, the outside one focuses on the common characteristics of the project at hand with past ones, allowing planners to avoid thinking just to the particular project and analyse it from a similarity distribution point of view. Given the experience-oriented nature of construction industry, this view is gaining more and more recognition, and governments have started to release guidelines and policies indicating how to adopt the outside view on forecasting in the most effective way according to the latest developments in the topic’s research.

The outside view, in fact, can be implemented through the adoption of a variety of methods, the most known of which is the reference class forecasting (RCF), belonging to the family of the so-called case-based reasoning methods (CBR). Such a method considers and weights the result of past tasks in projects to elaborate a more precise forecast of the project at hand (Flyvbjerg, 2008; Ji et al., 2011). This technique has been developed as a consequence of the theory of optimism bias by Khaneman and Tversky (1996) and consists of three main phases.

The first phase is to identify the most relevant reference class, in which dissimilarities are considered in order to identify the reference class that presents the highest correlation to the task into consideration. An important thing to bear in mind while performing this step is to account for a relatively large sample of projects in order to have a result that is statistically meaningful, at the same time the reference class should not be excessively large, otherwise, the comparability of data between projects could be negatively impacted.

The second phase of RCF, is the establishment of a probability distribution, in order to have a maximum, minimum, median points and, if any, clusters of data. As a matter of fact, some authors such as Ji et al. (2011) have noticed that this phase may present a lot of challenges, such as the capability of project appraisers to correctly establish the position of the project at hand on the distribution chosen and consequently, assessing the reliability of the prediction when computing the correlation with historical data. To answer this problem, the Green Book (2013) suggests some standard mitigation and contributory factors to optimism bias divided by different categories of projects, embedding some pre-set probabilities in the calculations. In the next section, with the help of the so-called “support theory” I will investigate if the use of objective probabilities can represent a correct tool to utilise in the context of cost forecasting for projects to help reducing biases arising from cost underestimation.

The third and last phase of RCF, encompasses the placement and comparison of the project with the reference class; generally, in most construction projects, the placement assumption is very close to the median point (Flyvbjerg, 2008).

The RCF technique presents some limitations, such as the hard accessibility of precise and reliable cost data, projects pertaining to different areas grouped in the same cluster (as in Flyvbjerg database) and the variability of the sample given the geographical location of the various projects (Salling and Banister, 2009). However, this technique has been proven to be effective with an overwhelming level of statistical significance (Flyvbjerg, 2018). Many efforts have been made in order to further develop these techniques, even though, most of them emphasise the importance of taking an outside view on the project at hand, risking to overlook the analysis of the unique characteristics of every project, and to exclude some categories of projects, such as the vanguard ones (Frederiksen and Davies, 2008). This conduct may eventually lead to reduce the pivotal activity of initial cost forecast of a project to a mere statistical and distributional exercise. For this reason, this research wants to explore the possibility to adopt a framework able to have a more holistic approach, grounded on the same theories of the two perspectives introduced in this section, that perhaps, rather than being opposite perspectives represent two faces of the same coin.

2.5. Enhancing the positive impact of the outside view with attributes of the inside view: Support Theory and subjective probabilities

As mentioned earlier, the outside view presents many positive features and can mitigate some aspects of the planning fallacy by considering similar past decisions or endeavours. In fact, by considering a probabilistic rather than deterministic approach on estimations, the outside view, is able to give a more complete picture on the different scenarios that are likely to happen given past similar events (Buheler et al., 1994). Applying this perspective to cost forecasting techniques, has been showed to be effective in partially mitigating, the optimism bias usually arising when appraisers make cost estimations for projects. However, academics and practitioners that are usually favourable towards this approach, disregard an important matter: adopting an inside view when making forecast of any kind does not always produce an outcome of planning fallacy (Kruger and Evans, 2004), indeed, in some cases, the opposite may be true. In fact, it may happen that by adopting a perspective that is “too external” and does not enter in the known details of the task or project that needs to be estimated can result in an increase of the planning fallacy, as appraiser’s perspective may be diverted excessively from thinking about the specific endeavour.

From a theoretical perspective, it is possible to find supporting evidence that explain the above-mentioned supposition in another work based on decision-making analysis and judgement capabilities in the behavioural economics literature, the support theory (Tversky and Koheler, 1994). The starting assumption of this theory is the fact that when a person is asked to make a decision, it will base this on subjective probabilities, or in other words, on the degree of belief expressed both in form of direct judgment or as a choice between different events (Kahneman and Tversky, 1983). The decision will be, therefore, dependent on many variables, such as past experience, different opinions or simple intuition; there will be, on the other hand, also other factors at play during this decision, and this theory, by recognising the nonextensionality of subjective judgements (i.e. objects of decisions are not deemed to be equal even if they have the same external properties) postulates that probabilities associated with those judgements are not linked to events but to the way the events are described. Following this logic, the resulting probabilities’ outcomes, will not be equal to the probability that the event occurs, as predicted by probability theory (Fischoff et al., 1978) but will be assessed in terms of the support corresponding to a specific hypothesis derived from the description of the event of the judging probability.

As a consequence, the main assumption of this theory is that unpacking an event into subcategories is able, in general, to positively influence its support: for example, if the event to be considered is “a building collapse”, and two subcategories are “building collapses because of foundation failure” and “building collapses because load is heavier than expected”, then the support relative to the two disjoint events will be equal or greater than the support relative to the event that does not have further cause description. This principle, as Tversky and Koheler (1994) mention, is not only related to probability judgement but is something that can be applied at a greater level as a founding characteristic of human judgement.

Before exploring further how the attributes of the support theory can enrich the analysis on cost forecasting, a clear differentiation between the concepts of unpacking and decomposing a task or project should be outlined. Following what Kruger and Evans (2004) report, when practically operating on a task and the requirement is to decompose it, next step is to divide the task into subcomponents and to make separate forecasts for each of them that will be subsequently aggregated (not important the actor that performs the aggregation task). Unpacking, on the other hand, requires only to break down the task in a figurative way, altering the description or representation of it in order to enhance the accessibility to the parts that constitute the whole task, so that the forecaster is able to elaborate a single judgement and not numerous as in the case of decomposition.

In light of this, unpacking (and relative support theory) rather than decomposition principle has been chosen in order to create the conceptual framework of this paper, because it yields to a single judgement elaborated by the same person that is required to complete the estimation task, as it happens in reality when appraising for a project.

As a result of these reflections, support theory may be presented as a possible route to follow in order to address the issue related to the underestimation of time and cost arising from the planning fallacy, as it accounts for factors that are oriented towards the inside view rather than focusing solely on the outside view. Further to this, unpacking can give a more specific perspective of the task at hand and in the case of projects, could give to forecasters the opportunity to gain better and more specific insights by focusing on how to divide into subcategories the estimation to be performed rather than just focusing on similar past projects. Indeed, unpacking a determined project or task, by offering a different outlook, may remind forecasters of possibilities they would have not accounted for and at the same time, may give importance to the various milestones to be achieved in order to conclude the endeavour in a way that a more precise snapshot of the object of the estimation can be built in the forecaster mind before performing the appraisal.

With this in mind, it becomes clear that, if the main attribute of support theory (i.e. unpacking), is integrated with the outside view on forecasting, some of the shortcomings of this technique might be mitigated. When considering the case of RCF, one of the weaknesses of the method is related to errors in the positioning of the project in the reference class (and the selection of the reference class itself): if the appraiser, before conducting the analysis on the selection of the reference class and the relative position of the project in it, would unpack the project at hand, this might have the effect to unveil characteristics of the project that were not considered before, resulting in an overall improvement of the quality of the appraisal. In the same way, unpacking may help in isolating characteristics of the project at hand that could enhance the comparability between the project at hand and past projects in the reference class, which, as discussed, is a fundamental issue when it comes about adopting an outside view to develop cost forecasts (Lovallo et al., 2012).

Also, another relevant attribute of support theory in the context of cost forecasting, is the fact that it entails the use of subjective probabilities: earlier on, it was discussed that in order to reduce optimism bias the Green Book (2013) suggests that some contributory and mitigating factors can be used in order to consider if the optimism bias final calculation on the project at hand can be reduced. However, contributory factors constitute a standardised measure given by policymakers not aimed at being changed and therefore represents an objective probability (they are expressed under the form of a pre-set description that cannot be changed by the forecaster without altering the result of the calculations). Differently, unpacking, is based on subjective probabilities, so that the application of it on forecasting may have the opportunity to make estimations more reliable, as, with the possibility to manipulate the probabilities related to the contributory factors, it may focus on aspects of the project at hand that would be overlooked by using standard measures.

Moreover, support theory, through unpacking, has the potential to improve forecasts accuracy and, as a consequence, reduce the impact of optimism bias; unpacking, indeed, is more effective when performed on complex tasks

(Kruger and Evans, 2004), and this is the reason why it may seem particularly suitable in the appraisal of projects' costs and schedule.

The above discussion, suggests therefore that, integrating some attributes of the inside view on the outside view on forecasting, have the potential to have a beneficial impact in relation to the reliability of the final estimates and that, the issue of the planning fallacy, may be analysed by looking at it from a more comprehensive perspective.

2.6. Integrating different perspectives on cost forecasting: the Holistic view

Prospect theory and support theory have a different interpretation of how to overcome the planning fallacy, this, however, does not mean that the two explanations are mutually exclusive. On the contrary, they both contribute to give insights helpful to clarify the issue of project/task estimation. Prospect theory, in fact, informs on the importance of comparing the task of estimation at hand with similar past cases adopting a probabilistic mind set, whereas support theory informs on the opportunity to increase the awareness on the components that constitute that given task. Interpreting the theories in this way gives the opportunity to look at the issue of the planning fallacy in a more holistic way, opening new routes of exploration both at a general level and at more specific cases such as the one of cost forecasting for projects.

In this setting, I introduce the conceptual framework that this dissertation proposes: analysing the problem of cost underestimation and its corollary of making more accurate estimates, should not split between two different and incompatible perspectives, as in the current literature (especially in the project management field); it should focus on a more holistic approach that capitalizes on the positive features of the two perspectives to improve the precision of forecasts.

As can be observed in fig. 3 a process of integration (Phase B), both at a practical and at a theoretical level is promoted in order to create new knowledge in this field and create a set of techniques, methods, approaches and regulations. Those should be able to consider a “holistic” view on cost forecasting rather than two scattered perspectives that even though providing relevant insights, have been proved to have many limitations, as previously discussed.

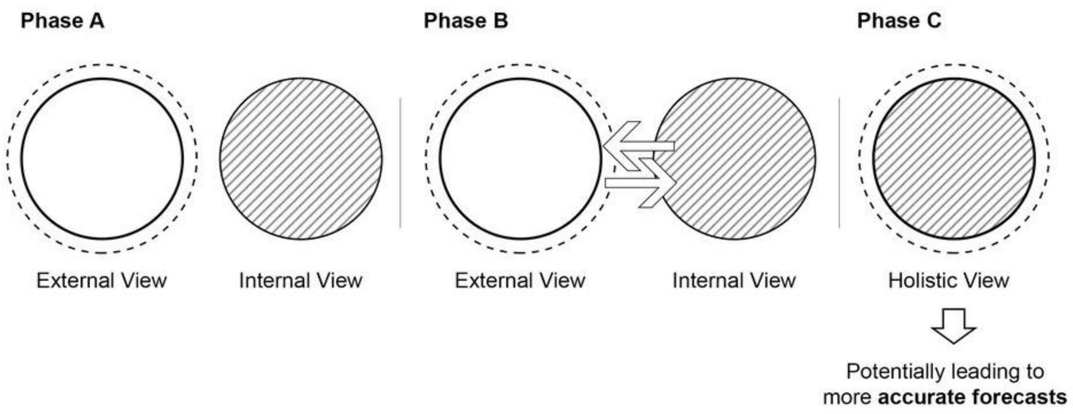


Figure 3 - Integration of the External and Internal view to create a Holistic View on forecasting (by author)

This framework may help in devising what are the strengths and weakness of each perspective to operationalise them into the cost forecasting techniques that are currently being used to make them more precise and effective decision-making tools. The framework is intended, therefore, to provide a ground for analysis as suggested by Semiatycki (2009) in blending qualitative and quantitative approaches to develop innovative techniques to collect and analyse data from different sources able to contribute to an advance in the understanding of the mitigation of optimism bias in the project management context. A critical assessment of those techniques will be the subject of the second part of this literature review.

2.7. Theoretical implications

One of this literature review objectives, was to introduce a new conceptual framework in order to analyse the problem of cost underestimation and forecast accuracy during the phase of project planning. To formalise the framework, I first discussed the existence of cognitive bias in projects using decision theories from behavioural economics. This background was used to pave the way for the psychological reason behind cost overruns, present the planning fallacy and investigate a possible explanation of it thanks to the prospect theory. After that, I analysed a perspective that looked at the appraisal of projects as a distributional historical analysis of past project characteristics, the outside view, to understand its main positive and negative features.

As opposed to the outside view, I explored the inside view on forecasting, to understand whether some attributes offered by the support theory and its main assumption related to unpacking may provide new insights on how to mitigate the issue of cost underestimation. Furthermore, I explained that unpacking a task, or, in other words, figuratively breaking down a task and manipulating the description or representation of it, has the potential to reduce the planning fallacy, and make estimates more accurate. Also, thanks to the use of subjective probabilities, it may unveil contributory risks that are not part of the standardised measures found, for example, in the Green Book (2013) and therefore helping in getting more reliable estimates.

In light of these findings, I proposed that, in order to gain new insights and to approach the problem of cost underestimation, a holistic view rather than inside or outside view on forecasting should be adopted, creating a conceptual framework that integrates the two perspectives considering the two theories presented.

Moreover, it should be mentioned that the proposed framework supports some precedent contributions to the field, such as the one of Koch (2012), in which the author advises to integrate some aspects of the outside view, in specific the RCF, with some socio-technical features that the inside view can provide in order to improve the quality of investments' decision-making for engineering projects. Even though Koch only focuses his studies on engineering projects, given the previous discussions on the presence of the issue of underestimation throughout different industries, these considerations can be generalised for the sake of this research.

Further to this, some limitations on how to operationalise this integration may be devised: first, in practice, no forecasting technique embedding internal and external view has ever been put in practice for more than one project, an example of it is the Salling model (Salling and Banister, 2009) that considers distributional and deterministic calculations (no mention of unpacking though), therefore there might be resistance from practitioners to adopt it. In order to solve this issue, the necessity to make more research on the topic and to empirically provide a practical ground for the validity of this framework should be carried out.

Also, the proposed framework is based on behavioural economics constructs, which, acknowledging the limits of human rationality also make a point on the difficulty to study patterns related to human behaviour: every forecaster and every decision-maker will adopt different behaviours in different circumstances. However, if specific patterns are isolated for study, they may lead to the formulation of techniques that even if not able to solve completely the issue of optimism bias (something that is unfortunately impossible to achieve) may help in mitigating it in a more efficient way. In this sense, the advice is to devote more research efforts in exploring appropriate and innovative laboratory and out of laboratory simulations able to mimic in the best possible way what happens in reality.

Finally, I mentioned how the two theories and the two perspectives presented differ from each other's; however, I should highlight that they both emphasise how at the base of the inaccuracies in estimations, human behaviour and judgement play a central role. This, therefore, constitutes an appropriate ground in order to analyse, further, with the help of the experimental methodology, how cross-fertilising and integrating those two theories may offer an enhanced "toolbox" for project appraisers to get better estimates.

Before that, it is however necessary, to further analyse the different existing perspectives when it comes about exploring over-optimism in forecasting methods. In this way, we will be able to make a more detailed analysis on overruns theories, that will help to develop a more comprehensive discussion on the existing methods I analyse in the context of this research.

CHAPTER 3. COST AND SCHEDULE OVERRUNS: DEFINITIONS, PERSPECTIVES, TECHNIQUES AND POLICIES

3.1. Schedule and cost overruns: different literature perspectives

At the beginning of this literature review, I mentioned that one of the first assumptions of this research is that ignorance is bad for projects: in other words, the assumption is that when decision makers are prompted to take a decision on the approval or rejection of a project, it is beneficial for them to have as much information as possible. In saying this, the aim is not to forego the fact that at the preliminary stages of a project the level of uncertainty and ignorance is quite high (Hirschman, 1967). However, in the context of this research, I put emphasis on the fact that following a direction that overcomes ignorance represents a good way to get more accurate forecasts and that has the potential to help in reducing the inefficient use and allocation of resources decreasing the likelihood of delays (Flyvbjerg and Sunstein, 2016).

Further to this, I want to highlight that, ignorance, as intended in this dissertation, does not have an unilateral definition which the literature usually refers to as absence of “true” knowledge (Greisdorf, 2003). Ignorance may also refer to other traits and biases which may be driven by social constraints, contextual situations or being beyond the cognitive bounds of an individual, in accordance with what has been mentioned earlier about people’s architecture of mind (Kutsch and Hall, 2010; Slovic et al. 2002).

Is considering these initial thoughts that in this section, I introduce three different perspectives, all looking at overruns originated by the over optimism in schedule and resources forecasting and the way it should be perceived and dealt with: the "*Hiding Hand*" perspective, the Evolutionist perspective and the behavioural perspective.

This theoretical introduction to the issue of overruns, represents the link between the conceptual framework just discussed and its application to the infrastructure construction industry. Next objective is, therefore, to establish a definition of cost overruns, the frequency of it, the measures certain countries have decided to put in place to mitigate it and what other techniques may be used following the internal and external perspectives; these topics will cover most of the second part of this literature review.

3.1.1. The "Hiding Hand" perspective

A. O. Hirschman was the first scholar to emphasise the importance of project appraisal in the context of development projects, by proposing that the World Bank participate in the attempt to systematise project evaluation (Picciotto, 1994a; Willoughby, 2003). The result of this joint effort was the publication of the book "Development Project Observed" in which the economist presents right at the first chapter the "Principle of the Hiding Hand". This theory created disagreements with the multilateral organisation that supported the Professor's research, as it recognised uncertainty as a constituting element of the decision-making process, a principle in clear contrast with World Bank operational focus (Alacevich, 2014).

Hirschman, in fact, was more interested in exploring the overarching principles of project appraisal, especially in the context of development economy, rather than providing an operational analysis aimed at the betterment of the forecasting project, which was the main objective of the World Bank. Therefore, the result of his field research, based on a sample of 13 projects located in four different continents, reflects the above-mentioned aim of the scholar (Rodwin and Schön, 1994; Woolcock, 2019).

According to Hirschman, ignorance is an important condition for stakeholders to initiate a project in the first place, as, if all the risks associated with the project would be known from the beginning, many endeavours would not start, particularly in the case of developing and underdeveloped countries. Further to this, the economist highlights that project estimators do not only overlook potential risks and, consequently, underestimate costs (partly identifying and recognising the account of the planning fallacy that will be presented twelve years later by Kahneman and Tversky, 1979) but in most cases they do not consider their capabilities to overcome the challenges to be faced during the project. In this sense, appraisers, do not contemplate the degree of resilience and creativity that potentially will help stakeholders in delivering the project notwithstanding the existing challenges (Lynch, 2019).

As a result, the Hiding Hand is capable of hindering threats and through its "*providential ignorance*" to make sure that the deliverables of the project will be in line with estimators' forecasts (Alacevich, 2014). It is interesting to highlight at this point that, Hirschman refers to the Hiding Hand as a beneficial situation that is able to accelerate and in some mysterious ways optimise the decisions' engine of planners, which is crucial not only for those projects to be initiated in underdeveloped nations but for all projects as a general category (Hirschman, [1967], 2015).

The literature, however, provides many examples that would indicate the fact that the Hiding Hand is not inherently benevolent, as there are cases in which the underestimations of the potential challenges and costs are not outweighed by the underestimations in terms of creativity to solve the problems and potential benefits arising from the projects (Flyvbjerg and Sunstein, 2016; Williams, 2017; Honig, 2018; Lynch, 2019). In light of this, Streeten (1984) first and Picciotto (1994b) afterwards, started to postulate the existence of a second hiding hand, with a more malevolent nature, that, as a matter of fact, has the same attributes of the planning fallacy, presented earlier on in this literature review. The authors did not substantiate this claim with strong data, and, as the studies were from officers of the World Bank, many practitioners and scholars disregarded the issue for a long time, considering the research biased given the disagreements between the World Bank and Hirschman.

The first and more comprehensive study that looks at the dichotomy of benevolent and malevolent hiding hand, is the research of Flyvbjerg and Sunstein (2016) that tested the principle against a sample of 2062 projects, a far greater sample than the one of 13 projects used by Hirschman in 1967 to elaborate the principle. The data were analysed looking at the difference between estimated versus actual costs and estimated versus first year benefits, by considering whether the net benefits of the projects outweighed their final costs even in those cases where the final costs were higher than what was initially estimated, as theorised by Hirschman. In other words, Flyvbjerg and Sunstein were looking at the variance between cost and benefits overruns and in the case of a Benevolent Hiding Hand, the average benefit overrun would have exceeded the average cost overrun. The results of the study showed that on average, for the projects analysed, benefit overrun was not more than cost overrun, and that in many cases, there was a total absence of benefit overrun, or, in other words, a benefit deficiency. The two authors went further ahead, analysing the percentage of project presenting a benevolent hiding hand and the one presenting a malevolent one: they found that 78% of projects presented the action of the malevolent hiding hand, de facto discarding Hirschman's thesis on the benevolence of the Hiding Hand. Following Flyvbjerg and Sunstain, other authors studied the incidence of the malevolent Hiding Hand, reaching very similar conclusions, as in the case of Williams (2017) that studied public projects in Ghana with a sample of 14000 projects and in the case of Honig (2018) that used an even more extensive project sample base from different geographical areas.

Having said that, it is necessary to clarify that even if multiple studies rejected the predominance of the Benevolent Hiding Hand, this does not mean that there are no examples of projects in which the phenomenon is verifiable, but these examples, rather than representing normality as speculated by Hirschman, represent the minority of cases.

The consequences of this are straightforward: evaluating a forecast on the grounds of the Benevolent Hiding Hand is misleading and the likelihood that the project will present a benefit shortfall and a considerable cost overrun is quite high. For underdeveloped economies, this situation, may represent an even bigger challenge as a project failure is likely to have a long-term impact on their economy. Indeed, considering as viable the Hiding Hand perspective both at a policy level and at a project management level does not seem to fit the reality and may result in huge monetary, time and/or skills losses leading in some cases to even worse outcomes, such as projects' failure.

Finally, this perspective, even if from one side recognizes the propensity of people being over optimist in their forecasts, does not recognize the negative impact that his might have, overvaluing the "problem solving" capabilities of project's stakeholders. Moreover, those capabilities will only be exhibited by stakeholders during the project execution phases and for this reason, the Hiding Hand account is not deemed to be able to portray in a correct way the forecasting process of cost, schedule, resources and benefits.

3.1.2. The Evolutionist perspective

The Evolutionist perspective takes a step back: it discards the assumption of individuals over optimism and, in general, denies human behaviour as being the founding phenomenon that negatively affect cost underestimation, cost overrun and ultimately the project performance. As a direct consequence, it rejects the explanation according to which the planning fallacy is the main cause that drives cost and schedule down in initial estimates (Love and Ahiaga-Dagbui, 2018).

For this reason, many studies that have been made following the direction of this perspective do not consider overruns as resulting from biases, but errors, which are related to actions rather than behaviours that were not initially accounted for (Aibinu and Pascu, 2008; Amadi and Higham, 2017; Locatelli et al. 2017; Love et al. 2010, 2014; Odeck, 2004; Park and Papadopoulou, 2012). These actions, as described in Love et al. (2011) can be defined as "*pathogens*" and are very likely to occur between the initial stages of the project and the delivery of it. Examples are represented by changing in scope, errors in estimations and increased cost of raw materials. Pathogens, as a consequence are liable to drive up costs, therefore, all efforts in order to reduce the variance between initial estimated cost and final costs should be dedicated to the identification, analysis, and study of the interdependencies between these factors (Love et al. 2016).

This belief has been translated in several studies aimed principally at understanding the main causes of cost overruns in terms of project dependent variables, that, as a matter of fact emphasise the vision of cost variances as being highly dependent from endogenous factors of the projects and discard as a minor cause the account of the planning fallacy (Love et al. 2019). Hence, these studies are not oriented at finding the overarching principles that explain the phenomenon, but at presenting a more operational view, attempting to solve the issue with "hard tools" that, if adopted, would allegedly create a more effective ground for cost estimation discrepancies mitigations (Ahiaga-Dagbui and Smith, 2014).

Following, I analyse studies that endorse the evolutionist account in order to gain better insights on the evolutionist perspective and understand whether it represents a viable perspective to adopt in the context of this research or not.

Jahren and Ashe (1991) conducted one of the first studies attempting to detect predictive causes leading to cost variations between the initial and completion stages of a project. In this study, data from 1576 construction projects were considered from two different geographic areas ranging from $25000 - $1M.

Findings highlighted that among the most prominent reasons why cost overruns occur are the type of construction, the level of competition during the bidding phase and the dimension of the project. In the context of these results, a statistical analysis in which the distribution of cost overruns for the sample of projects was studied as well: it is interesting to see how this distribution is non normal and how the authors suggest that this consideration could be helpful in order to build parameters, methods and models able to mitigate the issue. This is a pivotal consideration to better understand the issue of cost underestimation and the phenomenon of cost overrun arising from it and therefore, a more in-depth analysis of the distribution of cost overruns is done later in this work. One limitation of this study is the fact that the size of projects analysed is quite small. In this sense, the motivations highlighted in the paper for cost overruns might be more linked to the small dimension of the projects, inherently less complex than bigger ones, where issues like over optimism and strategic misrepresentation are likely to be less present.

Among the supporters of the evolutionist perspective there are also Bordat et al. (2004), which studied the incidence of cost overruns and time delays for transportation projects in Indiana. Their main findings were that cost overruns are mainly caused by situations such as design changes, different than expected site conditions and alterations in the scope of the projects. When talking about design changes, the authors suggest that the predominant reason for changing orders is “errors and omissions” (Bordat et al., 2004, pag. 65) which would suggest that at the basis of the variance between initial estimates and final count of monetary resources used, there is a component of mistake which derives from consultants, designers or other stakeholders.

From this instance it is clear that evolutionists, even if recognising the human component in the issues of over optimism and cost overrun, tend to operationalise it into project tasks, labelling it under different categories of actions, same actions that Love et al. (2011) call “pathogens”.

Odeyinka et al. (2010) conducted another study looking at the different kind of risks associated with cost overruns in order to create a model able to mitigate the phenomenon. As in the previous study they reported that the major risks associated with cost overruns are given by changing in scope, in design and unexpected conditions on site. Perhaps, the most important contribution of this work, is not related to the categorisation of the factors potentially affecting cost overruns, confirming the findings of Bordat et al. (2014); but is related to the recognition that to effectively tackle cost overruns is necessary to adopt a tailored approach for every project. This consideration, could be interpreted as emphasising the importance of concentrating the forecasting efforts on the potential risks that could arise from the project at hand, as advised by the inside view. On the other hand, it could also open to the use of project-specific techniques or models that are not only endogenous and that by adopting a more holistic approach could represent a more efficient way to mitigating over optimism, as advised in Ahiaga-Dagbui and Smith (2014) and as baselined throughout this research through the conceptual framework proposed.

In a more recent study, Love et al. (2019), analysing a database of transportation projects in Hong Hong with a cumulative value of around HK$115 billion completed between 1999 and 2017, found that cost overruns do not follow a normal distribution, confirming the findings in the first study analysed (Jahren and Ashe, 1991). They also suggested that more rigorous quantitative tools (e.g. QRA, Montecarlo simulations) should be adopted in order to minimise the risks of cost overrun in transport infrastructure projects.

This study, like all the above just examined, presents a mechanistic view of how the initial optimism in project forecast may translate into cost overrun, enabling to identify contributory factors and causes of it. Identifying contributory factors, even if providing many interesting insights, still has not helped in improving accuracy in estimations (Flyvbjerg, 2005). This is perhaps symptomatic of the fact that the evolutionist perspective, by focusing all the efforts on actions or pathogens that lead to cost overruns, might not give enough importance to less mechanistic and more behaviourally oriented factors. Recognising the importance of human behaviour not only during the forecasting stage of the project but also during the whole project life cycle is, in fact, vital in order to adopt a more holistic view and mitigate the issue of cost and schedule overrun.

As a result, the evolutionist perspective, even if providing a number of insights on the topic of over optimism in forecast is not deemed to be strong enough to mitigate the issue of cost and schedule underestimation and, as a consequence, a cross-fertilisation with another perspective is advised: the behavioural perspective.

3.1.3 The Behavioural perspective

As can be deducted from the name of this last perspective under analysis, the “Behavioural perspective”, emphasises the importance of individuals’ behaviour in forecasting as postulated by some of the studies in the field of behavioural sciences (Kahneman and Tversky, 1979, Gilovich et al. 2002). These are some of the studies that were introduced at the beginning of this literature review, in order to elaborate the model of the "architecture of mind" which helped in formalising the conceptual framework for this research work.

As a matter of fact, the so-called Kuhnian revolution of behavioural science, based on the model of heuristics and biases, is the archetype for the studies of the authors embracing the behavioural perspective that taking the outputs of these studies operationalised them into frameworks, models and even forecasting techniques, such as the Reference Class Forecasting (Flyvbjerg et al. 2009).

In the previous section, I reported that according to evolutionists, overruns result from pathogenic actions that can be categorised and operated upon. Even if at a first glance this explanation may seem perfectly acceptable there are some other factors, besides the one analysed above, that might need a bit more attention when adopting this perspective. These pathogens, certainly, are present during the project life cycle, but rather than representing the emergent factor of the problem, more reasonably, may be regarded as a manifestation of certain issues that happened during the front-end decision-making phase of the project. Those issues, originated from biases such as overconfidence, de facto, do not constitute the general causes of the overruns as in the case of the pathogens, but their root causes.

The first difference between the evolutionist and behavioural account is the level of analysis used in order to tackle the problem of overruns. Following an example of the proposed causal chains of event that leads to cost/schedule overruns according to the two different perspectives, in order to better understand this differentiation in terms of level of analysis of the issue, derived from Flyvbjerg et al. (2018), Love and Ahiaga-Dagbui (2018) and Love et al. (2019).

The cost arising from the compliance to a new regulation can be seen as a general cause of cost/schedule overruns, as can be found in the studies of Odeyinka and Perera (2009) and Potts and Ankrah (2008), therefore seems to be a fitting situation for the example. When considering the causal chains of events that leads to overruns, according to evolutionist there will be a three-step process leading to those as showed in fig. 4 next page.

The causal chain starts directly with the isolation of the cause and how the specific issue reacts in relation to the other pathogens that can be identified in the project, leading finally to the problem of overrun; as mentioned earlier in this case rather than bias, evolutionists talk about error. Even if some studies by the authors proposed methods to mitigate the problem considering this model, no final results on the applications to real projects substantiates these ideas (see for example Ahiaga-Dagbui and Smith, 2014).

The causal chain according to behavioural perspective supporters, on the other hand presents a different starting point, given by the root cause and not the contingent cause of the problem. Also, provides an interesting differentiation between the effect of underestimation during the planning phase of the project and the non-accountability of the same factor during the delivery phase, as represented in the fig. 5 next page.

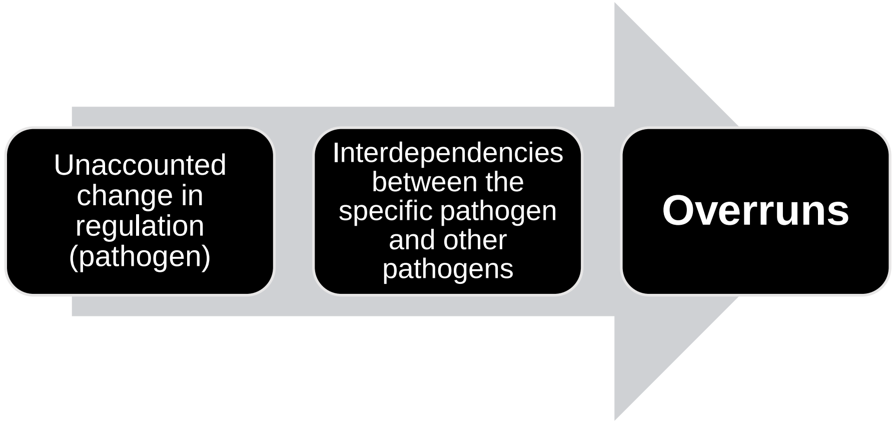


Figure 4 - Causal chain leading to overruns according to Evolutionists (by author)

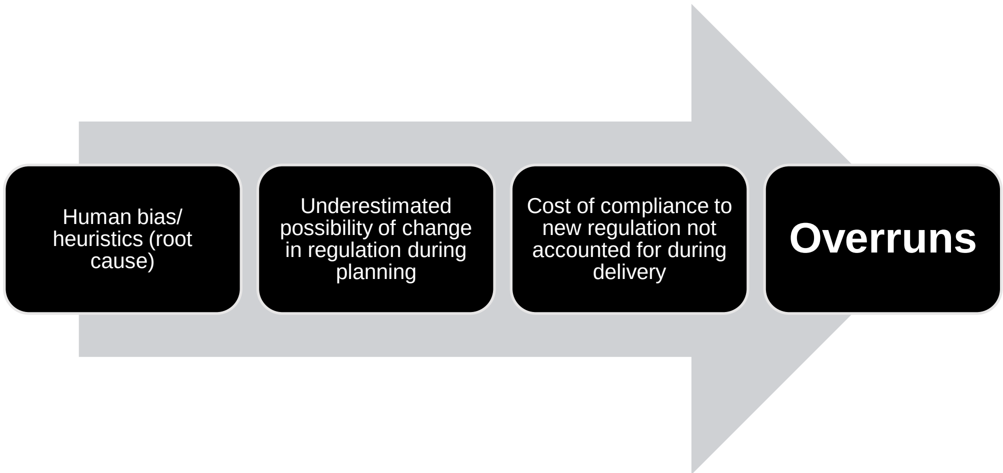


Figure 5 - Causal chain leading to overruns according to Behavioural perspective supporters (by author)

The behavioural perspective, based on this, claims that in order to effectively mitigate the detrimental effect of overruns for projects it is necessary to put effort into the understanding of the root causes of the problem (human behaviour), given the fact that, notwithstanding the considerable number of studies devoted to the general causes of overrun, this has yet not produced results in practical terms (Flyvbyerg, 2013).

In one of the previous sections, I introduced the planning fallacy, with a table (Table 1) reporting possible reasons for overruns: these reasons according to the behavioural account, are summarised by two concepts: delusion and deception (Flyvbjerg et al., 2002; 2003; 2004(b); 2005; 2009; 2013 (a); 2018). The term delusion, is related to those explanations that in Table 1 are internal, therefore arising from human biases such as overconfidence and anchoring. Deception, on the other hand, refers to intentional behaviours driven by political or economic motives leading eventually to overruns.

Delusion and deception do not represent a mutually exclusive dichotomy, as posed by the evolutionists in Love et al. (2018), but they are complimentary explanations for overruns, that in the context of a project can appear simultaneously (Flyvbjerg, 2009). Evolutionists, at the same time, discard the notions of delusion and deception, disregarding the abundant number of studies that show how deceptive behaviours brought to the underestimation of cost/schedule and consequently to overruns (Wachs, 1986; 1989;1990; Cantarelli et al. 2010). Thus, limiting their analysis only to “mechanical” pathogens that might not be enough in order to solve the problem of overrun. When it comes to the analysis of the concepts of delusion and deception, therefore, it seems that the behavioural perspective presents a more complete explanation. Furthermore, with the differentiation between cause and root cause for overruns, the behavioural perspective provides a more theoretically solid explanation (bias and heuristics theory by Kahneman and at the time deceased Tversky won a Nobel Prize in 2002). For this reason, the same constructs are used as well throughout this research.

Another trait of the behavioural perspective is that it firmly recognises the importance of taking an outside view when prompted to make a forecast of any kind as initially advised by Kahneman and Tversky (1979). As an answer to this, Flyvbjerg (2008), developed a forecasting technique called Reference Class Forecasting, that by analysing the project at hand in comparison with past similar projects, has the principal aim to give more reliable forecasts. Even though there are instances throughout the literature of the positive impact this forecasting technique has (Batselier and Vanhoucke, 2016; Bordley, 2014; Kim and Reinschmidt, 2011; Liu and Napier, 2010), it should be pointed out that, as in some of the studies cited above, RCF, has been deemed to be an optimal tool to be used also in addition to some other techniques such as EVM (Earned Value Management). In this sense, even though the behavioural perspective mainly advises the adoption of an outside view in forecasting, it should be more open towards the use of mixed techniques with the inside view, in an effort to improve forecasting with an integration process, which is ultimately the theme of this research.

The behavioural perspective, finally, when compared with the Hiding Hand perspective, seems to offer a more compelling argument backed up by stronger data. In a way, the Hiding Hand account, may be deemed to be the starting point for the development of the current theories on over optimism and overruns that once divested from its “mysticism” by studies such as the one of Flyvbjerg and Sunstein (2016) created a fertile ground to build the blocks of the behavioural theories.

Even if acknowledging the limits of the behavioural account as mentioned above, in the context of this research, I consider this perspective to be the one that is able to explain the phenomenon of overruns and over optimism in the most complete way for the reasons explained throughout this literature review. In the next section, I will establish the definition of cost overruns, the frequency and magnitude of it and how to mitigate it, considering mainly the set of principles advocated by the behavioural perspective, with some differentiations and/ or inclusions from the other two perspectives when deemed necessary.

3.2. Application of the theories to the infrastructure construction industry: insurgence of cost overruns

In section 2.1, I introduced the mechanisms underlying the constitution of the phenomenon of cost overruns, providing also a number of reasons why this issue occurs, in line with what the literature on the topic states. That preliminary analysis, together with the further considerations made in the last section about the “delusion and deception” models (Flyvbjerg, 2009), even if framing the issue, do not give any other information. For this reason, after discussing more extensively the relationship between cost underestimation and cost overruns, the next step is to provide a definition of cost overrun adopted to conduct this research and how to correctly measure it. Following, I will discuss the frequency and size of the issue in the infrastructure construction industry and the possible actions to be taken in order to detect in advance the issue and mitigate it. Finally, to appreciate the importance of this topic for Governments, Public Authorities and policy makers, I present an overview of policies on the matter in a few different countries.

It should be mentioned, moreover, that most of the considerations in this section are applicable, in general, as well to schedule overruns. However, given the fact that the majority of existing literature focuses only on cost overrun, I adopt the same line so that reported research on the topic is not distorted to fit a concept, schedule overrun, which even if in principle very similar, is different in practice.

3.2.1. From cost underestimation to cost overrun

Fig. 5 in the previous pages, shows the causal chain, according to the behavioural perspective, leading from human bias to cost overruns. Perhaps, the most interesting dissimilarity with the causal chain of the evolutionist account, is the distinction between cost underestimation and cost overrun, a crucial point in order to understand how to tackle the problem under analysis. Indeed, by considering human behaviour the root cause of the problem, the emphasis on the resolution of it is not posed anymore on the overrun itself (i.e. the outcome of the process), but on a step before: cost underestimation. The behavioural perspective, consequently, advises that in order to address the issue of cost overruns, it is necessary to consider the real problem underlying it, which is the cost underestimation (Flyvbjerg et al. 2018).

The relationship between cost overruns and cost underestimation is, therefore, of a temporal nature, and, as showed in the causal chain the latter comes before the former. As a matter of fact, cost underestimation will reveal itself a long time before overruns and if corrective measures are put in place in order to detect it and de-bias it, likelihood is that this will reduce the risk of overrun (Flyvbjerg, 2013 (b); Kahneman et al., 2011). This consideration is also supported by a study from McKinsey which, analysing more than a thousand business investment decisions concluded that, when de-biasing techniques are put in place during the decision-making process, this can yield higher returns (Lovallo and Sibony, 2010). In this sense, analysing the impact biases have during the appraisal process by taking into consideration past similar occurrences has the potential to help in getting this estimation right, as promoted by the outside view and the use of RCF. The use of de-biasing techniques, furthermore, scales down the risk of cost underestimation, decreasing the amount of uncertainty present at the time of the decision and, as a consequence, diminishing the probability of overruns at the end of the project (Leavitt et al. 1993).

Uncertainty and lack of information, however, are constituting features of the front-end phase of any project and notwithstanding the many techniques that can be adopted in order to reduce it, the risks associated with them will never be completely eliminated (Williams et al. 2019). Cost underestimation is therefore the product of the many risks, uncertainties, missing pieces of information and behavioural biases that might occur during the initial phases of the project, culminating eventually into cost overruns and sometime in the failure of the project itself. Once established why is important to consider cost underestimation and how this phenomenon leads to cost overrun it is now the time to define cost overrun and establish what is the best way to measure it in the context of this research.

3.2.2. Defining and measuring cost overrun

The term "cost overrun" has been used in the literature to indicate the increase in cost from the initial budget and in some instances has been used interchangeably with the term "cost escalation", an example of this can be find in Flyvbjerg et al. (2002). However, in Love et al. (2015), a difference is drawn between cost overrun and cost escalation, defining the latter as the anticipated variation of the initial cost given the effect of non-controllable time and market forces (e.g. inflation). Recalling the different reasons for cost overrun outlined in section 2.2, considering the two terms as indicating two different phenomena seems to be reasonable, therefore, in the context of this research, the terms cost escalation and cost overrun are not used interchangeably. In this way, every time during the course of this dissertation that the term cost overrun will be adopted, this will exclude from the analysis non-controllable market and time externalities, focusing on the four reasons exhibited in Table 1.

This clarification is particularly useful when analysing the results from the experiments designed for this research. In fact, in line with the initial sections of the literature review, I will focus on the psychological reasons for cost overrun in the analysis of the experiments. This distinction helps in reducing the potential noise arising from the consideration of variables that are not relevant for the analysis of the experiments. At the same time, it identifies the first characteristic of cost overrun as not being dependent from non-controllable time and market forces when measured. Finally, this distinction should be considered as a rule of thumb for all those studies on overruns that seek to compare different project cost overruns in relative terms (i.e. looking at percentage variations), so that more reliable figures for analysis may be created. Once established the importance of distinguishing between cost overrun and cost escalation, is now time to look at some of the different definitions of cost overrun that can be find in the literature.

Conventionally, cost overruns are calculated in relative or absolute terms: in absolute terms, the calculation will be done considering final cost of the project minus the estimated cost. In relative terms, cost overruns will result from computing either final cost as a percentage of estimated cost or as a ratio of the final cost divided by estimated cost (Invernizzi et al. 2018).

In the literature, references to the final cost variable are usually found under the name of "actual cost", indicating according to Flyvbjerg et al. (2002, p. 5) *"real, accounted construction costs determined at the time of project completion"*. Across studies, there is usually agreement on this definition, whereas for the estimated cost there are many different perspectives, given by the variability of the point in time of the project when this estimated cost should be considered. One of the consequences of these considerations is the formulation of different definitions for cost overruns. Merrow (2011), for instance, defines cost overruns in relative terms by mathematically dividing the actual cost of a given project with the estimated cost at the time of full fund authorisation adjusted for escalation variables.

In another study, Sloan et al. (2014), analysing the case of affordable housing projects in the UK and focusing on the issues relative to the cost estimation for these kinds of projects, defined cost overruns as the absolute discrepancy between initial estimated cost and final estimated cost. Interestingly, the academics believe that cost overruns can be calculated even before the actual end of the project thanks to the continuous whole-life costing technique, but unfortunately, the claim is substantiated only by the analysis of the project type specified above.

Further to this, Jadhav et al. (2016), in the context of a qualitative investigation on the most prominent causes of cost overruns, define at the beginning of their study, cost overruns as the amount of absolute difference between forecasted and the actual construction costs. The common feature of the last two definitions, is given by the fact that the authors consider the calculation of cost overruns in absolute terms. However, calculating cost overruns in absolute terms may be liable of creating inaccuracies in the comparison between projects pertaining to different geographical locations and time periods, undermining as well the precision in estimations of cost risks (Flyvbjerg et al. 2018).

For this reason, other researchers prefer to calculate cost overruns in relative terms, such as Locatelli et al. (2017) that define a project to exhibit cost overrun whenever the actual cost of it adjusted for inflation is at least 10% higher than the initial forecasted cost. In this case, the initial forecasted cost is considered to be the cost that is the closest in time to the first formal activity undertaken in the project.

Flyvbjerg et al. (2018, p. 175), defines cost overrun as "*the amount by which actual cost exceeds estimated cost with cost measured in local currency, constant prices, and against a consistent baseline*". This study also points out that overruns should be measured as a percentage of the forecasted cost and that if this percentage yields a positive value there is a situation of cost overrun, whereas for negative values there is a situation of cost underrun. This definition highlights a very important trait of cost overrun measurement; in fact, it says that it needs to be calculated against a consistent baseline. As mentioned earlier, this baseline (i.e., estimated cost) is at the center of an ongoing debate in the literature, as there are many opinions on which one should be selected in order to calculate cost overruns.

Hinze et al. (1991), for example, sustain that taking into consideration tender estimates would be inaccurate, as competition among bidders is likely to drive down the proposed cost to appeal more the client. In this case, by taking post-tender estimations as a base for calculation of overruns, the result would be to have an analysis on the efficiency of the winning bidder in terms of cost performance for the project. Following the same line of thought, Odeck (2004), deems the correct baseline for cost overruns calculation the level of cost declared at the detailed planning stage, given the fact that specifics of the project are mostly ruled out. Love et al. (2011; 2014) consider for the calculation of cost overrun the baseline as being the estimated cost at the contract award phase, as they believe that using estimates that are precedent to that point, is likely to unnecessarily increase the size of the cost overrun. As above, the baseline chosen by the last three described studies, constitutes a good reference point in order to calculate the performance of the contractor in the project.

Some other studies following the evolutionist perspective, finally, such as Gil and Lundrigan (2012) pose emphasis on the continuous evolution of the project from inception to delivery and deem the comparison between estimated cost and final cost as a potentially deceptive unit of analysis.

Academics pertaining to the behavioural perspective adopt a different baseline to measure cost overruns: their chosen baseline is the budget at the decision to build (Flyvbjerg et al. 2018; Cantarelli et al. 2012; Ansar et al. 2014). As a matter of fact, the use of this reference point, instead of aimed at measuring the performance of the contractors as appropriate for the studies mentioned above, wants to look at how accurate is the decision-making process during the phase when a go or no-go decision is taken for a project. With this in mind, it is straightforward to understand why any other estimated cost after the moment of the decision to build is irrelevant, as the decision to go ahead with the project is already taken when reaching the contract award stage.

All in all, there is not a right or wrong baseline for the calculation of cost overruns and, what is important, is to choose the focus of the analysis and couple it with the relative baseline. This discussion results particularly useful when I introduce in the method and analysis chapters the focus of this research and the corresponding baseline I choose. In the next section, some more details about the magnitude and frequency of cost overruns are reviewed.

3.2.3. Magnitude and Frequency of cost overruns

Many studies considered the issue of magnitude and frequency of cost overruns in the construction industry. Interestingly the size and occurrence of this phenomenon has a great degree of variability across studies, in part, also because of the effect of the baseline chosen to quantify cost overruns, as explained in the previous section.

One of the first studies undertaken in order to understand the magnitude of cost overruns is by Pickrell (1990): the study, among other things, analysed the appraisal of costs related to ten projects completed during a period of twenty years straight before the research. The report, in accounting for cost overruns, considered as a baseline the forecast at the time the decision-makers took the decision to build. The findings unveiled 90% of the projects under investigation to exhibit cost overruns from a minimum overrun of 13% to a maximum of 116%, the remaining 10% of the projects (corresponding to only one project) instead presented an underrun of -11%.

In a later paper focusing on the improvement of the project management processes, Anderson and Tucker (1994) conclude that one third of the projects, on average, do not meet expectations in terms of forecasted costs, schedule and initial objectives. This paper, even if not reporting a thorough study on a dataset and analysing it against a baseline to provide a quantification of cost overruns, was one of the first ones to acknowledge the relationship between the use of best practices and projects' performance, which set out a path for later studies in the field.

A decade later, Odeck (2004), examining cost overruns in the context of road projects in Norway, reported that projects are more likely to suffer from cost overruns than cost underruns and, in the data set of the projects analysed, the range of underruns and overruns was from -59% to +183%. The projects' database used for the research, however, not only comes from project promoted by a single public authority (Norwegian Public Roads Administration) but considers also small road projects with a minimum initial budget of 15 million NOK (≅ 1.4 Million GBP). Therefore, the external validity of the study may be threatened by the above-mentioned factors and the use of the study to infer general principles for transport mega projects as done in some other works (Love et al, 2015; 2019; Love and Ahiaga-Dagbui, 2018) may be potentially misleading. Further to this, another noteworthy finding of Odeck (2004), relating the size of the projects to the magnitude of cost overruns, showed that likelihood of cost overruns is higher on smaller rather than bigger projects. This result may be counterintuitive, as the bigger and the more complex the project, the higher the number of variables that may impact the final cost of a project, I will show, if, in other studies this finding is confirmed or not.

In another study, Olatunji (2008), reviewed 137 projects (95 construction and 42 supply projects) delivered in Nigeria: the projects unveiled a great variability when it comes about cost overruns and underruns, ranging from -91% to +101%, but again the findings supported previous research on the topic, confirming the predominance of overruns in respect to underruns. Data used in order to build the findings, are from a secondary source (i.e. meta-analysis), a choice which, some years later, will be subject of criticism from Love and Ahiaga-Dagbui (2018) when commenting research findings from Flyvbjerg et al. (2002), as being faulty of “*cherry-picking*” the data. Peter E. D. Love, in so doing, seem to disregard the work of O. Olantuji among the studies that support the evolutionist account on the calculation of cost overruns and that O. Olantuji was one of the co-authors of his paper “Understanding the landscape of overruns in transport infrastructure projects” in 2015. Notwithstanding these considerations, the study of Olantuji (2008) has been included in this literature review because of the interesting results it produced in terms of variability of cost overruns and underruns.

The research just cited, as mentioned above, is characterised by the use of secondary data as a mean of analysis of the phenomenon of cost overrun as in the case of the study from Flyvbjerg et al. (2002) in which the authors’ first conclusion is relative to the frequency of cost overruns: 9 out of 10 projects, according to them, suffer from the impact of final cost alteration versus initial budget forecasts. Similar numbers came out from past research such as the one of Lee (2008), Love et al. (2014), Pickrell (1990), Sovacool et al. (2014), Wachs (1986), however, in the case of this study the sample of projects used not only is much bigger (258 infrastructure projects) but it also comes from different geographical location (Europe, US, Japan and ten other developing countries). This strengthens previous findings in terms of external validity, as most projects from other studies came from a single geographical location. Another finding of this paper worth a mention, relates the incidence of cost overruns per type of projects: the authors conclude that all kind of large projects and not only the transport projects are likely to suffer from cost overruns, generalizing in this way the phenomenon to all types of projects, one of the assumptions of this research. This finding, therefore, is in stark contrast with Odeck (2004), stating that small projects are more prone to exhibit cost overruns.

Another research in contrast with Odeck findings, is the one of Terril et al. (2016), analysing from a consultancy secondary data source 542 completed infrastructure projects belonging to the Australian territory. According to their study, larger projects are more likely to suffer cost overruns; more precisely, they state that a 10% increase in projects size in general results in a 6% increase of risk for the project to exhibit cost overrun upon completion. This study is also in contrast with most of studies cited until now, as it reports that the majority of projects are delivered more or less in line with initial figures of budgeted cost. A closer look at the data used by the researchers, however, unveils an average cost overrun rate for the projects analysed of about 24% from the initial estimates which would not endorse the above-mentioned conclusions. Also, the authors state, in one of the notes, that their final data set did not account for 41% of the projects analysed of cost estimates forecasted before the commencement of construction, therefore, they are likely to have missed a big chunk of cost overruns in their final results, potentially justifying the mismatch of this study with the other ones considered, as reported by Cantarelli et al. (2012).

Terril et al. (2016) pose, finally, that cost underruns are rare in relation to cost overruns, in line with other studies cited and with a more recent study on the field performed by Love et al. (2019). In this paper, the authors analysed 63 projects all delivered in Hong Kong territory and reported that “only” 47% of those projects felt pray of cost overruns, presenting it as evidence against the paper of Flyvbjerg et al. (2002) stating that 9 out of 10 projects are delivered over the initially estimated cost. The paper, however, displays a number of shortcomings: they provide their analysis based on different assumptions than the study they criticise; they specify that no access was given to initial estimates at the feasibility stage of the projects, but only to approved budget figures and assume the two figures to be the same. This assumption, given the discussion on the baseline to compute cost overruns the same lead author made only a year before (Love and Ahiaga-Dagbui, 2018) and summarised in the previous section, seems to be imprecise, and may be the source of the discrepancy between the two research. Furthermore, in 2014, another paper from Love et al. was published, reporting that 95% of the 58 Australian infrastructure projects analysed suffered from cost overruns. The discrepancy of results shows that perhaps, geographical differences may come at play when analysing the incidence of cost overruns in infrastructure projects, as found by Flyvbjerg et al. (2002), therefore comparing the findings of Love et al. (2019) with Flyvbjerg et al. (2002) may be conceptually incorrect, and, as a matter of fact, one research cannot invalidate the findings of the other one.

The takeaway from all mentioned research is, finally, that cost overruns are very frequent in infrastructure projects and that their size may vary dramatically according to every situation. For this reason, the phenomenon has for long being under the attention of Public Authorities and Governments, which, in some countries, resulted in the implementation of policies to mitigate the detrimental effects of cost and schedule overruns. In the next section, I will give a closer look to some of these policies.

3.2.4. Policy landscape

When it comes to the systematic imprecision of estimations both from a time and cost points of view culminating in cost and schedule overruns, Flyvbjerg et al. (2003) specify that no major change has happened in the past 70 years.

One of the explanations given by the scholars is that the issue has always been approached with the wrong assumptions and treated as an error rather than a bias, as explained in the last sections. However, also other explanations may be found in the literature, such as the one offered by Thurairajah et al. (2018), which, through a qualitative investigation with a sample of fourteen semi-structured interviews concluded that cost and time estimations are by nature imprecise and should be interpreted as having a fluid dynamic that always changes according to new information during the project life cycle. This approach of using estimates as a “soft tool” rather than an “hard tool” shifts the importance of getting estimates right from the deterministic mentality of standard forecasting methods to more comprehensive ones able to put together the technical and human judgement aspects. As the paper further states, nowadays, the vision of estimates as hard tools is still the predominant one, however, some Governments and some Public Authorities (local and/or regional) have implemented systems and policies in order to mitigate the issue of cost overruns; the most relevant ones identified for the sake of this research are: benchmarking, external quality assurance and Optimism uplifts’ guidelines.

Benchmarking systems aim to analyse companies’ output (e.g. contractors’ performance) by comparing them between each other’s in order to use this information as a decision-making tool to award projects. By creating a scoring system, governments provide the incentive to maximise contractors’ performance in a way that the benefits are not only exploited for the project at hand but are also incubated for future projects. In this way, those contractors are encouraged to use better forecasting techniques that may have the power to reduce the impact of optimism bias (Siemiatycki, 2010). This technique has been developed as a part of front-end estimation systems in many countries and regions, such as Hong Kong, Singapore, Denmark and Ontario (Canada).

Ontario's Government, for example, developed a benchmarking system for contractors delivering public projects, assessing their performance in terms of time, cost and quality against contract specification. When the government needs to decide on the award of an upcoming project and issues an RFQ (Request for Quotation), they benchmark different contractors with a ranking system that accounts for 50% their past performance and the other half the bid of the current project. Past performance is calculated based on the average of the contractors' appraisal in the past 3 years (MTO qualification committee 2006, 2017).

This system presents many interesting features, such as the fact that when deciding which contractor to award the project to, Ontario government, not only considers the specifics of the project at hand but also the past performance of every bidder, in a way that the uncertainty related to this decision can be decreased by adopting a different view on the project at hand. Ontario Government, furthermore, is the first public authority that adopts a benchmarking system in the context of infrastructure projects (mainly transports). In other countries, such as Hong Kong and Denmark, the system is employed only for housing development projects in the former case and all projects but the infrastructure ones in the latter case. In the Danish case, benchmarking is carried out by a private company, Byggeriets Evaluerings Center, that sells evaluations to both project initiators and contractors (Siemiatycki, 2010; Tam et al. 2000).

Benchmarking systems, according to Siemiatycki (2010) even though have the opportunity to inform decision in a more complete way by adding some useful variables in the process, have not showed, alone, to notably reduce optimism bias in forecasts. Therefore, some synergies with other techniques and the incorporation of incentives should be promoted for benchmarking to yield a positive impact into the decision-making process and to the problem of cost and/or schedule overruns.

Another example of policy aimed at helping the decision-making during the front-end phase of projects is represented by the so called "State Project Model" in Norway. Firstly, introduced in 2000 under the form of an external quality assurance system focusing on cost estimation and cost control named QA2 and after five years integrated with a set of norms pertaining the choice of concept for the projects named QA1 (Samset and Volden, 2013). Given the focus of this research on cost estimation, next, I will only discuss the processes underlying QA2 system.

With the QA2 scheme, government agencies request through an enquiry to external quality assurers a proposed cost frame for the project under scrutiny, provided after an independent analysis, under the form of stochastic values (to mirror uncertainty in implementation during the project life cycle) that may or may not be considered for the final decision by the Parliament. In providing forecasts for investment costs, external quality assurers provide a cumulative probability function as represented in fig. 6, next page. The function represents, at every given level of P(X) the likelihood in percentage that the final investment cost for the project under analysis will fall at that level of below it. Usually, the Parliament takes into consideration the P85 elaborating, at the same time, a list of possible reductions to put in place during the implementation phase of the project, in case of unanticipated expenses so that the risk of exceeding the approved cost frame is decreased. However, the investment budgeted that the executing agency should target is usually at the level of P50 (the median) in a way that the agency is discouraged from the use of contingency reserves (Muller, 2016). Consequently, with this process, the recommended budget considered by the Parliament is at the P85 level whereas the targeted final cost of the project should be set at P50 level from the executing agency's side (Volden and Andersen, 2018).

As mentioned earlier, the Parliament is not obliged to follow the suggestions of the external quality assurance investigators and may opt to use different values, however, as Samset and Volden (2013) indicate, in almost all cases Parliament follow those suggestions.

QA2 main aim, therefore, is to help ensuring the operational success of the project from a time, cost and quality perspective. As a consequence, external quality assurers need to be extremely expert in all the mentioned project management areas. Moreover, to partially mitigate the behavioural biases that may arise from external quality assurers, a stochastic approach is adopted and promoted which has the potential to foster estimation practices as well.

Samset and Volden (2013) reported an analysis on the impact of the use of the QA2 scheme: they analysed 40 projects and showed that around 80% of those met the Parliament approved cost frame, proving that the scheme, if correctly implemented, has the potential to improve projects' operational success. The authors, also remark that the scheme does not add a lot of bureaucracy for the agencies and for the decision-making bodies, however, Norway, decided to make the system compulsory to use only for very big projects, around 20 per year. As Odeck (2014; 2004) found out, in Norway, the projects that have the highest cost overruns are small projects, hence, is reasonable to assume that the application of the Scheme also to smaller project would benefit their operational success as well, however, no research on this topic has been carried out while writing this.

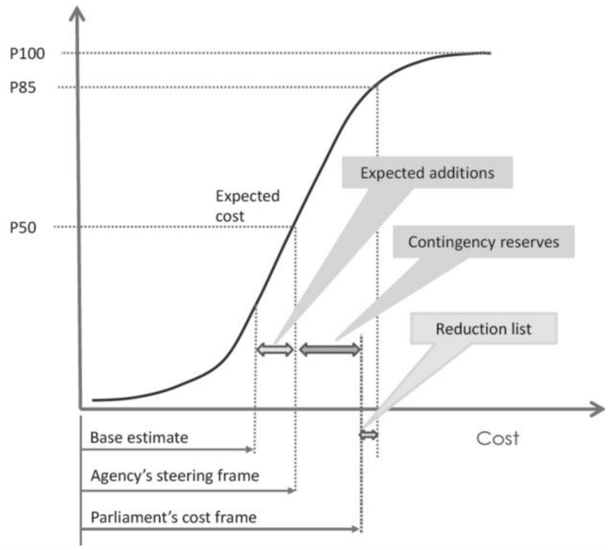


Figure 6 - Cumulative probability function of investment cost

As Volden and Samset (2017), Trafikverket (2014) and Olsson et al. (2019) report, Norway and Sweden are the only two countries in which probability-based estimation is used and regulated. Sweden introduced the system in 2012, however it is adopted only to appraise transportation projects and is less structured than the Norwegian scheme just described. All in all, those systems have showed to have the potential to improve estimation practices and mitigate some of the biases that may arise during the decision-making process, however, the use of them is restricted to either a specific type of project (Sweden) or to very big projects (Norway), therefore, a further adoption to smaller and different kind of projects should be promoted in order to better ascertain the validity of those models.

Another example is Denmark, that, when deciding on the implementation of a public project, relies on one of the most complete cost databases of past projects; however, only a basic cost estimate is considered that is increased of 10% for the agency and a 20% for the ministry under the form of a "*general supplement*" (Ministry of Transport and Building, 2015).

Unlike Norway, the UK, even if having a policy addressing the issue of cost overruns (HM Treasury, 2013), does not differentiate between budgeted and target cost, instead, an uncertainty level is chosen for every project depending on the willingness of the government to assume more or less risk of cost overrun to happen. After the choice of the level of uncertainty, the so called "*optimism bias uplift*" is applied to the estimation. The uplift is not a general supplement and of the same magnitude for every project like in Denmark but based on certain defining characteristics of the project at hand.

The *Supplementary Green Book Guidance* relative to optimism bias establishes as its aims the ones of adjusting estimates from cost, benefits, schedule and value perspectives in the first place and as a consequence helping in achieving more precise and realistic estimates. The guidance has been shaped as the result of a study undertook by Mott MacDonald (2002) on the magnitude and causes of cost and schedule overruns.

The guidance articulates in five steps how to correctly use the data provided in order to improve estimations task, providing a table that reports for each of the project types identified adjustment ranges both for Work Duration and Capital Expenditure as per Table 2 in the next page. First step is the selection of the right project type to use; given the complexity of some projects the guideline advises to be very cautious in determining the type of projects and clearly defines the characteristics of every project's typology. The second step advises on the use of the upper bound as a starting point to estimate optimism bias. After that, considerations on whether the optimism bias level could be decreased considering standard contributory factors should be done, calculated through a mitigation factor to be applied to the calculation of step two. The fourth step recommends the application of the computed *optimism bias factor* to the base estimate of the project at hand. The final step entails the review of the calculations, also from independent sources, to ascertain the strength of the estimates. Same principle is applied for the calculation of the work duration optimism bias factor.

From this process, the result is an adjustment related to the *Base Case* of the project estimates. The guidance, for this reason, advises sensitivity and stochastic analyses to be performed around the central value connected to the uncertainty value chosen before conducting the cost and/or schedule analyses. Some methods have been developed in the past years in order to support the calculation process just presented, following the outside view perspective, in the next section I will highlight the most relevant ones for the context of this research.

Project Type	Optimism Bias (%)			
	Works Duration		Capital Expenditure	
	Upper	Lower	Upper	Lower
Standard Buildings	4	1	24	2
Non-standard Buildings	39	2	51	4
Standard Civil Engineering	21	1	44	3
Non-standard Civil Engineering	25	3	66	6
Equipment/Development	54	10	200	10
Outsourcing	NA	NA	41	0

Table 2 - Optimism bias uplifts (from HM Treasury 2013)

Not a lot of research has been devoted into the understanding of the impact of optimism uplifts on the precision of estimates. However, Jennings (2012), highlights an interesting fact about the possible repercussion uplifts may have on the outcome of the project (in terms of time and cost), as it could legitimize changes during the implementation phase of the project given the “extra” budget applied to the initial forecast. In support of this idea, the concept of the “fudge factor” can be cited, pertaining to the corporate finance realm, which, as in the case of the optimism bias uplift, is used in order to decrease uncertainty related to forecast imprecisions. As Brealey et al. (2012) report, the use of a fudge factor may lead to create a distorted image of the future cash flows for a given capital investment project and may hinder some of the characteristics of this investment that may be pivotal for the go or no-go decision to invest in the project in the first place.

Building up on these considerations, during this research it will be explored, through the chosen method, whether or not optimism uplifts may present the drawback to negatively affect the final outcome of a project and/or task, so that a preliminary analysis on the effectiveness of the UK policy just described can be carried out.

3.3. Mitigating cost overruns through the outside view on forecasting with Case Based Reasoning Methods: A focus on Reference Class Forecasting techniques

In section 3, I stated that the outside view can be implemented through various methods, the most known of which is the reference class forecasting (RCF), part of the so-called case-based reasoning methods (CBR). This method, as mentioned, considers and weights the outcome of past similar projects to construct a more accurate appraisal of the project at hand (Flyvbjerg, 2008; Ji et al., 2011).

RCF activities, to summarise, can be divided in mainly three phases: during the first phase the most relevant reference class is identified. In the second phase, the probability distribution is established, so that it is possible to determine a maximum, minimum, median points and, if any, clusters of data. The last phase encompasses the placement and comparison of the project with the reference class.

Further to that, I briefly discussed some limitations of the technique, such as the hard accessibility and precision of data and the possible heterogeneity of the sample of projects in the databases used to compute the reference class given, for example, the geographical location of the different projects (Salling and Banister, 2009). However, in the last section, I showed how a number of governments believe in the advantages this technique brings to the appraisal of projects and how this resulted in the implementation of policies on the matter. RCF, therefore, represents an interesting area that bridges academic research, forecasting practices and policymaking. For this reason, other methods have been developed based on RCF, with the aim to reduce uncertainty and optimism bias during the forecasting process: the Mott MacDonald method, the Flyvbjerg method and the Salling method; following, I will address each one of them.

3.3.1. Mott MacDonald method

Mott MacDonald, in 2002, delivered a study commissioned by HM Treasury, in which 50 large public projects were analysed to understand the impact of optimism bias and investigate how to mitigate this phenomenon in the best way. A method based on RCF was developed, which is still used in the UK to measure and compute the dimension of the uplift as indicated in Table 2, in the previous page. This method can be considered as the pioneer of the application of Reference Class Forecasting to construction projects. Moreover, the consultancy, in response to the enquiry by the UK government on how to implement an outside view in forecasting for construction projects, created a database of past public projects (Mott MacDonald, 2002).

One of the novelties of this method is related to the introduction of the concept of “uplift”, as it recognises that cognitive biases have an impact on the construction projects, and they must be accounted for in the appraisal process in order to create more reliable estimates. In the previous section, I argued how the introduction of an optimism uplift might create unintended behavioural dynamics in project stakeholders, such as project managers and appraisers, however, at the time I am writing this, no data substantiating this claim exist.

In practice, in order to adopt the method in the correct way, Mott MacDonald guidelines advise to first consider a point estimate, which is usually the first estimates project appraisers come up with. Afterwards, following the uplift table, the highest percentage (or upper bound) should be used in order to calculate the maximum level of appraisal reached by applying to the base estimate the upper bound uplift. At this point, starting from this higher value this is subsequently decreased (phase 3) by analysing the way a correct procedure of risk management can be implemented considering five principal risk areas (fig. 7 and Table 3, next page).

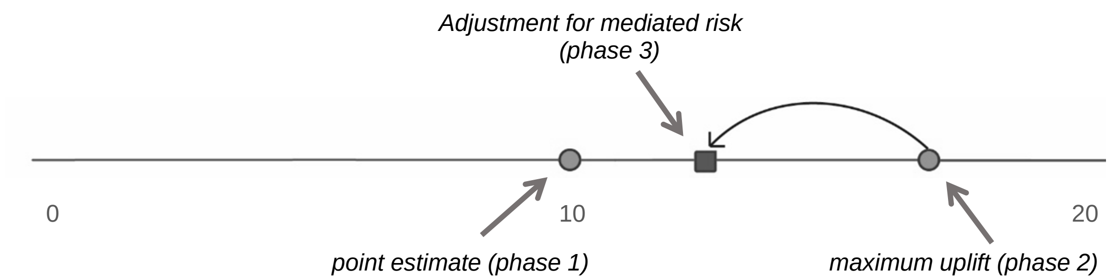


Figure 7 - Mott MacDonald method (by author)

Project Risk Areas
Procurement
Project specific
Client specific
Environment
External Influences

Table 3 - Project risk areas of Mott MacDonald method (From Mott Macdonald, 2002)

An interesting feature of this method is that it allows the forecaster to account for specific project characteristics. On the other side, the database is not very comprehensive; therefore, the sample of data may be smaller than expected and the number of projects per project characteristics slightly differ (HM Treasury, 2003). Calculation procedure for this method is depicted in the next page (fig. 8, next page): Mott MacDonald method is the predecessor of Flyvbjerg method, the calculation process is, for this reason, equivalent for both methods. As I will show later on, however, there are few important differences between Flyvbjerg's technique and the method just discussed.

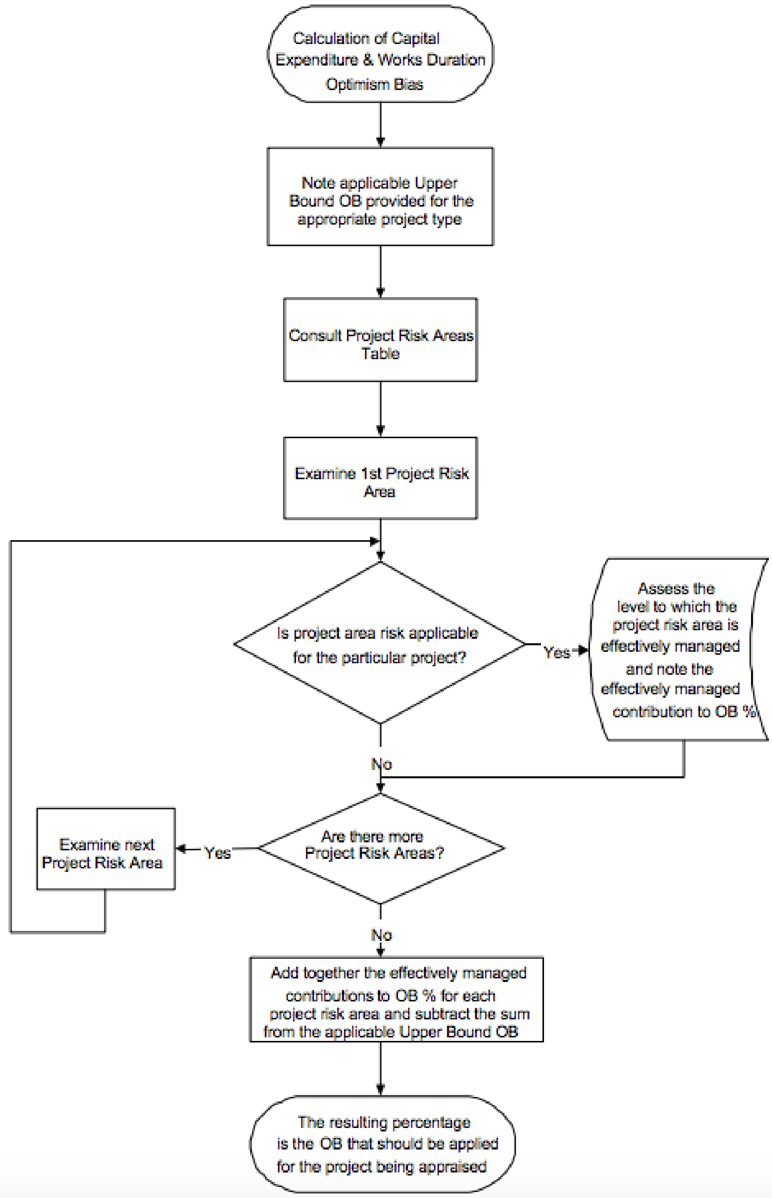


Figure 8 - Mott MacDonald and Flyvbjerg methods' calcutation procedure (Mott Macdonald, 2002)

3.3.2. The Flyvbjerg method

On the basis of the study of Mott MacDonald, Flyvbjerg (2003) and Flyvbjerg and COWI (2004c) developed a guidance document, partly referring to a different database of projects, in which the principles of Reference Class

Forecasting are set out. The objective of these documents is, moreover, to advice on the use of specifics optimism bias uplifts belonging to certain reference classes of infrastructure projects. The guidance, however, does not cover uplifts relative to work duration (as in the case of the previous method), but only the ones related to capital expenditure, as the author reported a lack of data to establish statistically significant uplifts for the former case.

The starting assumption of the method is linked with the level of uncertainty the decision makers are willing to undertake in the forecast they receive. If for example, a public authority is interested in developing a new fixed link between two cities and when asking for a forecast is willing to take a risk on the precision of the forecast, because of external pressures that may derive from different agents interested in the delivery of it, then the level of uncertainty considered in the calculation would be of 50% or 60% percentile. This means that the estimate received will correspond to the final cost of the project respectively in the 50% and 60% of the cases. On the other hand, if a project promoter would not be willing to undertake a relevant risk on the precision of the forecast, it would be more likely to choose an uncertainty level of 10-20%, or, in other words the 80-90% percentile. In general, the amount of risk it should be considered for stand-alone public projects should be of around 80% as the funding for those types of projects may be limited, considering also that the main source of those funding are taxpayers (the UK Department of Transport, for example, in most cases considers this uncertainty level). In the case of projects that are part of a bigger portfolio where a cost overrun in one might be offset by a cost underrun in another, project promoters might choose to accept a higher risk as the one given by the 50-60% percentile (Flyvbjerg, 2006). Therefore, the method suggests optimism bias uplifts related to the willingness to risk of the promoter according to the project category types as showed in table 4.

Category	Types of projects	Source of optimism bias uplifts	Applicable Optimism Bias Uplifts (% Percentile)				
			50%	**60%**	**70%**	**80%**	**90%**
Roads	Highways Trunk roads Local roads Bicycle facilities Pedestrian facilities Park and ride Bus lane schemes Guided buses on wheels	Reference class of **172** road projects contained in the Flyvbjerg database (of which 128 are British)	15%	24%	27%	32%	45%
Rail	Metro Light rail Guided buses on tracks Conventional rail High speed rail	Reference class of **46** international rail projects (of which 3 are British)	40%	45%	51%	57%	68%
Fixed links	Bridges Tunnels	Reference class of **34** international bridge and tunnel projects (of which 4 are British)	23%	26%	34%	55%	83%
Building projects	Stations Terminal buildings	Mott Macdonald - Non-standard Building Capital Expenditures	4-51%				
IT projects	IT system development	Mott Macdonald - Equipment/Development Capital Expenditures	10-200%				

Table 4 - Categorisation of projects, sources of optimism bias uplifts and optimism uplift per % percentile (Adapted from Flyvbjerg and COWI, 2004)

Uplifts are obtained by looking at the normalised distribution of cost overruns' database. Optimism bias uplifts, moreover, even if semantically alluding to optimism bias consider the total realisation costs of the past similar projects in the reference class. This means that, in a way, optimism bias uplifts encase all kind of biases and underlying costs that lead to a cost overrun, respecting the argued differentiation between cause and root cause that generate this phenomenon as per the theoretical discussion reported in the past section. For this reason, according to Flyvbjerg (2008), this method has the characteristic not only to mitigate the impact of optimism bias, but also to perform a more complete analysis grounded on past project with similar characteristics.

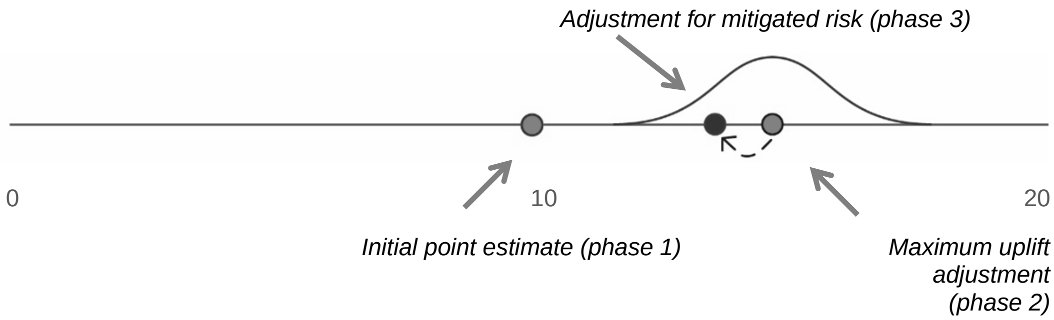


Figure 9 - Flyvbjerg method 2004 (by author)

Another case to further elaborate on how to implement this model in practice, is for example if it is adopted in the context of a tunnel project and the mean cost overrun for this category of projects is 30%, it is estimated that the probability of the project to be completed within the 130% of the expected cost is close to 50-60%. At this point, the initial point of the estimate from which the uplift will be computed is derived but, unlike the previous method, only the type of the project is analysed (tunnel, bridge, road, rail or facility) and no others specific project characteristics are considered. Afterwards, optimism uplift is applied and subsequently it is re-adjusted for mitigated risk (fig. 9): however, uplift can be reduced only if there is evidence that an effective risk management and mitigation procedure have been implemented from the pre-construction phase of the project, otherwise the initial uplift adjustment should be considered as the final estimate (Flyvbjerg, 2008).

Finally, this method is the one that the supplementary guidance on optimism bias of the Green book (2003) explicitly refers to. As a matter of fact, starting from the second half of 2004, this method is strongly recommended for adoption for all major infrastructure by the UK Department of Transport and HM Treasury (Flyvbjerg, 2006).

3.3.3. The Salling Method

This method was introduced by Salling in 2009 (Salling and Banister, 2009) and it brings together the RCF probabilistic approach discussed for the previous methods mixed with the stochastic characteristics of Quantitative Risk Analysis (QRA) and the deterministic approach provided by Cost Benefit Analysis (CBA). Given the *"hybrid"* nature of the model, it represents a good example to understand how and if the integration of the perspectives given by the outside and inside view have the opportunity to lead to an improvement of the forecasting techniques used during the project appraisal process. Before introducing the method, however, I will briefly define and discuss the main characteristics of CBA and QRA, so that a more complete overview of the method is provided.

3.3.3.1. Cost Benefit Analysis and Quantitative Risk Analysis

CBA is one of the most widely used forecasting techniques in public infrastructure projects, especially when it comes about assessing transport infrastructure projects (Grant-Muller et al., 2001; Rotaris et al., 2010). CBA translates the impact of the envisaged positive and negative features of a project in monetary terms. After that, the positive features are juxtaposed with the negative ones and formalised into present values. These present values are usually combined into a single indicator, so that a comparison between different investment opportunities may be operated; examples of those indicators are: Net Present Value (NPV), Internal Rate of Return (IRR) and Benefit-Cost Ratio (BCR) (Brent, 1996; Mouter, 2014).

In order to monetize the calculated benefits, in the analysis, three main actors are considered: "*the users*", which represent the category of actors that will be the main receivers of the benefits arising from the implementation of a project. *"The operators"*, representing those who take care of providing the service associated with the infrastructure being built (for an airport the operators would be the airlines for example) and "*the authorities"* which are the bodies the construction costs and the infrastructure is allotted to (Salling and Banister, 2009).

In many countries, such as the Netherlands, CBA is deemed to be one of the best decision-making tools, in fact, since more than a decade the technique is a compulsory requirement for all infrastructure projects funded by the Dutch national government (Annema et al., 2007). For this reason, many research efforts have been devoted into the analysis of this tool and possible ways to improve. Research outputs have reported, however, some critiques on the use of this technique: one of the biggest ones being that not all benefits can be monetised, and this means that important factors may be excluded during the decision-making process leading to potential unanticipated risks when developing the project under scrutiny (Beukers et al., 2012; Odeck and Kjerkreit, 2019). Moreover, in a recent study Flyvbjerg and Bester (2021), discussed how estimations deriving from CBA are highly inaccurate and that in order to use this tool in an effective way, corrective measures have to be put in place. Those measures are related to the implementation together with CBA of de-biasing techniques derived from behavioral science, of independent audits to check the analysis and to the formalisation of a system of incentives for the forecasters. In this sense, NPV, IRR and BCR might not be able, alone, to encapsulate the complexity of a project, and other methods are needed to support the decision-making process in the right way.

Some studies, for instance, advocate the necessity of including in the decision-making process stochastic techniques able to offer “what if” scenarios, or in other words doing a sensitivity analysis on the uncertainty inputs. Sensitivity analysis is usually performed by considering different inputs (e.g., possible benefits arising from the delivery of a project) and elaborating a model that considers different scenarios (e.g., best and worst scenario) (Salling, 2008). QRA, in this sense, is similar to a sensitivity assessment and it focuses on determining the degree of influence of certain risks. QRA provides as a result an interval of values that may help in devising a range of possible outcomes rather than single and deterministic outcomes as in the case of CBA (Salling and Banister, 2009).

3.3.3.2. The CBA-DK Method

The Salling method (or CBA-DK method), as said, comprises not only the deterministic features provided by CBA, but also the probabilistic ones RCF offers and the stochastic ones given by QRA. In practice, the calculation process is represented in the figure below (fig. 6). First step is to calculate the point estimate through CBA, after that the estimate is adjusted for the optimism bias uplift, as in the method previously described. Following, the distribution of cost overruns of the Flyvbjerg database is compared with several other types of distributions (not just the normal one but also PERT, Erlang, etc.) through quantitative risk analysis (QRA) to account for distributional dissimilarities. Afterwards, a minimum and maximum costing point for the project is defined and the right distribution function is assigned in order to run a Monte Carlo simulation to understand the most likely interval of result of the forecast.

This method combines the positive features of the first two methods described, with a more deterministic approach, which in principle could reduce the uncertainty given by the forecast providing stakeholders with a more precise answer and at the same time taking into consideration cognitive biases.

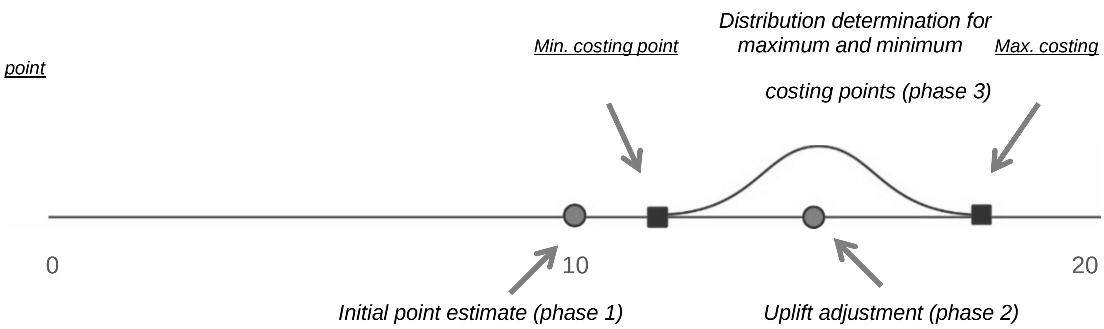


Figure 10 - CBA-DK method (by autor)

In this sense, it could be said that the method involves the use of techniques coming both from the inside and outside view, attempting to integrate them. Unfortunately, I have not found any evidence of its use in real projects except for the study of Salling for the airport and transportation links in Greenland's Capital City (Salling and Banister, 2009). One of the reasons for this could be the high amount of work that implementing the method for real projects require. Also, the method, even if involving an "objective" measurement as QRA, results in an interval of outcomes; this would be interpreted through expert judgement (as clarified by the authors) that may result, in the end, just on the choice of an interval of values that is anchored to the already calculated value from the CBA. This might be the case especially for practitioners using the method for the first time, as they could be subject to the status quo bias. In order to avoid this, a system of external assurance should be put in place and/or a process of thorough justification of the rationale used in the calculation to get to the final estimate made by the appraiser.

To further clarify the calculation flow relating to this method refer to fig. 11:

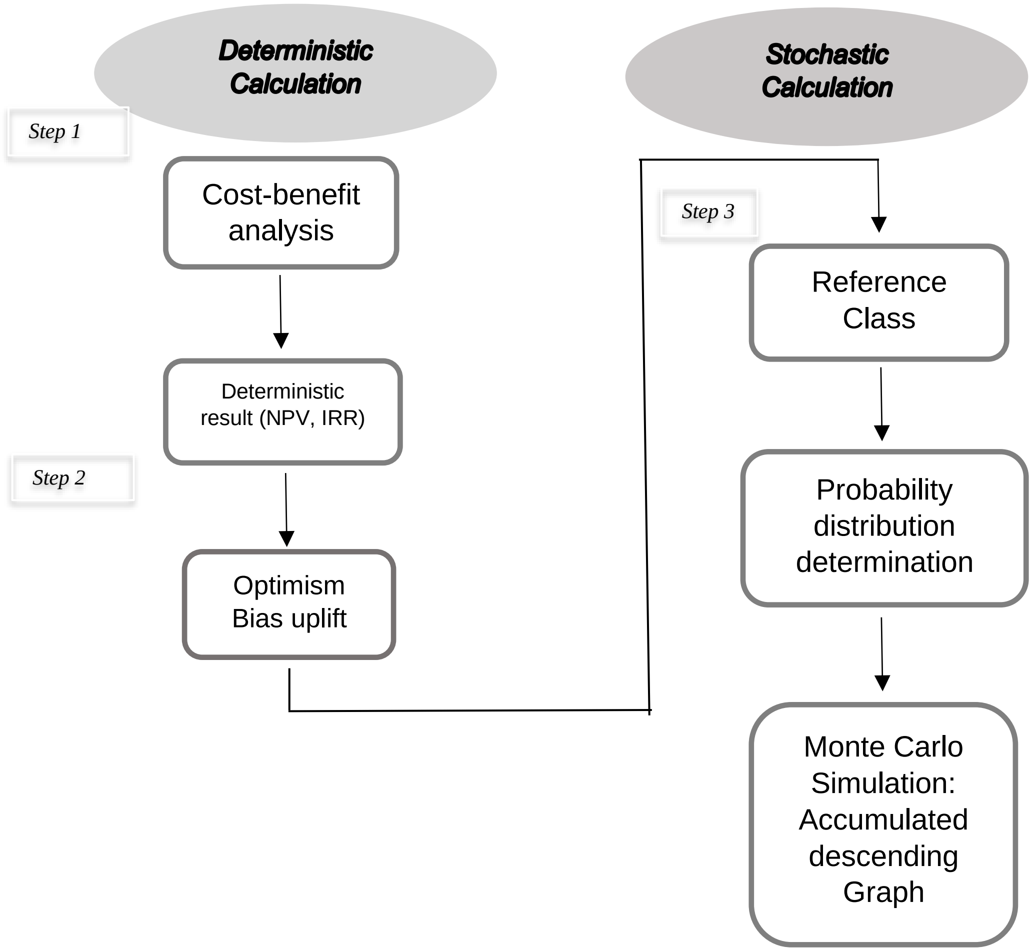


Figure 11 - Salling method calculation procedure (adapted from Salling and Banister, 2009)

3.3.4. Further Considerations

In section 3.3, I highlighted different ways in which Reference Class Forecasting can be adopted in practice by discussing three different methods: the Mott Macdonald method, the Flyvbjerg method and the Salling method. The main characteristics of the three methods are summarised in the table below:

	Mott MacDonald method	***Flyvbjerg method***	***Salling method***
Sample for database	*Study of 50 Large public procurement projects in the UK*	*1000+ projects from different countries collected through a variety of different studies*	*Uses Flyvbjerg's database as theoretical ground*
Calculation procedure	*Optimism uplift adjusted by risk mitigating factors*	*Optimism uplift adjusted by risk mitigating factors*	*Integration of uplift of risk mitigating factors with CBA and QRA*
Adopted in practice	*Yes*	*Yes*	*No*
Benefits	*- First to adopt an outside view* *- Accounts for specific project characteristics*	*- Database is the most extensive both from a geographical and project type points of view* *- Uplifts validated at an high level of statistical significance*	*- Integration of stochastic, probabilistic and deterministic approach* *- Reduces uncertainty given by the forecast through QRA* *- Systematises appraisal through CBA taking into account cognitive biases*
Drawbacks	*- Reference classes entail different categories of projects* *- Similarity function is extremely volatile given the large Reference Classes*	*- Project are not defined by specific characteristics but just type (rail, tunnel, road). Comparability may therefore, be biased and too abstract*	*- Taking as starting point CBA, re-introduces internal view at early stages of decision making. Hence, the mitigation effect and risk management procedures need to be accurate in detecting potential biases*

Table 5 - Main characteristics of the methods discussed (by author)

When it comes about the sample of project considered to build the uplifts the Mott MacDonald method used, it is straightforward to understand that the number of projects might not be deemed sufficient in order to establish effective uplifts. This is the reason why among the resulting reference classes of projects are quite heterogeneous and might result in the similarity function to be not perfectly relevant for the project at hand. The variability present in the reference class poses therefore a threat into the application of this method. Same limitation might be reported for the Flyvbjerg method (and by proxy to the Salling method), which, even if based on a much higher number of projects might still present cases in which the specification of the Reference Class makes the comparability with the project at hand too abstract.

As mentioned in the previous pages, comparability, is a fundamental issue when it comes to the adoption of an outside view on cost forecasting methods (Lovallo et al., 2012) and represents a big strength of those. Comparability issue might refer not only to the type of project included in a reference class but also to the dimensions of projects part of the Reference Class. As a matter of fact, databases include in the same reference class projects of different dimensions; this means that larger projects result to have a greater impact on the calculation of uplifts. In order to avoid this problem and standardise the weighting function related to a project, one possible way to operate, as suggested by practitioners in a previous qualitative research I made, would be to reduce every project into standard unit measures. Salling and Banister (2009) declare that one of the primary aims of their method is to find an optimal balance between the adoption of deterministic and stochastic perspectives, in order to effectively merge this method with CBA.

In order to have statistical meaning, one could argue that it might seem reasonable to reduce projects into unit measures so that every project in the database is entirely comparable with the project at hand. This reduction into unit measures would allow calculating more precise uplifts because it would not be biased by the different dimensions of projects in the database and at the same time, it would make it less abstract by providing a measure that could be customised to the project. By reducing every project into unit measures, moreover, it would be easier for costs specialists to elaborate estimations and, at the same time, the interpretation of the comparison between the project at hand and past projects would be more intuitive for all stakeholders. This would facilitate the flowing of information among project actors and would strengthen the value of calculations coming from RCF by making them more trustworthy data grounded on substantial and measurable assumptions. These issues have been taken into consideration to develop the propositions and hypotheses of this research that I will discuss in depth in the method's chapter.

Overall, even if these techniques have some limitations, they present many positive features that I highlighted in the table above and during this literature review. Taking an outside view in forecasting might be therefore considered a viable root when considering how to improve the forecasting process. Integrating it with other techniques that consider the inside view such as the one proposed in Salling and Banister (2009) and advised by the conceptual framework I introduced in the previous chapter might be, however, a more complete approach to understand how to improve the appraisal process.

3.4. Remarks and research gaps

In this literature review, I analysed the psychological process behind a decision, and how it is influenced by the context, according to the model of ecological rationality and the use of a fast decision-making tool called heuristics (Baddeley, 2013; Gigerenzer and Selten, 2002; Todd and Gigerenzer, 2012). After that, I investigated the use of heuristics, explaining why people tend to fall prey of decisional gaps, defined by the literature “cognitive biases” (Kahneman and Tversky, 1979).

These cognitive biases, when translated into a project management perspective, cause various problems, the most cited in the literature being the phenomenon of the planning fallacy and optimism bias, which in many cases leads to the underestimation of costs, eventually causing cost overruns. The literature proposes different reasons causing cost overruns; the focus of this research is the psychological reason for cost overruns, hence, I chose to adopt the lens of behavioural economics to analyse the issue, formalised in two theories: prospect and support theories.

These two theories interpret differently how to overcome planning fallacy as they refer to two different perspectives: the inside and outside view on forecasting. The two perspectives are deemed to be mutually exclusive in the project management literature, however, I investigated the possibility to cross-fertilize them given the fact that both present positive features. An example given was to integrate unpacking with Reference Class Forecasting (RCF). This led to the introduction of the conceptual framework of this research, promoting an integration of the two perspectives. I summarise benefits and limitations of the framework in table 6, next page:

Benefits	Limitations
✓ Promotes integration both at a practical and theoretical level	○ No forecasting technique adopting the two perspectives ever used in practice
✓ Identifies internal and external views as complementary perspectives	○ Difficulty to study human behaviour
✓ Helps in devising strength and weaknesses of each perspective	○ Possible resistances by academics and practitioners
✓ Provides a ground for analysis based on blending different approaches	
✓ Potential to develop innovative forecasting techniques	

Table 6 - Benefits and limitations of the proposed framework

The initial limitation described, brought to the identification of the first devised gap in literature, which is relative to the lack of studies and practical adoption of a system embedding the two perspectives and recognising that rather than different perspectives, they are two sides of the same coin. In the past, there has been a call for further research to study blended approaches also coming from different perspectives (Siemiatycki, 2009) such as the inside and outside view in order to improve the current forecasting situation, however, to my knowledge, few efforts have been devoted in this direction.

After introducing the conceptual framework, I discussed different perspectives, approaches, policies, techniques and definitions present in the literature when considering the issue of cost and schedule overruns. In order to liaise the concepts captured by the theoretical framework with their application to the infrastructure construction industry I discussed the “Hiding Hand”, evolutionist and behavioural perspectives. The behavioural perspective seems to be, in the context of this research, the account that better explains cost and schedule overruns, highlighting the differentiation between root causes and causes for the insurgence of those two phenomena.

Once this is established, cost overrun was defined as "*the amount by which actual cost exceeds estimated cost with cost measured in local currency, constant prices, and against a consistent baseline*" (Flyvbjerg et al., 2018, p.175). I explained how the choice of the baseline impacts the result of studies reporting the size and frequency of cost overruns to demonstrate that, no matter the chosen baseline, cost overruns are very frequent in infrastructure projects and that their size may vary dramatically according to every situation.

This consideration, moreover, proves why governments and public authorities have been so invested into solving the issue; to support this, I presented policies applied in seven different countries (Hong Kong, Singapore, Denmark, Sweden, Norway, UK and Canada) aimed at reducing cost and schedule overruns. These policies mainly fall into three categories: Benchmarking, External Quality Assurance and Optimism bias uplifts. At this point another gap in the literature might be identified: in the case of the Optimism Bias Uplift policy in the UK, no study has been carried out showing that uplifts are a fully positive measure and might not, for example, trigger the legitimization of the changes during the implementation phase of a project given the "extra" budget applied to the initial forecast (Jennings, 2012).

In the final section of the literature review I introduced three methods aimed at reducing optimism bias in forecasting, the Mott MacDonald Method (2002), The Flyvbjerg method (2003) and the CBA-DK method (2009). All methods have the common characteristics to use optimism uplifts in the forecasting process using Reference Class Forecasting, a method that considers past similar projects in order to build an estimate for the project at hand. The databases used to build the uplifts are generally quite old, therefore, future research should be aimed at reviewing those uplifts. Especially in the case of IT projects, given the fast-paced nature of the industry, this should be treated as a priority, so that more relevant projects can be considered in the formalisation of the uplifts and better forecasts can be achieved.

Besides these recommendations for future research, with this literature review I highlight two main gaps: the validity of optimism uplifts in improving the final outcome in terms of cost and schedule of a project, and the impact the use of a method embedding characteristics of the internal and external views might have on the final outcome of schedule and cost of a project. The aim of this research is to shed some light onto these two interconnected matters with a method able to isolate variables and establish connections between them, experiments.

CHAPTER 4. RESEARCH METHOD

4.1. Research Strategy

The starting point of my research strategy was to conduct a literature review divided into two parts. The first part of the literature review focused on developing and proposing a new conceptual framework, to be used as a ground for further future analyses and improvement of current forecasting methods that have been put in place by organisations in the construction industry (Mott MacDonald, 2002), academia (Flyvbjerg, 2004(a); Salling and Banister, 2009), policy makers and multilateral organisations (Green Book, 2013) to mitigate the impact of optimism bias. The second part of the literature review, supported by the established conceptual framework, described different perspectives, approaches, policies, techniques and definitions present in the literature when considering the issue of cost and schedule overruns.

The literature review is useful not only in order to investigate what has been written on the topic of this research but also to understand how to cross-fertilize different perspectives by offering a novel conceptual framework. In this way I can highlight research gaps that will be addressed with this dissertation or will constitute dimensions of analysis for future research, thereby creating a research agenda on the topic.

After this, I conduct four experiments in order to test the relationship between optimism bias and the variance between expected versus final cost to conclude a task, exploring the role of optimism bias in cost and schedule estimation. During these experiments I investigate firstly whether an higher level of dispositional optimism leads to an higher variance between expected versus final cost to conclude a task. After that I explore whether or not more information on historical data of similar tasks is deemed to be value adding in terms of assessing the cost of the project at hand. In doing so, I unveil further insights on current policies that are based on adding an “optimism uplift” to their forecast. The next experiment studies whether unpacking a task into subcategories without the use of any other mitigation tool, is a powerful tool in order to reduce optimism bias. Finally, the fourth experiment looks at combining the tools used in the previous two experiments (unpacking and optimism uplift) and add them to the forecast and if this has the potential to make estimation more precise than if only one method is used.

In this chapter, I explain the philosophical stance and epistemological and ontological assumptions adopted in the context of this research, after that I elaborate further on the choice of experiments as a viable methodology of inquiry for the research topic. Following, I discuss the design of the experiments with their relative power analyses. Finally, I establish research hypotheses and propositions as a result of the considerations made on the literature review and on the experiments design described.

4.2. Philosophical stance

Remenyi et al. (1998) point out that when approaching research, it is necessary to reflect on many questions: the central question that needs to be addressed, however, is “why research?”. This question entails a variety of considerations starting with the choice of the researcher to establish a personal philosophic perspective which mainly focuses on two factors: how society is perceived and how science is perceived (Burrel and Morgan, 1979). For example, society can be seen as a rationally evolving system or as a system that evolves as a result of continuous human conflict. Science, on the other hand, might be approached following either a subjective or an objective research procedure that will in turn be influenced by researcher's assumptions on reality (ontology) and what constitutes valid knowledge (epistemology) (Holden and Lynch, 2014; Babbie, 2015). This as Creswell (2009, p.6) states, will delineate a *"basic set of beliefs that guide action [of research]",* this set of beliefs are commonly known as "paradigms" (Easterby-Smith et al., 2012), that guide inquiry decisions made by the researcher (Greene and Hall, 2010.

4.2.1. Ontological and Epistemological assumptions

As can be inferred from above, the different assumptions made when establishing the framework of a research inform on the worldview of the researcher but also on the way a certain research is carried out and why a certain research approach is selected over the others (Guba and Lincoln, 1994).

One of the constitutive assumptions characterising a paradigm is the view of the researcher on the nature of reality and what are the mechanisms governing it, conceptualised in the term "ontology" (Saunders et al., 2009; Blakie and Priest, 2017). Through the ontological assumptions, the researcher communicates his or her vision of reality, generally ascribed with the dualism of subjectivism versus objectivism (Bryman, 2012).

The objectivist approach, originally deployed in the natural sciences has been subsequently adopted in the context of social sciences as well and has its roots in the realist philosophy (Pollack, 2007). Following this philosophy, in fact, proponents of objective reality recognise it as being independent from one's perspective, beliefs and set of values (Holden and Lynch, 2004). According to objectivists, therefore, the meaning is an inherent characteristic of the reality and every object in it has properties that are independent from the examiner's mind. These properties, furthermore, can be systematically studied, quantified and verified, so that the researcher main task is to build upon that meaning and explain it, allowing, in this way, to generate new knowledge (Guba and Lincoln, 1994). Objectivists, usually prefer to adopt quantitative methods aimed at studying how certain determinants and variables affect social matters (Candy, 1989).

As opposed to objectivism, subjectivism perceives reality as being the result of the influence of individuals' actions, therefore, the study of interconnections between those actions among different actors is the emphasis of this approach (Lincoln et al., 2011). As reality does not exists outside one's mind, the prerogative of the researcher following this approach is to ascertain the dynamics underpinning actors' interpretations, assumptions and beliefs when investigating social issues (Easterby-Smith et al., 2012). This perception, in most of the cases, will lead to use qualitative methods as tool for inquiry.

Ontological assumptions alone, however, are not sufficient to explain the way knowledge is captured in the context of research and what represent valid knowledge, hence another set of assumptions that encapsulate these considerations are needed: epistemological assumptions (Saunders et al., 2009). Epistemological assumptions have also the aim to analyse the relation between the researcher and what is researched.

Is considering this context that in modern times different ontological and epistemological views resulted in the formalisation of different paradigms (or philosophies), Saunders et al. (2009) describe five of them: positivism, critical realism, interpretivism, postmodernism and pragmatism. At the edges of the spectrum between subjectivity and objectivity in ontology and epistemology, there are two of the above-mentioned philosophies, positivism and interpretivism (Collis and Hussey, 2013). Following, a table reporting the fundamental characteristics of the two paradigms is showed.

Philosophical assumption	Positivism	Interpretivism
Ontological assumption	Social reality is objective and external to the researcher	Social reality is subjective and socially constructed
	There is only one reality	There are multiple realities
Epistemological assumption	Knowledge comes from objective evidence about observable and measurable phenomena	Knowledge comes from subjective evidence from participants
	The researcher is distant from phenomena under study	The researcher interacts with phenomena under study
Methodological assumption	The researcher takes a deductive approach	The researcher takes an inductive approach
	The researcher studies cause and effect, and uses a static design where categories are identified in advance	The researcher studies the topic within its context and uses an emerging design where categories are identified during the process
	Generalizations lead to prediction, explanation and understanding	Patterns and/or theories are developed for understanding
	Results are accurate and reliable through validity and reliability	Findings are accurate and reliable through verification

Table 7 - Assumptions of positivism and interpretivism (from Collis and Hussey, 2013)

In general, positivists believe in an objective reality and that in order to measure it, those methods used in the natural sciences can be used as well in those inquiries falling in the field of social sciences. On the other hand, interpretivism, recognises reality as being a socially constructed concept, dependent on the various interpretations of the different social actors (Blaikie, 2007).

The approaches suggested by interpretivism and positivism, may seem to be overly deterministic. Further to this, as McKerchar (2009) poses, in order to deliver efficient results as a researcher, the choice of the philosophical stance to adopt (and all the consequent decisions) should be dependent on the nature of the issue under analysis and flexibility to research under different set of assumptions is therefore advised. Is considering this that following, I introduce the research philosophy chosen for this study.

4.2.2. Selected Research Philosophy

For this research, considering its aim and objectives I designated pragmatism as research philosophy. Pragmatism owes its origins to the early twentieth-century philosophers William James, Charles Pierce and John Dewey. This philosophy, by recognising the importance of the practical deliverables of a study, using knowledge as an enabler, aims to conciliate objectivism and subjectivism and to achieve successful action through the process of research enquiry (Saunders et al., 2009).

In the context of social research, pragmatism is a route that has been widely explored in the past (Howe, 1988; Patton 1988); in many cases, however, this paradigm has been connected with the use of Mixed-Method Research (MMR) (Teddlie and Tashakkori, 2010; Onwuegbuzie and Johnson, 2006; Pearce, 2012). Moreover, as Morgan (2014) reports, there is a lot of literature incorrectly linking certain paradigms with specific methods, hence, a more flexible approach that disentangles research philosophies with certain research methods should be promoted. Admittedly, there are several methods leaning towards determined paradigms (as can be seen in table 7), however, there is no set rule on the matter. In this sense, pragmatism should not be considered as a philosophical stance just useful when using MMR, but as a philosophy that has the potential to support and develop good results in social research in general. This is one of the factors that lead me to the selection of this philosophy, as, this flexibility, is a valuable asset when exploring transformative and/or disruptive matters such as the one under discussion in this research. Specifically, given the aim of this research to improve current forecasting techniques, this philosophy not only calls for the use of controlled trials as being the best tool to isolate variables and show the connections between them, but also aids in emphasising the importance of reaching a practical solution for the problem under analysis.

Also, flexibility, is one of the necessary conditions for cross-fertilisation and integration to happen, and this is particularly true in the context of this research, with its starting point provided by the conceptual framework presenting the “holistic view” in forecasting. Adopting a pragmatic approach would make this process easier in practice, as the philosophy advises to use the tools that best suit the problem to be analysed and/or solved (Saunders et al., 2009).

In addition to this, pragmatism conceives research as an experience which is grounded on the opinions, values and acts of the researcher. In this sense, is very different from the positivistic account, which considers research to be value-free and independent from the researcher action (Elkjaer and Simpson, 2011). Indeed, pragmatism maintains that it is not possible for theories to be value-free, hence neutral, because in those theories there are contextual implications caused for example by personal political beliefs or external historical circumstances (Dewey, 2008[1920]). When providing recommendation on how to improve current project forecasting practices, it is, in my opinion, necessary to take this perspective, so that those recommendations are based not only on the results of the experiments but also considering the political and managerial context where the forecast decisions are taken and where certain policies have been created.

When it comes to the metaphysical assumptions adopted in the context of this philosophy, the general view is that there is no restriction towards a certain type of ontological or epistemological belief and that those issues are not at the center of pragmatism but creating knowledge usable in practice through the action of research is (Morgan, 2004). Even if at a first glance this perspective might seem naïve and not inclined towards the complexity of metaphysical reasoning embodied in the other philosophies, this is not the case, because it is possible to find in the literature clear indications of what constitutes valid knowledge for pragmatists and how they perceive reality (Elkjaer and Simpson, 2011).

Pragmatism, strives to connect through research reality, knowledge, rationality and empirical practices, recognising the dynamicity between those concepts under a phenomenological perspective so that they become relevant only when they make action relevant, which is the goal of this philosophy (Kelemen and Rumens, 2008). In this sense, epistemological and ontological considerations are not assumptions adopted in the context of the philosophy, but they become aspects embodied in the philosophy itself.

Wicks and Freeman (1998), for example, by defining the concept of *“usefulness”* as the most important epistemological characteristic of pragmatism, sustain that both knowledge created and used for researching on social phenomena (their work is especially focused on organisational studies) needs to be reliable, consistent, well founded, significant and applicable so to offer practical solutions. This element is particularly important in this research, as behavioural insights arising from optimism bias are studied so that they can be applied in the practical context of the appraisal of a project. Moreover, this research, by analysing current policy that instruct on the use of optimism uplifts in relation with the above-mentioned behavioural insights, in accordance with the pragmatist lenses, aims to offer practical solutions to improve current forecasting techniques and, consequently, policies on the matter.

Pragmatism, however, is not only invested of epistemological aspects like the one just described, but it also showcases a specific view of reality. As Shook (2000) reports in his book discussing Dewey’s theories on reality and knowledge, pragmatists see reality not as a static phenomenon, but as a mechanism that is constantly in the making, which will never be perfectly finalised. This is another consideration that leads the way this research is conceived and conducted: from one side, the recognition of reality as a continuously developmental phenomenon, makes the study of the subject of this research more interesting as it provides not only one dimension of analysis but multiples: a socio-ethical dimension, a macro dimension for policy and a micro dimension for individuals, as will be showed when describing propositions and hypotheses of this research. This could not be achieved if positivistic lenses were to be used. From the other side, it makes the researcher acknowledge that the study can only provide a small step forward in solving the issue under analysis. In this sense, throughout this research, I always used the term “mitigating” optimism bias instead of “eliminating” optimism bias, so to recognise that there is no way to completely eliminate the issue given the social, behavioural and ethical complexity of this issue.

Finally, pragmatism, as mentioned, allows to use the best method in order to deal with the problem at hand, which in this study I recognized to be experiments. In the next section, I will explain further the reasons that brought me to choose controlled trials which will be conceptually liaised with the philosophical stance adopted, the analysis performed with the literature review and the aim and objectives of this research.

4.3. Why experiments?

When it comes about researching on phenomena originated and/or subject to human behaviour (as is the case of most social sciences studies) Selten (1998) states that it is very important to gather empirical insights. They are an important feature for those seeking to analyse an issue through the lens of pragmatism. Those insights can either be obtained through field data, which present the characteristic of being hardly obtainable and interpretable, or through laboratory experiments (Scazzieri, 2003). Laboratory experiments, present many strengths: for example, unlike field data they give the possibility to observe and control variables, simplifying the process of data interpretation (Thomas, 2004).

According to Jaquemet and L'Haridon (2018, p. 94) experiments can be defined as *"a controlled situation in which many features of the environment are implemented by design, so as to observe the resulting individual decisions and interactions."* From this definition, it is possible to identify some of the main characteristics of experiments: emphasis is put on the controlled nature of the experiment, or on the fact that the variables considered in the study can be in large part manipulated given the design of the experiment (Kadzin, 1978). This characteristic addresses some of the shortcomings of field data as data resulting from experiments are in general easier to retrieve and to interpret.

Experiments, moreover, unveil the relationship between variables in a way that the phenomenon under study can be better understood. In experiments, there are usually two different types of variables: the independent variable, which represents the variable against which the outcome of the experiment is measured and the dependent variable which, as the word say, depends on the independent variable. Hence, controlled trials, in their simplest form, test the influence of a hypothesized independent variable on the dependent variable by mean of comparison between two groups, named as the experimental group and the control group. The experimental group is the group in which the independent variable under study is present (or absent, depending on the case) and the control group is the group in which the independent variable is absent (or present, depending on the case, Marczyk et al., 2005).

Besides these structural features, experiments present many other interesting characteristics, such as the one to have the capability to establish a mutual link between theory and reality. This can be achieved, considering the *"controlled"* and at the same time empirical nature that inherently characterises controlled trials (Croson and Gächter, 2010). Indeed, as in the case of theoretical models, experiments concentrate on certain factors of analysis to better understand the issue under scrutiny, with the intent to underline few effects they might have on the environment, and not emphasizing on every single aspect that might have certain consequences in the environment. Concurrently, as in the "real world", whatever occurs in laboratory might be the result of infinite uncontrollable causes and consequences (Jaquemet and L'Haridon, 2018). Is in this sense that this methodology can capture and link theory with reality.

Building up on these considerations, it is possible to identify three main purposes of controlled trials as reported by Roth (1988): firstly, they can help in testing a theory, by simulating the structure of a theoretical model and observe how certain actors behave in relation to it. Secondly, experiments may be able to substitute the analysis provided by certain theories offering contextual data on the functioning of the issue under study. Finally, controlled trials can help in directing decision-makers by supplying different insights on the functioning of certain dynamics in a determined environment.

Given the established relevance of experiments as a methodology able to provide invaluable insights to the research process, it will not come as a surprise that they are gaining more and more ground as a mean to generate knowledge, not only in the field of natural sciences but also in the one of social sciences (Webster and Sell, 2014). In social sciences, nonetheless, the advent of experiments was much slower, mainly because of the internal resistances of social scientists to adopt them in research settings. This reluctance is based primarily on the fact that some researchers believe that experiments' outputs lack generalizability, given the simplification in understanding of the reality they provide (Borgatta and Bohrnstedt, 1974; Samuelson and Nordhaus, 1985). As Falk and Heckman (2009) claim, however, those perceived limitations are originated by the poor understanding of the meaning of scientific evidence and of the nature of insights experiments provide. Interestingly, some of the detractors of this methodology (Samuelson and Nordhaus, 1992), revised their idea and acknowledged the relevance of experiments in their field of research, economics.

Experimental economics has, since then acquired a lot of recognition, until, in 2002, Daniel Kahneman and Vernon Smith received the Nobel Prize in Economics for the contribution they made in establishing experiment as a methodology for empirical economic analysis (Belyianin, 2003). Their analysis, started in the late Fifties, together with several other economists in Germany and in the UK was based on the rejection of the classical assumption of a perfectly rational "*homo oeconomicus*" and, in doing that, they brought aspects of psychological research into economics (Selten, 2003).

Here lies the first reason that led me to identify experiments as the best tool for my research. As a matter of fact, the starting point of this research was considering the psychological process behind a decision from a theoretical perspective. This lead, eventually, to the rejection of the assumptions of perfect rationality as in the above-mentioned case. By following behavioural economics constructs, I introduced the architecture of mind model according to which we take decisions. Those constructs are the result of experimental research, as seen in Kahneman and Tversky (1979) and, for this reason, pursuing this path of investigation, seemed from the very first stages of this research a relevant route, substantiated by a strong body of research.

Still, given the lack of experimental studies in the field of project management decision-making and the minimal presence of experimental studies about optimism bias in the project management context (Prater et al., 2017), basing the choice of the methodology only on the ground of theoretical strength would have been short-sighted. Hence, I started to analyse the links between the above-mentioned purposes of experiments by Roth (1988) and the objectives of my research.

My first objective is to establish and further understand the role of optimism bias in the forecasting process. This objective is linked with the purpose of "*searching for facts*", as, with the use of experiments, I can artificially create an environment in which to manipulate dispositional optimism (further details on this in the section below). The second objective is the one to bridge theoretical constructs with practical forecasting methods, which, in other words, is one of the characteristics of the experiments as described by (Croson and Gächter, 2010). Third objective is to investigate how to blend different approaches to develop innovative forecasting techniques; in order to achieve this is necessary to test different theories and practices and understand how they work together, again one of the purposes of experiments devised by Roth (1998). The fourth objective is to analyse current forecasting policy based on optimism uplifts, to understand how effective they are and if their implementation resulted in any unintended consequences. This objective is backed up by the last of the described purposes for experiments, which is the one to inform and direct decision-makers (in this case policymakers).

Further to this, the risk of participants to be deceptive in their forecasts is minimised through the controlled environment the experiment offers, as there is no self-interest or hidden interest from other stakeholders in giving misleading forecasts. This means that the possibility to identify optimism bias impact given by “delusion” (Flyvbjerg et al. 2005) is higher. Also, through the use of controlled trial it is possible to isolate the variables which are of interest for the research (optimism and expected vs final result) and attempt to manipulate them, as it will be shown during the description of the experiments. The use of experiments, furthermore, is able to generate an improved understanding on how the phenomenon of optimism bias can lead to cost underestimations and eventually to cost overruns; in fact, even if experiments can present an “oversimplified” perspective on the phenomena explored, it is able to give valuable insights on the basic processes underlying them (Weberer and Camerer, 2003). This “oversimplification”, in fact, is not necessarily a weakness of the methodology. As Plott (1991) states, when testing theories through experimentation and artificially creating an environment, it is possible to generate knowledge thanks to a principle of embeddedness of the general theory to the specific case studied in the laboratory. This means that some of the dynamics present in the general theory must pertain to their simpler projection in the experimental environment.

Finally, another reason why I chose experiments is because there are little to no studies applying it in the research context of project forecast. Hence, this research, has the potential to deliver innovative insights, contributions and solutions on the phenomenon under analysis.

4.4. Experiments Design

As mentioned above, I designed four experiments for this research: the first experiment explores whether a positive relationship exists between level of optimism and resource/schedule underestimation, by manipulating the level of optimism on subjects to observe whether this influences final estimates and to what extent. The second experiment uses data built during the first experiment (from the control group) to assess the impact of an optimism uplift on the final estimated data points and its usefulness in getting more reliable estimates. The third experiment aims to analyse the impact of unpacking a task into subcategories before performing it, in order to understand whether a method based on this practice may help in elaborating a more precise cost estimation model for projects. The fourth experiment combines unpacking with optimism uplift manipulation to appreciate the impact this would have on the final forecast. The objective of this last experiment is to analyse whether a method based on this practice might be able to yield a more accurate forecast for the task/project at hand as advocated by the theoretical framework I created for this research and discussed in the literature review.

4.4.1 The experimental platform and experimental task with instructions

The experiments should have been run in a laboratory, however, because of the insurgence of the Covid-19 pandemics, laboratory experiments that were scheduled to take place from mid-March of 2020 were cancelled. Instead, in order to cope with the new rules of social distancing and keep progressing with my research, I developed an online platform that includes a 3D building game tool. In this way, the design of experiments remained mostly unchanged as the one presented in the next sections, with the only difference being that subjects, rather than building the structure of the task with physical building bricks did that through the online building tool.

As a general task for all participants in the experiments, I asked them to build a brick structure (online) for which they had complete information on dimensions in terms of length, height and width. I provided the information about the structure through an imagine of the structure as per Figure 12. Participants were not given information regarding the correct number of bricks to complete the structure, which was of 35 bricks. However, participants were provided with a full list of bricks they have available, each with its measures, so to compare it with the measures given for the structure. Figure 13, is a screenshot of one of the experiments' web pages participants went through in order to complete the experiment, containing the user instructions for the building game they were asked to do.

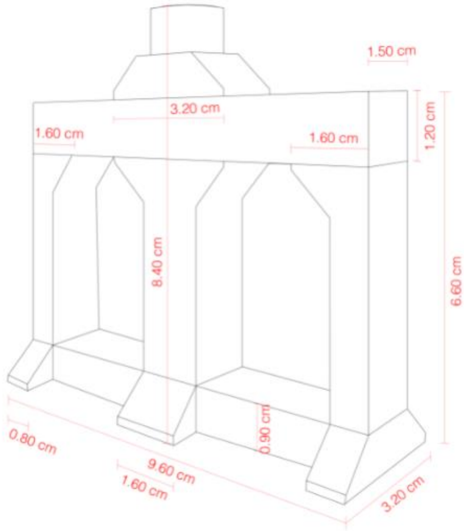
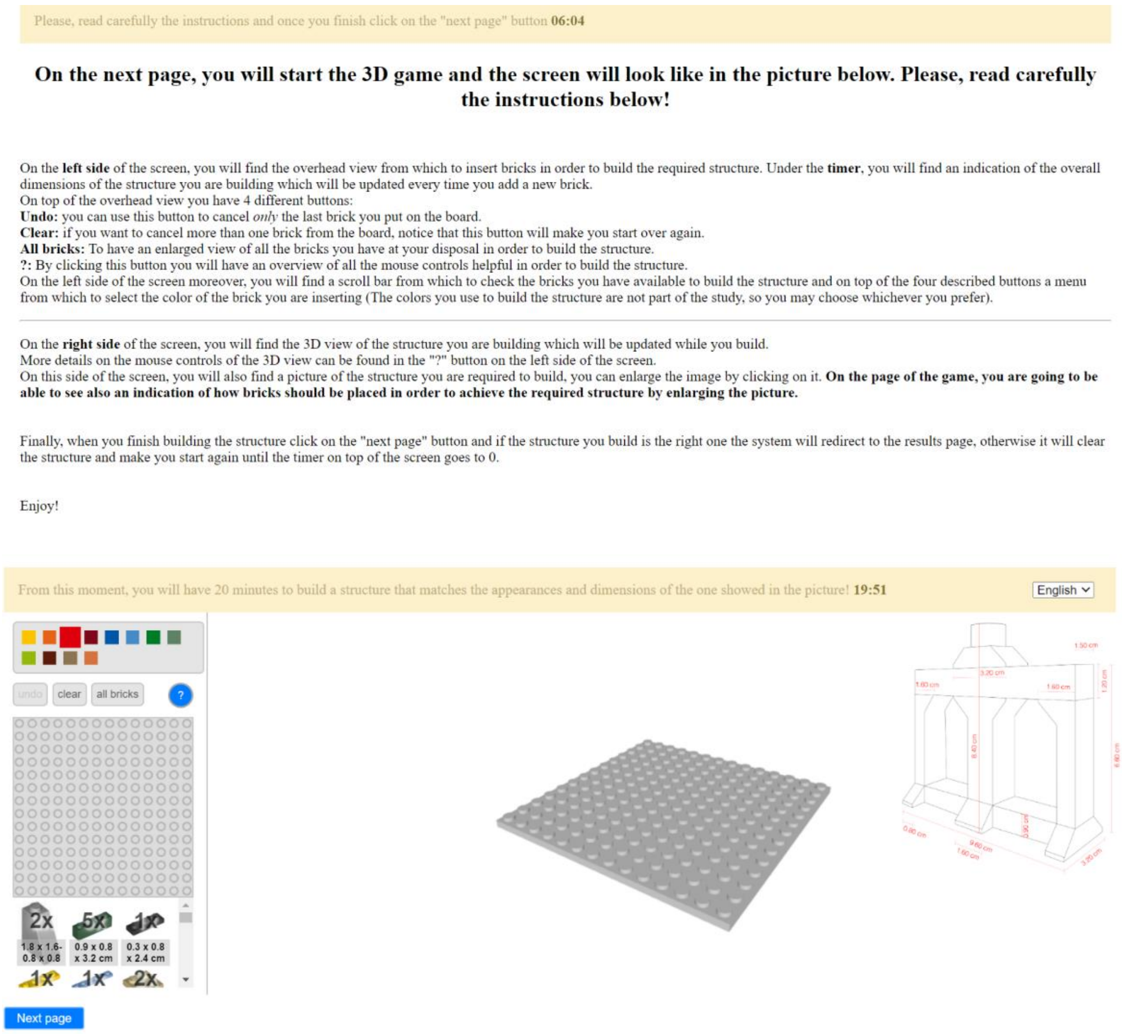


Figure 12 - Experimental task's structure with measures (by author)



Please, read carefully the instructions and once you finish click on the "next page" button **06:04**

On the next page, you will start the 3D game and the screen will look like in the picture below. Please, read carefully the instructions below!

On the **left side** of the screen, you will find the overhead view from which to insert bricks in order to build the required structure. Under the **timer**, you will find an indication of the overall dimensions of the structure you are building which will be updated every time you add a new brick.
On top of the overhead view you have 4 different buttons:
Undo: you can use this button to cancel *only* the last brick you put on the board.
Clear: if you want to cancel more than one brick from the board, notice that this button will make you start over again.
All bricks: To have an enlarged view of all the bricks you have at your disposal in order to build the structure.
?: By clicking this button you will have an overview of all the mouse controls helpful in order to build the structure.
On the left side of the screen moreover, you will find a scroll bar from which to check the bricks you have available to build the structure and on top of the four described buttons a menu from which to select the color of the brick you are inserting (The colors you use to build the structure are not part of the study, so you may choose whichever you prefer).

On the **right side** of the screen, you will find the 3D view of the structure you are building which will be updated while you build.
More details on the mouse controls of the 3D view can be found in the "?" button on the left side of the screen.
On this side of the screen, you will also find a picture of the structure you are required to build, you can enlarge the image by clicking on it. **On the page of the game, you are going to be able to see also an indication of how bricks should be placed in order to achieve the required structure by enlarging the picture.**

Finally, when you finish building the structure click on the "next page" button and if the structure you build is the right one the system will redirect to the results page, otherwise it will clear the structure and make you start again until the timer on top of the screen goes to 0.

Enjoy!

Figure 13 - Screenshot of task instructions from the online experimental tool developed by the author

Before starting with the experimental task, participants were asked to give a forecast on the number of bricks needed in order to finish the task and about the time it would take according to them to complete it.

When giving those instructions, subjects, were also informed that, if the final structure did not match the one asked to build, the system would restart from the beginning and that the bricks used until that moment would finish in the final count of the bricks used in order to complete the task. The system, moreover, allows to undo the last action performed by the subjects, in line with the thought that when errors are spotted at early stages, they are easier to solve and do not have a big impact on the overall performance of the task/project. Subjects were instructed to consider this factor when doing initial forecasts (in this way complexity and uncertainty are artificially added to the experiment). To complete the task participants had 20 minutes.

For the reader to have a complete understanding of how the experimental platform I developed works, I provide the link and password to access it, so that you can experience the experimental flow in the same way the participants to this research project did. Notice, you will be re-directed to the experimental flow of participants belonging to control groups of the last three experiments, as this is the session which is faster to complete: https://gscbe.com/en password: VIVA2022.

The link just provided shows the front-end part of the developed platform, which is what the participants of the control groups found when they were asked to complete the experiments. The back end of the platform, from where I could organize the different pages of the experiments, visualise the results of participants and program the structure to be built by participants as the experimental task, was developed not to be specific only for these experiments, but also to use it in other studies. In fact, every page is completely customisable as well as the URL, and, if needed, a large number of structures can be programmed, as well as new bricks (if not already present). In this way, different types of experiments could be run at the same time and, in case I want to do more experiments in the future, I can do that without incurring in any additional cost involved in the use of other platforms. Moreover, a plug-in to account for random (or not) allocation of prizes or incentives has been included, which was not used in the context of this research.

Figure 14 shows a screenshot of the back-end menu of the platform developed which is on top of the page whereas the bottom part of the page shows the way I visualised participants' results once they were finished with the experiment, with all information in one place (entries of the participant in the figure were erased for privacy and ethical reasons).

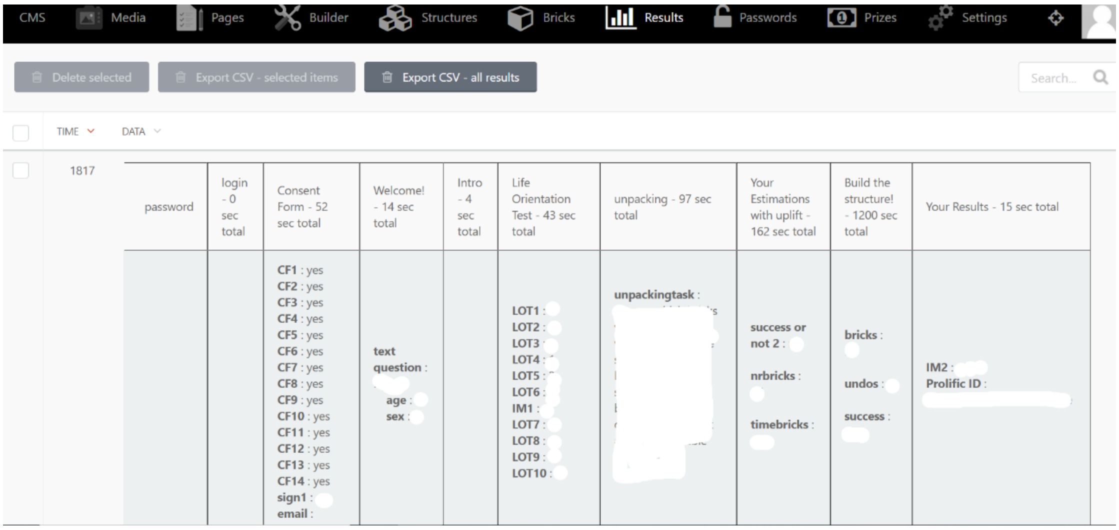


Figure 14 - Screenshot of back-end menu and results' visualisation

4.4.2. LOT-R test

Subjects were provided with a Life Orientation Test (LOT-R, as per Scheier et al. 1994) to measure their level of dispositional optimism.

LOT test was developed in the first place by Scheier & Corver (1985) and originally comprised twelve questions aimed at evaluating general expectations for positive and negative outcomes; of the twelve question eight are scored items the remaining four fillers: the eight sentences part of the scoring system are divided into two groups of four questions, one group worded in a positive manner and the other in a negative one. After an analysis carried out during 1994 by Scheier, Carver and Bridges, aimed at re-evaluating the predictive validity of the test in relation to dispositional optimism, the authors created a revised version of the test eliminating two items of the test because they were not directly related to the expectation of positive or negative outcomes and therefore impacting the predictive validity of the test. For this research, therefore, was used the LOT-R (revised) version of the test, as it has been proven capable of measuring level of dispositional optimism. Table below shows the item comprising the LOT-R test.

	Items Composing the Revised Life Orientation Test
1.	In uncertain times, I usually expect the best.
2.	It's easy for me to relax. (Filler Item)
3.	If something can go wrong for me, it will.*
4.	I'm always optimistic about my future.
5.	I enjoy my friends a lot. (Filler Item)
6.	It's important for me to keep busy. (Filler item)
7.	I hardly ever expect things to go my way.*
8.	I don't get upset too easily. (Filler item)
9.	I rarely count on good things happening to me.*
10.	Overall, I expect more good things to happen to me than bad.

Table 8 - Items composing LOT-R(Adapted from Scheier et al. 1994)

**These items are reverse scored before scoring and analyses*

4.4.3. Experiment 1

During the first experiment, participants were asked to carry out the task and to express their forecasts. Participants in the experimental group (randomly selected by the system among the whole number of participants), before conducting the task, were asked to think at the Best Possible Self in terms of ambitions for career, family, and self-realisation. Aim of this activity is to induce optimism, Best Possible Self Manipulation is in fact considered to be among the most effective techniques to induce positive future thinking and therefore temporarily alter the level of dispositional optimism (Peters et al. 2010, Peters et al. 2016, King 2001, Fosnaugh et al. 2009).

The reason why this rather than other manipulations has been selected, is because Peters et al. (2016) were able to prove that the maintenance of positive experiences coming from the BPS (even with a single session) is able to affect the performance of unrelated tasks for approximately 20 minutes: time frame highly generous to allow participants to make the initial estimates on time and resources needed to complete the task. Straight after BPS has been carried out, LOT – R test was asked to be completed by participants in order to understand the average level of dispositional optimism after the manipulation in comparison with the control group, which did not receive the manipulation. Participants from the experimental group then performed the task and when completed, the system reported back to them the actual time and the number of bricks used in order to complete the task and if they succeeded or not.

Control group has been asked to carry out the task without the manipulation; this group performed exactly the same task that, in real life, forecasters are asked to do when appraising for a project. Data coming from the control group were used to build a single reference point in terms of task completion and resources used to complete it in order to mimic optimism uplifts elaborated by the British government (HM Treasury, 2013) to mitigate the impact of optimism bias in forecasts.

4.4.4. Experiment 2

During the second experiment, statistics from the first experiment were used to create a reference class in order to build an "optimism uplift". It should be remarked, at this point, that the reason why for every experiment was selected a different group of people, is a further effort to mimic what happens when forecasters are asked to perform appraisal of projects following the optimism uplift guidelines. Optimism uplifts have, in fact, been built on the ground of past projects gathered in a database by Mott Macdonald in 2002 and in general, it is safe to assume that the stakeholders part of those projects were different from the ones that need to carry out the appraisals nowadays.

New participants from the experimental group (randomly selected from the whole group of the participants to the experiment) were, therefore, asked to carry out same task as participants of the first experiment (and to complete LOT-R test), but this time they were given data of past experiment under the form of a single point information both for resource and schedule overrun or underrun (if any). They were asked to include this number in their forecasts and subsequently to complete the task. Data of expected versus final forecasts was compared and matched with the results from the control group.

Moreover, in an effort to make the hypothesis further proved or rejected, the results were compared and matched as well by looking at the results from the control group of the first experiment following the replication principle (Babbie, 1998; Robson and McCartan, 2016). Final aim is to understand if the use of the uplift is able to give closer results between expected and final value. For this experiment, emphasis was put only on the optimism uplift per se, without considering any adjustment for mitigation or contributor factors to optimism bias (differently than what advised in the supplementary guidance for the green book) so that an analysis on how the optimism uplift on its own affects the way subjects take decisions can be carried out.

Preliminary assumption about the result of this experiment, is that optimism uplift may give more "breath" to forecasters and at later stages of the project to all relevant operational stakeholders (during the construction phase for example) resulting in an even higher cost or schedule overrun than what would have been if the uplift would have not been applied. A supporting argument to this preliminary assumption, as I discussed in the literature review, is that optimism uplifts, can be considered "fudge factors" a concept pertaining to the corporate finance literature. In general, this stream of literature considers fudge factors as not helpful in terms of forecast accuracy and do not deem them as capable of leading to a greater possibility of project success (Braeley et al. 1988). This assumption could be further strengthened if through the comparison of LOT-R test scores, level of dispositional optimism when using optimism uplift will show to be higher on average than in the first experiment.

4.4.5. Experiments 3

For the third experiment, the task subjects were asked to complete remained unchanged; this time however, experimental group, was asked to divide into subcomponents the task to be completed, following the logic of the unpacking manipulation used by Kruger and Evans (2003). The theoretical background of this manipulation does not take into account the outside view as suggested by the green book through the use of optimism uplifts. On the contrary, it considers the inside view as a mean to correctly explain the planning fallacy. In fact, the authors suggest that unpacking can give a more specific perspective of the task at hand and that forecasters, are able to gain better and more specific insights by focusing on how to divide into subcomponents the task at hand rather than just focusing on similar past projects. Unpacking manipulation is based on the support theory that I introduced in Chapter 2, by Tversky and Koehler (1994), which considers subjective probabilities as being able to give different outcomes resulting from different initial judgements. In this case, the hypothesis to be tested is, if it is true that unpacking the task gives better results in terms of forecasted precision or not so that a subsequent analysis on how to adopt this in practice can be done.

Indeed, unpacking is based on subjective probabilities, as mentioned above, so that the application of it on forecasting may give new insights on how to balance the internal and external view on project appraisal, as it considers a different perspective than the one promoted by the Reference Class

Forecasting. This experiment, therefore, aims to find out whether unpacking can be considered a good empirical background to propose an enhanced version of RCF. Once the experimental group finished with the unpacking task, they were given LOT-R to complete and were asked to make their forecast so to subsequently perform the task.

4.4.6. Experiment 4

Subjects in the experimental group of experiment 4 were asked to combine the two manipulations of the previous experiments, optimism uplift and unpacking in order to understand if the claim of the conceptual framework introduced in Chapter 2, promoting an holistic view on forecasting, has the potential to lead to more accurate forecasts. The control group of this experiment performed the same task as in the case of the previous experiments.

4.4.7. Power Analyses of the experiments

In order to have a preliminary idea of the number of participants needed for the experiments, a power analysis has been performed using the equation suggested by Rosner (2011) for sample-size calculation in the case of a two-sided alternative:

$$K = n_2/n_1 = 1$$

$$n_1 = n_2 = \frac{(\sigma^2_1 + \sigma^2_2/K)(z_{1-\alpha/2} + z_{1-\beta})^2}{\Delta^2}$$

Where

$\Delta = |\mu_2 - \mu_1|$ = absolute difference between two means

σ_1, σ_2 = standard deviation of mean #1 and #2

n_1 = sample size for group #1

n_2 = sample size for group #2

α = probability of type I error (for these calculations considered 0.05)

β = probability of type II error (for these calculations considered 0.05)

z = critical Z value for a given α or β

K = ratio of sample size for group #2 to group #1

Enrolment ratio = enrolment between the two groups (for these calculations 1)

I adapted the original equation presented in Rosner (2011) to fit the fat tail distribution (skewed towards the right) which characterizes cost/resource overruns as suggested in Flyvbjerg et al. (2018). It should be mentioned, also, that generally standard deviation is not considered an accurate measure of uncertainty for those types of distributions and the Median Absolute Deviation (MAD) should be used instead. The MAD is considered a stronger measure of dispersion whenever there are outliers in the distribution that are likely to heavily impact the resulting measure as in the case of the standard deviation where the distances from the means are squared (Taleb, 2014). However, in this specific case of power analysis, standard deviations have been adapted in order to reflect more realistically the distribution of data for cost overruns, where the incidence of overruns is higher than the incidence of underruns and the assumption of equal standard deviations in the two groups in case of a normal distribution has been relaxed (Newbold, 2013). In other words, differently from the standard equation where $\sigma_1 = \sigma_2$, for these calculations I considered $\sigma_1 \neq \sigma_2$, so that the equation above can still be used to compute sample size for the designed experiments. The analyses have been carried out considering a study group design of two independent study groups (represented by the control and the experimental group) and a continuous primary endpoint, which in this case is defined by the distributions of final count of resources used in order to complete the task both groups are asked to perform during the experiments. Anticipated means and standard deviations considered for these calculations are relative to a variable, that I will call "difference in bricks" representing the difference between estimated and actual number of bricks used in order to complete the experimental task.

Further to those considerations, some other general assumptions have been made for all the experiments: the structure participants will build during the experiments will be, as said, made from 3D representations of building bricks. It was considered, for the sake of the calculations that the optimal level of bricks used to build the structure will be of around 30 bricks and that every participant will be given a bit less than double that number of bricks. Participants, moreover, will have an image of how the final structure should look like. Finally, the anticipated mean and standard deviation of the control groups across experiments are considered to be the same as the task to be performed does not vary. However, it is likely that there will be fluctuations around the average value in the different control groups given by the randomization of the sample and the impossibility to account for all data points of the population under investigation.

4.4.7.1 Experiment 1

I made the calculations for the sample size of the first experiment, considering the general assumption described and the anticipated size of the effect. The anticipated mean for the experimental group in the Best Possible Self condition (group 1) is considered to be double than in the control group, as a result of the higher level of dispositional optimism that is expected to be present in those participants generating, following the experimental hypothesis, a big impact in terms of optimism bias exhibited by the participants. Following study parameters and sample size tables for the experiment:

Study Parameters	
Anticipated mean group 1	28
σ_1 group 1	19
σ_2 group 2	13
Anticipated mean group 2	14
α	0.05
β	0.05
Power	0.95
Enrolment ratio	1

Sample size	
Group 1	35
Group 2	35
TOTAL	**70**

Table 9a and 9b - Power experiment analysis 1

4.4.7.2 Experiment 2 and 3

For the second and third experiment, anticipated means for the experimental group is considerably less than the one in the first experiment, as the level of dispositional optimism of subjects will not influenced by the BPS manipulation. Moreover, considering the fact that the literature reports both unpacking and optimism uplift as effective tools to mitigate optimism bias and no comparative study has been done in order to ascertain whether one tool is more powerful than the other, I assumed power calculations for the two experiments to be equal. For the second experiment, the manipulation to the experimental group is represented by the adoption of an optimism uplift generated from the data of the control group of the first experiment. Participants in the experimental condition of experiment 3 will receive the unpacking manipulation, therefore, before conducting the task they will be asked to write down the different actions they are planning to do in order to conclude successfully the task. The standard deviation of the first group for both experiments (Optimism uplift and unpacking respectively) is assumed to be less volatile than in the case of the first experiment because of the use of the optimism bias mitigation tools.

Study Parameters	
Anticipated mean group 1	3
σ_1 *group 1*	9
σ_2 *group 2*	13
Anticipated mean group 2	14
α	0.05
β	0.05
Power	0.95
Enrolment ratio	1

Sample size	
Group 1	26
Group 2	26
TOTAL	**52**

Table 10a and 10b - Power analysis experiment 2 and 3

4.4.4.3. Experiment 4

For the last experiment, the anticipated mean of the experimental group (or group 1) given the unpacking and optimism uplift manipulations combined, is expected to be less than the anticipated mean of group 2 of the two previous experiments. Indeed, assuming that, as the literature reports and in line with the experimental hypotheses of this study, both unpacking and optimism uplift reduce the planning fallacy, a combined effect of the two tools should be able to reduce the difference between estimated and actual bricks to complete the task. Therefore, as the manipulation should make estimations more accurate, the volatility around the central value is expected to be lower than in the other experiments.

Study Parameters	
Anticipated mean group 1	2
σ_1 group 1	8
σ_2 group 2	13
Anticipated mean group 2	14
α	0.05
β	0.05
Power	0.95
Enrolment ratio	1

Sample size	
Group 1	21
Group 2	21
TOTAL	**42**

Table 11a and 11b - Power analysis experiment 3

4.4.7.4. Conclusion

As can be seen from the tables above, according to the power analyses performed, the total number of participants needed for all experiments is of at least 216 people. This quantity represents the minimum number of participants needed in order to yield valid and usable outcomes from the experiments. The fact that for the first experiment the sample needed is higher than the other ones, is in line with the type of manipulation used as, even if theory shows its sound validity in altering the level of dispositional optimism, it is not directly related to the tasks participants will have to perform as in the other three cases.

4.5. Further considerations on the experiments

Once all the experiments were carried out, all data were gathered and I performed an analysis on the relationship between optimism bias and cost/schedule underestimation/overestimation, as well as a validity study on the use of optimism bias uplift and possible impact it could have on the result of the project. The aim of this is to find innovative solutions of implementation of a cost forecasting model able to aid stakeholders into making more informed decisions when starting a project. Also, the experiments have the potential to show if there is a direct relationship between optimism and resource/schedule underestimation.

Fig. 15, shows two important aspects of this experimental design: first, by choosing to provide the experimental groups with LOT-R after the manipulation it is possible to cross-check whether or not the manipulations on average had an effect to alter level of dispositional optimism, factor that, together with the analysis on expected versus final time/resources used is able to further strengthen possible considerations relative to a causality between optimism and cost/schedule underestimation. Secondly, it shows that, as mentioned earlier, control groups across experiments performed exactly the same flow of activities, so that, thanks to the replication principle, results of the experiments can be further reinforced, cross-checked and analysed.

Participants were randomly selected by the online recruiting platform used in order to run the experiments: Prolific (www.prolific.co). The background of the participants is not important as no particular skills are needed in order to perform the task. In line with what Kahneman (2011) writes in his book “Thinking, fast and slow” the results from the experiment may have external validity for the category of those who work professionally as cost estimators, as it has been showed by the author that having more experience in forecasting past projects, on average does not increase the precisions of them. Interestingly, the study says that given the overconfidence coming from the fact that the forecaster may have a great amount of experience in undertaking appraisal tasks, this may generate even more discrepancies between estimated and final costs for the project at hand.

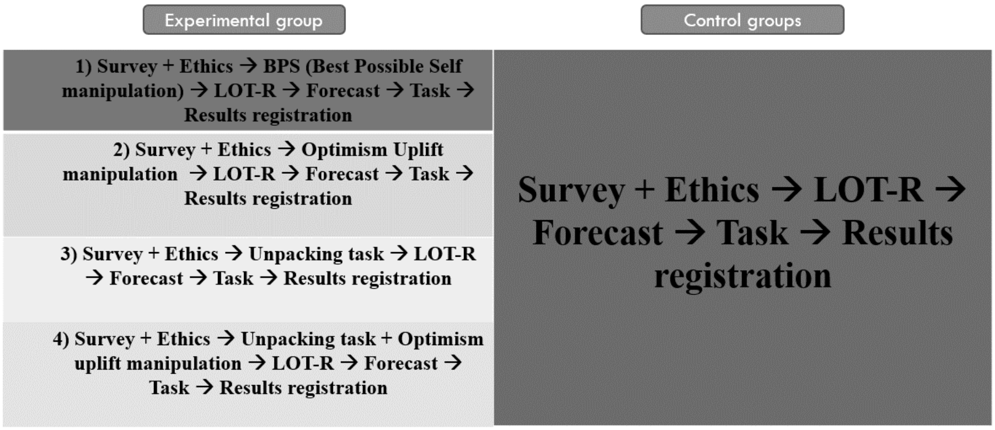


Figure 15 - Experimental Flow

Every participant taking part at any of the experiments was paid a flat rate of £5. No performance reward was awarded as at this stage it does not seem relevant for the sake of the research; performance rewards may indeed influence the participants in being more pessimistic in their forecast and playing "safe", not giving reliable and applicable results. It could, however, be the starting point for further research and experiments in order to understand how incentives influence optimism, optimism bias and forecasts accuracy.

4.6. Ethics in data collection

The experiments were carried out in compliance with UCL research ethics guidelines. All data generated from the experiments are anonymous and will be securely stored throughout the duration of the research. Every subject was informed about the possibility to withdraw from the study during the experiments. However, as data are anonymous, it is not possible to withdraw from the study after the end of the experimental session. This project ethics identification number is 15749/001, the ethics report has been submitted and approved by the Bartlett School of Sustainable Construction Ethics Committee.

4.7. Research Hypotheses and Propositions

The research hypotheses and propositions of this work can be illustrated with an inverted pyramid, whose sides are interconnected between each other's and create the structure for the three different levels of hypotheses and propositions. Indeed, there are three levels upon which the research aims to operate and for which hypotheses and propositions have been elaborated: a general level, a project management level and an infrastructure policy level.

Through the explanation of the three sides part of the pyramid in fig. 16, I will introduce the hypotheses relative to each level in order to appreciate the interconnection between them.

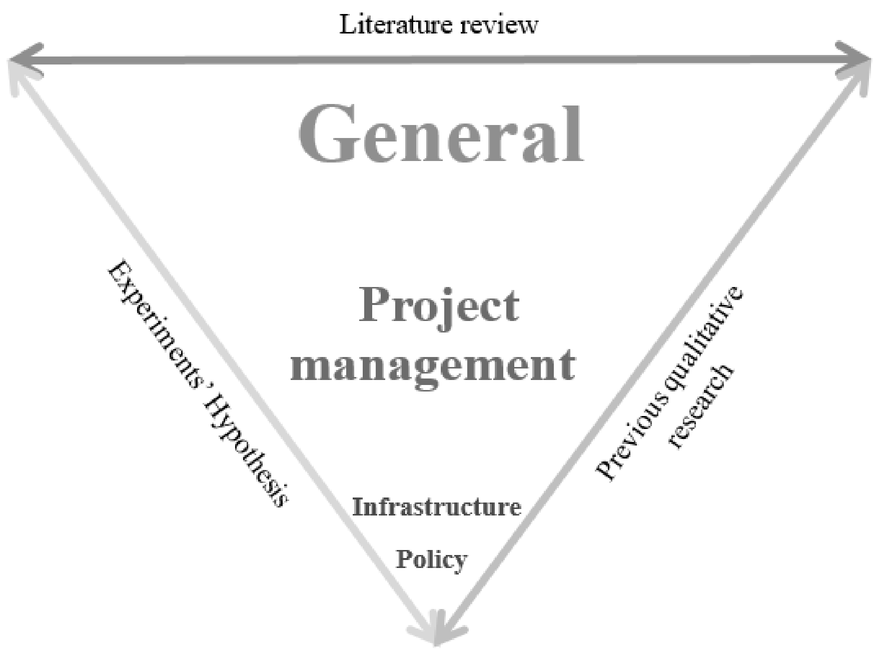


Figure 16 - Research Hypotheses and Propositions

4.7.1. Literature review

The first part of the literature review presents two theories, the support and the prospect theory. These two theories are usually perceived by the literature as an explanation of two different perspectives: the inside and outside view in forecasting. Even if some scholars advise to consider both perspectives when making an estimation, most studies made in order to tackle the issue of optimism bias only consider one of the two perspectives and never put them together. In this research, a theoretical framework that puts together the two perspectives is formalised, called the “holistic view” that, for example, proposes to merge some practices of the outside view (such as the use of Case Base Reasoning methods to get more precise estimates) with some characteristics of the inside view (unpacking the task at hand in order to get more reliable estimates). The first proposition is therefore based on the creation of this conceptual framework.

P1 → *Using a blended approach in project/task forecasting will make estimates more precise and help to better mitigate optimism bias.*

This proposition, at a general level means that when someone is prompted to make an estimation for a task, not only similar past tasks duration should be recalled but also an analysis of the various constituting parts of the task at hand should be made, so that a more reliable estimation can be provided. At a project management level, this proposition wants to look at the current forecasting techniques that are based on either the outside or the inside view and understand if through a systematic unpacking of activities at the front-end of the project together with the implementation of CBR (Case Based Reasoning) techniques estimation would be more accurate. At a construction policy level, this proposition wants to analyse why some regulations are going in one direction rather than the other and how the implementation of a policy advising the holistic view could benefit the industry.

The second part of the literature review looks at different perspectives on optimism bias from the literature (the evolutionist, the behavioural and the “hiding hand perspectives) and to the main techniques used in forecasting coming both from the inside and the outside view. The second part of the literature review not only wants to explore what are the current practices but also wants to reinforce the attributes of proposition 1 at all levels by analysing the strength and weaknesses of the different perspectives and most used techniques.

4.7.2. Experiments’ Hypotheses

The proposition coming from the literature review, directly influences the hypotheses underlying the design of the four experiments of this research.

4.7.2.1. Experiment 1

With the first experiment, the aim is understanding if there is a positive relationship between dispositional optimism level and resource/schedule underestimation.

Experimental Hypothesis 1 *A positive relationship exists between optimism levels and resources/schedule underestimation*

Experimental sub-hypothesis 1.1 *The higher the level of dispositional optimism, the higher the variance in resources/schedule estimation*

The hypotheses above, are related to the first two levels of the pyramid: indeed, at a general level and at a project management level, they indicate that the more a person is optimistic towards the task/project under appraisal the higher the inaccuracy of the estimation will be. At a construction policy level, if the hypothesis would be confirmed, it would mean that the direction policymakers are taking is correct and that policies aimed at reducing the level of optimism bias are beneficial in reducing inefficiencies coming from inaccurate project planning. From this, a second proposition can be formalised:

P2 → *Policies that address the issue of cost overruns in construction projects by considering optimism bias as one of the prominent causes of it go in the right direction.*

4.7.2.2. Experiment 2

The second proposition emphasises the issue of optimism bias, therefore, the next step is to analyse if the current policies based on the adoption of an optimism uplift based on historical data of similar projects, represents a viable option in order to get more reliable estimates. In this sense, experimental hypothesis 2 can be formalised:

Experimental Hypothesis 2 → *Including an Optimism Uplift in initial forecast/prediction will make predictions of the final time and resources used to complete a task/project more accurate in absolute value*

<u>Experimental sub-hypothesis 2.1 →</u> *Including an Optimism Uplift to a task/project forecast will, in general, increase the final count of resources and overall time to complete a task/project, in respect to a situation where optimism uplift is not included in the initial forecast for the same task/project*

These hypotheses derives from a practical thought: if the initial appraisal of a project is 2M, for example, and running the analysis on optimism uplift for similar projects it is discovered that the other projects had a final cost of 25% more, by adding this extra 25% on the initial forecast only on the basis of CBR, stakeholders may feel like they have more freedom on certain expenses as the budget available is higher. Therefore, they might end up spending more than if the optimism uplift would not have been added to the initial forecast.

The hypotheses, moreover, show the idea that at a general level, people should not only consider past similar tasks when making an estimation but also making an analysis on the task at hand (again using a holistic view), same consideration at a project management level. At a construction policy level, moreover, confirming these hypotheses would mean that current policies focusing on uplifts in order to mitigate optimism bias are not going in the right direction and even if good at emphasizing the importance that optimism bias has on cost and schedule overruns do not represent the best tool to mitigate this phenomenon.

Notwithstanding these considerations, optimism uplift may be regarded as a useful tool to get more accurate forecasts in actual value; in fact, by considering the above example, when adding the optimism uplift and having a budget of 2.5M rather than 2M the likelihood that the final cost will be closer to the value including the uplift is higher than the other way around.

P3→ *Including an optimism uplift to an initial forecast will make predictions of the final time and resources used to complete a task/project more accurate in relation to their final cost and time*

4.7.2.3. Experiment 3

Experiment 3, wants to look at the effectiveness that, unpacking a task in subcategories has in mitigating optimism uplift.

Experimental Hypothesis 3 → *Unpacking a determined task gives better result in terms of forecast precision (where forecast precision is defined as the minimisation of difference between expected value when making a forecast and actual value once the task/project is completed)*

Confirming these hypotheses, as in the case of the previous hypotheses I presented at a general and at a project management level, would mean that considering only the outside view when making forecasts might be misleading in terms of forecast accuracy. At an infrastructure policy level, it would indicate that efforts into developing these sorts of techniques and tools as well would be advisable to improve the current state of projects performance.

4.7.2.4. Experiment 4

Given Experimental Hypothesis 2 and P1, this research wants to propose some solutions that would make the use of CBR and optimism uplift more accurate. For this reason, I designed experiment 4, to test whether the use of the unpacking technique for a task combined with the application of an optimism uplift would help in mitigating planning fallacy arising from optimism bias.

<u>Experimental Hypothesis 4</u> → *Combining unpacking with optimism uplift gives better results in terms of forecast precision than using the two approaches separately*

This hypothesis, again, influences all three levels of the research. From a construction policy perspective, for example, it suggests an alternative way to look at the issue that would complement and improve the existing solution, by enriching the outside view with attributes of the inside view.

4.7.4. Previous qualitative research

The hypotheses and propositions just highlighted represent the result not only of an analysis of the literature on the topic and the design of the experiments, but also of a previous qualitative study I did as final work for my MSc. Indeed, from this previous qualitative research there are two findings that is worth mentioning for the current discussion: the first one is related to the perception of cost specialists and project managers that current models based on the outside view are too abstract and that, in order to improve the value coming from the use of these techniques, they should be more specific in the determination of the Reference class to use for the project at hand.

Another relevant issue that was brought up by practitioners, which is highly related to the one just mentioned, is the fact that, given the abstractedness of the techniques also the comparability between projects represents a risk when using RCF. Comparability is not only threatened by the abstractedness of the methods but also by the fact that in current databases in the same reference class are included projects of different dimensions. This means that larger projects result to have a greater impact on the calculations of uplifts. In order to avoid this problem and standardise the weighting function related to a project; practitioners suggested reducing every project into standard unit measures. This would mean not only unpacking the project at hand as suggested by H3 but also reducing in standard unit measure the projects in the database in order to enhance comparability. These findings reinforce the hypotheses above mentioned and are relevant especially at a project management level and at an infrastructure policy level. Findings also suggested a final proposition that could be cross-fertilised with the hypothesis relative to the efficiency of optimism uplifts use and can be formalised as:

P4 → *Optimism uplifts may be more efficient if calculated according to projects in the reference class whose weighting functions have been standardised; this would have the potential to yield a more significant value when comparing it to the unpacked project at hand.*

CHAPTER 5. EXPERIMENT 1

5.1. Introduction

In this chapter I present the results of the first experiment. I investigate the relationship between different levels of dispositional optimism and resources' underestimation or overestimation through a manipulation coming from another field of research: positive psychology. After providing more details about the manipulation and how it serves the purposes of the current research, I introduce the manipulation check and the analytical approach adopted to analyse the experiment's results.

Following, descriptive data for the experiment will be shown and variables for the analysis will be set out. I then present results in terms of statistical significance of the regression analyses of variables considered and to strengthen those, I add a section reporting robustness analyses of the models discussed.

Finally, I provide considerations regarding the theoretical implications of this experiment and how it helps to set out the scene for the other three experiments included in my dissertation.

5.2. Manipulation

During the first experiment, participants were asked to carry out the building task and to formalise their forecasts in terms of number of bricks needed in order to complete the structure and the amount of time it would have taken them to finish it. Subjects in the experimental group (randomly allocated), before conducting the task, were asked to think about the Best Possible Self in terms of ambitions for career, family, and self-realisation for one minute. After that they were given five minutes in order to write their thoughts about this, with the following instructions (adapted from King, 2001):

"Now, I will ask you to write about your best possible self for 5 minutes. ***The only rule you have about writing is that you write continuously for the entire time.*** *If you run out of things to say, just repeat what you have already written. Don't worry about grammar, spelling or sentence structure. Don't worry about erasing or crossing things out. Just write. The things you write will not be reviewed by the researcher and will be erased after this session once the system will check that you wrote something."*

The aim of this activity is to induce optimism, Best Possible Self Manipulation (BPS) is in fact considered to be among the most effective techniques to induce positive future thinking and therefore temporarily alter the level of dispositional optimism (Fosnaugh et al. 2009, Peters et al. 2010, Peters et al. 2016). The reason why this rather than other manipulations has been selected, is because Peters et al. (2016) through their experimental studies showed that the maintenance of the optimistic mood generated by the BPS (even with a single session) is able to affect the performance of unrelated tasks for approximately 20 minutes. Participants made the initial estimates on time and resources needed to complete the task immediately after completing the BPS assignment, thus well within the 20 minutes timeframe identified.

Furthermore, in recent studies, Heekerens and Eid (2021) and Carrillo et al. (2019), through their meta-analyses of thirteen different studies have confirmed how the BPS intervention has an impact on future expectations immediately after the manipulation has been administered. Moreover, Haakerens et al. (2020), with their analyses, showed how the effect of this intervention can have consequences on mood and future expectations for an entire day, de facto strengthening even more the effectiveness of BPS as an optimism induction tool. Also, another study has reported how BPS is a particularly flexible and useful intervention that can be successfully administered not only in the case of in-person studies but also in on-line ones (Loveday et al. 2018). Considering the nature of this experimental investigation, therefore, BPS seemed to be more relevant than other future thinking interventions investigated in the literature.

After the BPS intervention was administered, as a manipulation check, I asked participants to complete the LOT – R test in order to understand the average level of dispositional optimism after the manipulation in comparison with the control group, which did not receive the manipulation. Participants from the experimental group then performed the task and after this, the system reported the results of the individual session showing the actual time and number of bricks used in order to complete the task.

The control group, on the other hand, was asked to carry out the task with a similar manipulation than the one just described, however with no effect on positive future thinking and future expectations. This condition, called the Typical Day (TD), asked the participants to think about their typical day for a minute and after this, write for five minutes their thoughts. Instructions, adapted from Sheldon and Lyubormirsky (2006) were the following:

“Now, I will ask you to write about your typical day for 5 minutes. ***The only rule you have about writing is that you write continuously for the entire time.*** *If you run out of things to say, just repeat what you have already written. Don’t worry about grammar, spelling or sentence structure. Don’t worry about erasing or crossing things out. Just write. The things you write will not be reviewed by the researcher and will be erased after this session once the system will check that you wrote something.”*

Data coming from the control group were used to build a single reference point in terms of task completion time and resources used (i.e. number of bricks) to complete it in order to mimic optimism uplifts elaborated by the British government (HM Treasury, 2013) to mitigate the impact of optimism bias in forecasts. This optimism uplift was used as a manipulation for the experimental groups of two out of the three following experiments that will be described in the next sections.

5.3. Manipulation check

During the first experiment, I included a manipulation check, aimed at understanding whether the “Best Possible Self” treatment had an effect on the level of dispositional optimism exhibited by the participants under this condition. This manipulation check, administered straight after the manipulation, represented by the LOT-R test has been developed by Scheier, Carver and Bridges in 1994. The test’s main focus is the one to measure the level of dispositional optimism of those who complete it and, considering the relevance to the subject matter investigated in this experiment, it seemed the best tool to verify the strenght of the treatment and support the results deriving from it. As mentioned in the experiments' design section in Chapter 4, LOT-R comprises ten questions with four filler elements and three items that were reverse scored for the sake of this analysis (Table 12).

Given the online environment where the experiment took place, an extra question was added to the ten questions in order to screen out for random clicking with the following text: *"Please select 'strongly disagree' this is just to screen out random clicking"*. One participant did not select the right answer, therefore, results related to this subject were excluded from the current analysis.

Items Composing the Revised Life Orientation Test
1. In uncertain times, I usually expect the best.
2. It's easy for me to relax. (Filler Item)
3. If something can go wrong for me, it will.*
4. I'm always optimistic about my future.
5. I enjoy my friends a lot. (Filler Item)
6. It's important for me to keep busy. (Filler item)
7. I hardly ever expect things to go my way.*
8. I don't get upset too easily. (Filler item)
9. I rarely count on good things happening to me.*
10. Overall, I expect more good things to happen to me than bad.

**These items are reverse scored before scoring and analyses*

Table 12 - Items composing LOT-R test (Scheier et al., 1994)

Scoring system was based on a likert scale ranging from 1 to 5 and the same numerical values were given in order to score the answers, the inverse process was operated for reverse scoring questions. As per Scheier, Carver and Bridges indications scores were divided into three categories each corresponding to a level of dispositional optimism as per table 13.

How to interpret results of LOT-R: 0-18 points: *Low Optimism (High pessimism)* 19-23 points: *Moderate optimism* 24-30 points: *High optimism (Low pessimism)*

Table 13 - LOT-R scoring system (Scheier et al., 1994)

5.4. Analytical approach

For every participant in every condition the following variables were considered: the independent variable was represented by the level of optimism exhibited measured through the difference between forecasted number of bricks and actual bricks. The dependent variable were the two treatment conditions: BPS and TD. A further independent variable was created looking at the difference between the actual number of bricks and the "right" number of bricks, to have more insights on the participants' forecast precision. When studying the variables related to time, the same dependent variable was taken and a similar process for the independent variable was followed considering the independent variable as being the difference in terms of seconds between forecasted time to finish the construction task and actual time it took the participants to complete it. The three variables were studied through a linear regression analysis, creating six models, three of which adjusted for age of the participants. The analysis was performed using the R software, graphs were built using ggplot2 package and regressions summary tables using stargazer package (R Core Team, 2014; Wickham, 2016; Hlavak, 2018).

Furthermore, when looking at the descriptive statistics of the study, I also considered expectations in terms of succeeding or not to the task and actual success rate in relation to the manipulation received. This has the aim to better understand whether the success rate has been impacted by the manipulation received or not and, consequently, what was the impact of this choice when considering different levels of dispositional optimism.

5.5. Results

5.5.1. Manipulation check

As per results reported in table 14, summarising the study's descriptive data, participants in the BPS condition had a score that was on average 17.1% higher than participants in the TD condition. This indicates, therefore, that the level of dispositional optimism of participants in the BPS condition was higher than in those in the control condition, as can be also noticed from the boxplot in figure 17.

The relevance of these results is not only connected to the indication that the manipulation chosen for this study yielded results in line with what the literature argues, but it also informs on the applicability of this manipulation in different disciplines pertaining to the social sciences.

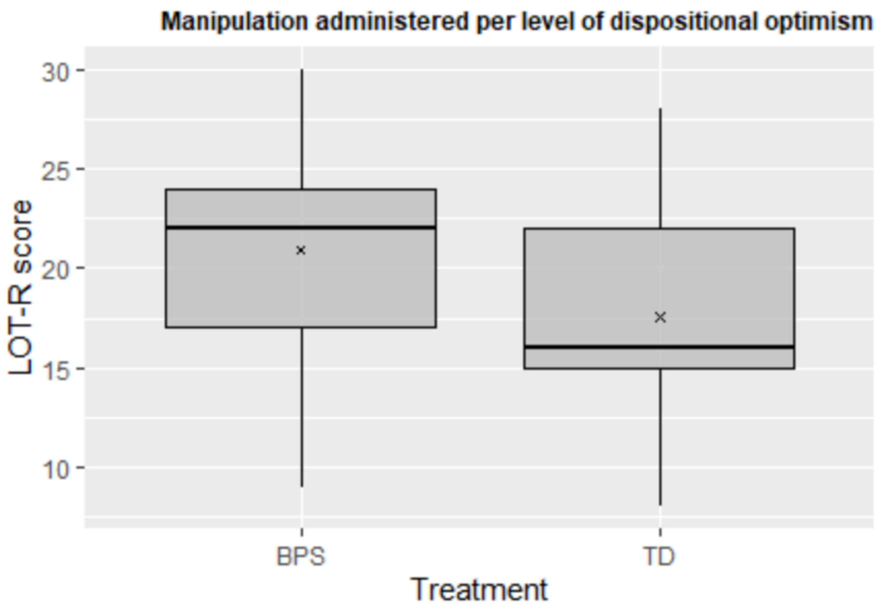


Figure 17 - Manipulation administered per level of dispositional optimism

5.5.2. Descriptive statistics

Further to the discussion on the manipulation check mentioned above, other descriptive data, reported in Table 14, were gathered to enrich the current analysis. Data presented in the table, are helpful in order to gain more insights into the issue investigated and are useful to prepare the ground for the hypothesis testing described in the next section. Besides the mean values of the variables under scrutiny, I provide the standard deviation as well, so to look at the dispersion of the data points from the mean and have a preliminary idea about the effectiveness of the manipulation administered in this experiment.

Participants in the experiment, after having received the treatment relative to the condition they were randomly allocated in, answering the questions of the LOT-R and presented the specific of the task they were asked to complete for the experiment, were prompted to provide estimations relative not only to the number of bricks and seconds it would have taken them to finish the task according to the indication given by the researcher, but also to consider whether they would have been able to conclude it or not. This information is summarised in the “success data estimate” row and, as can be seen, those in the BPS condition on average were more optimistic into estimating their likelihood of success by 16.2%. When this data is compared to the “success data actual”, reporting the actual result of participants (i.e. whether they succeeded in completing the task or not), it is interesting to notice that on average subjects on the TD condition were more successful than those in the BPS condition, supporting the claim of this research and of the literature on optimism bias that when participants’ level of optimism is higher they are not only less precise into estimating time and resources needed to complete a task but this also has an impact on the likelihood of success of completing the task at hand.

The other sets of data presented in this table are relative to the average of the estimated and actual bricks it took participants to complete the task in the allocated condition: the entries of subjects both succeeding and failing to complete the task during the given timeframe (20 minutes, more details of this in the method section) were considered. Subjects in the BPS condition estimated 10.4% fewer bricks to complete the task than the subjects in the control condition. However, they used on average 17.2% more bricks to complete it.

Experiment 1, n = 74				
General	m = 27	f = 47	average age = 29.5	
Manipulation	**BPS = 37**		**TD = 37**	
Statistic	***Mean***	***SD***	***Mean***	***SD***
Success data estimate	75.7%	-	59.5%	-
Success data actual	45.9%	-	51.3%	-
Average bricks estimate (n)	27.4	14.5	30.4	14
Actual bricks (n)	53.7	27.4	45.2	17.1
Average time estimate (s)	395	196	485	304.4
Actual time (s)	1017	272	926	339.7
Level of dispositional optimism	20.9	4.6	17.6	4.9

Table 14 - Experiment 1 descriptive statistics

For estimated and realised time for the participants to finish the task, subjects in the BPS condition were, on average, more optimistic in their estimations by 20.5% and took on average, 9.4% more time to complete it. Since failure rate of the task was rather high, the completion time is skewed towards 1200 seconds (the maximum time allowed). Participants who ran out of time are recorded has having taken 1200 seconds (20 minutes) in Table 14. This factor should be kept in mind for the interpretation of timing effects.

5.5.3. Hypothesis and sub hypothesis testing – preliminary analysis

Hypothesis and sub-hypothesis of experiment 1 are the following:

Experimental Hypothesis 1: A positive relationship exists between optimism levels and resources/schedule underestimation

Experimental sub-hypothesis 1.1: The higher the level of dispositional optimism, the higher the variance in resources/schedule estimation

In order to reject the null hypothesis relative to the hypothesis and sub-hypothesis presented above, I considered all the available variables and, taking into account the theoretical framework presented in the previous chapters of this dissertation, I built the statistical models presented in this section. With the aim to focus and simplify the analysis, I will present only three of these models at first and then, in the robustness check section, I will introduce three more models considering one confounding variable.

After having analysed the existing variables, in order to build regression models which could help in exploring the experimental hypothesis, three extra variables were created from the dataset of the experiment's results. Once the descriptive statistics showing that participants in the BPS conditions scored on average 17.1% more in the LOT-R test, the relationship between dispositional optimism levels and resources/schedule underestimation, was analysed by looking at the difference between the actual and predicted number of bricks in relation to the treatment received by participants. The same difference variable was built for the time's actual and predicted values, with unit measure in seconds.

When considering the experimental sub-hypothesis, an investigation using the above-mentioned variables would have given already some hints on whether a higher level of dispositional optimism presents a higher variance in resource/schedule estimation. However, by capitalising on the unique characteristics that experimental method offers, as the exact or "right" number of bricks needed to complete the structure is known to be 35 (check method section for more details on this), I created a further variable, considering the difference between the “right” number of bricks and the estimated bricks by participants, so to understand the magnitude of optimism bias by treatment. Unfortunately, the same cannot be done for the time variable as there is not a “right” time to finish the task, the 1200 seconds were only set as an arbitrary threshold to not prolong excessively the experiment.

Figure 18, through an histogram, shows the distribution of the created variable “difference in bricks” (actual minus estimated bricks) by treatment administered: looking at the Typical Day (TD) manipulation column, it is clear how most of the values falls in the range -20/+30, and the distribution curve is skewed to the right, as predicted by cost overrun literature (Flyvbjerg et al., 2018). BPS column, instead, presents more spread values around the average which seems also to be higher in relation to TD, both factors that would be in line with what I expressed in the experimental hypothesis and sub-hypothesis that would support rejecting the null hypothesis.

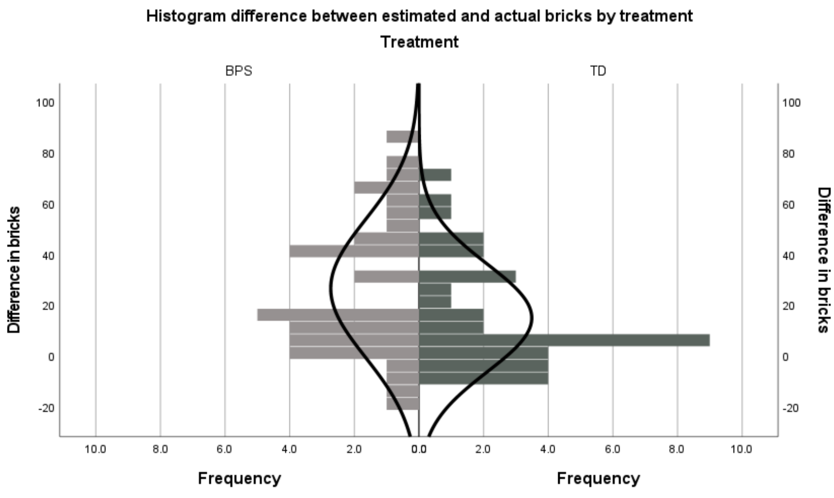


Figure 18 - Histogram difference between estimated and actual bricks by treatment

To visualise the variable created to describe the difference in seconds between the estimated and actual time it took for participants to finish (or fail) the task, Fig. 19, shows two boxplots, one per each treatment. As the failure rate of the task was quite high, values are naturally closer to the maximum amount of time the participants had at their disposal in order to finish the task (1200s). However, when looking at the range of the two boxplots, we notice straight away that the minimum difference value of the TD boxplot is considerably lower than the one of BPS (-750s vs -300s), which means that on average participants in the TD were much more precise in their estimations than subjects in BPS condition, as showed graphically with the cross in each boxplot, representing the average value of the observations.

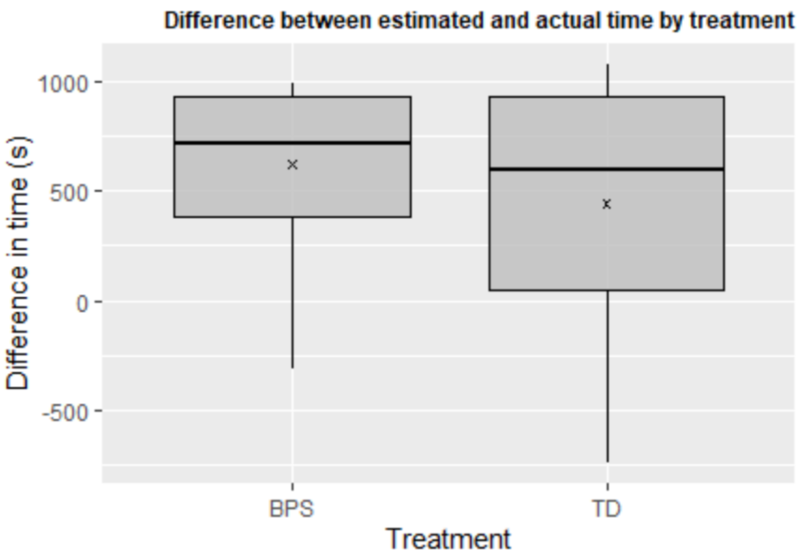


Figure 19 - Difference between estimated and actual brick by treatment

Finally, to better understand the distribution of the third variable created, representing the difference between estimated and “right” number of bricks by treatment, I created two boxplots reporting as well the average value (with the cross), in fig. 20. Whereas the participants in the TD condition had a distribution equally directed towards positive and negative values, indicating that people were on average optimistic in their estimation but with some subjects being conservative as well, it looks clear that those in the BPS condition were mostly optimistic in their estimation with very few values going over 0. This factor represents a further indication that the manipulation had an impact on the estimation bias by making participants less accurate in their estimations.

The preliminary analysis made on the variables, seems to support the underlying theory of the current research, however, the next natural step is to check if the patterns and differences devised in the two treatments are significant and, if so, what are the consequences in terms of rejection of the null hypothesis and sub hypothesis.

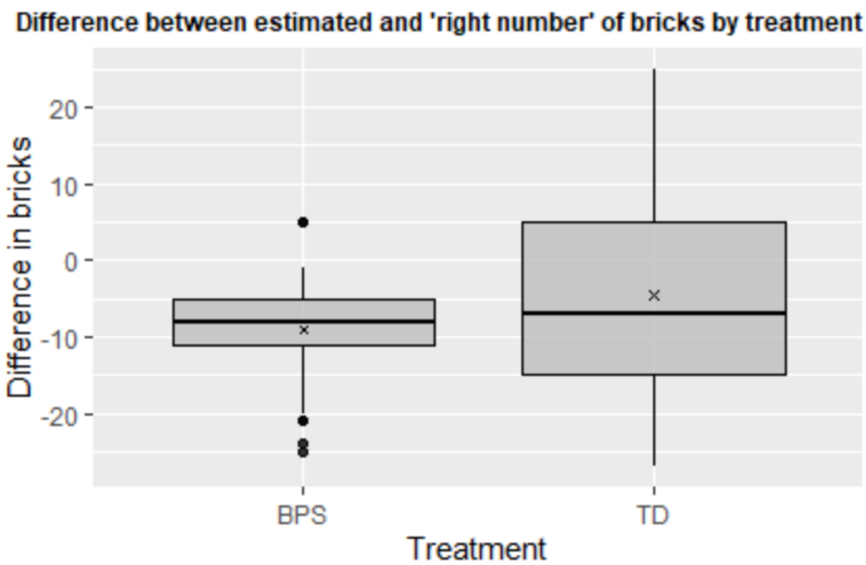


Figure 20 - Difference between estimated and “right number” of bricks by treatment

5.5.4. Hypothesis and sub hypothesis testing – regression analyses

Initially, as Table 15 shows, three simple models are examined, considering only one dependent and one independent variable: the dependent variables for each of the three models are those variables presented in the last section, whereas the independent variable is represented by the treatment received by the participants (Best Possible Self or Typical Day).

The decision to not include any other possible confounding variable at this stage of the analysis, is justified by the experimental design which aims to look at how different optimism levels impact the estimation capabilities of the subjects, net of any other factor that could alter their perception of the estimation and task in general.

Model 1 looks at the difference between estimated and actual bricks used by participants on average in relation to the treatment they were allocated to. Subjects that received the TD manipulation on average had a difference between actual and estimated value of 14.76 bricks (26.35 - 11.59) which corresponds to 56.4% bricks less than those who received the BPS manipulation. The significance level of this first linear model is 4.4%. This model supports the data observed through the descriptive statistics reported in the previous section, showing that participants in the BPS condition were on average more optimists when estimating and less accurate while performing the task. Significance is also confirmed by looking at the 95% confidence interval, which does not cross zero and as a consequence shows that the null hypothesis can be rejected when it comes to estimations in terms of resources.

Model 2, reports the regression analysis performed considering the difference between expected and actual time (in seconds) according to the manipulation participants were allocated in. Subjects that received the TD manipulation on average had a difference between actual and estimated value of 440.76 seconds (621.27 – 180.51) corresponding to 34% seconds less than those that received the BPS manipulation. The p-value of this regression model is 9.2%. In this case, the significance level does not meet the significance general threshold expectation set at 5% and looking at the confidence interval (95%) crossing the zero value, the null hypothesis of this experiment regarding the time estimations cannot be rejected.

The first two models therefore, show that when it comes to testing optimism in terms of resources and time there might be different behavioural factors and/or processes that come into play. Therefore, when studying optimism bias, it should not be assumed that the impact in terms of magnitude of optimism bias in estimations of resources and time is similar. The fact that this assumption is often incorrect should lead to the re-examination of prior research. However, the different patterns observed in the two models, might also be related to the fact that the difference in time variable, unlike the difference in bricks variable, is anchored to the maximum amount of time allowed to participants to finish the task (1200 seconds), therefore, the sensitivity of the model might be altered by this factor. This suggests that reproducing these results using a larger sample might be warranted.

	Dependent variable:		
	Diff. in bricks (n)	Diff. in time (s)	Diff. with real bricks (n)
	(1)	(2)	(3)
Typical Day	-11.59^{**}	-180.51^{*}	3.03^{*}
	(5.64)	(105.66)	(2.67)
Intercept	26.35^{***}	621.27^{***}	-7.62^{***}
	(3.99)	(74.71)	(1.90)
Typical Day 95 C.I.	(-22.85, -0.34)	(-391.14, 30.11)	(-3.58, 9.63)
Observations	74	74	74
R^2	0.06	0.04	0.05
Adjusted R^2	0.04	0.02	0.03
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 15 - Regression analysis and C.I. experiment 1

Model 3 explores the relationship between the difference of expectations in terms of bricks needed in order to complete the task and the “right” number of bricks to complete the task (35). Participants receiving the TD treatment, were on average closer to estimating the right number of bricks by 4.6 units (-7.62 + 3.03) or, in other words, 70.8% closer to the “right” number of bricks. This result, is in line with the other results seen until now, showing, once again, that BPS condition had an impact on estimation capabilities of the subjects, making them less accurate. When looking at the significance of the model, a 0.063 p-value is observed, which is not considered to be significant as the confidence interval reported crossing the zero value shows. However, considering the proximity to the significance level both in the p-value and in the confidence interval, it is possible that the sample considered for this analysis is too small, therefore more studies in this direction are recommended.

5.5.5. Hypothesis and sub hypothesis testing – robustness analysis

Even if some confounding factors are reduced in the controlled environment of the experimental setting and conditions' random assignment, there are some issues, such as the age imbalance between the two groups that might have had an impact on the observed results. This is the reason why, together with the models presented in the previous section, this section offers a few more details on the models, including a robustness check using the Mann-Whitney test and an analysis with a further independent variable, age.

Indeed, by considering the distributions of available data as non-normal, in line with what the literature reports, I performed Mann-Whitney tests for the variables under scrutiny in order to check for significance also through this method. The reason for this is that as it has been showed by Flyvbjerg et al. 2018, cost overruns, calculated by looking at the difference between estimated and actual resources used in order to complete a project, as in the case of the variables under scrutiny in this research, rarely follow a normal distribution. Indeed, they usually have a fat tailed distribution which indicates that "extreme" values are more likely to appear in the distribution in comparison to a normal distribution. In other words, overruns are more likely to happen than underruns when observing these kinds of distributions.

In this specific study, indeed, we have seen that extreme values related to over optimism in estimations are much more likely than the opposite, as can be seen also graphically with the figures showing the distributions of difference in time and difference in bricks variables. Therefore, statistically testing those variables following the assumptions of a normal distribution would not yield results useful in order to describe the phenomenon analysed in this dissertation. For this reason, instead of a t-test, I performed the Mann-Whitney test, that not only is able to account for the non-normality of the distributions under investigation but is also able to test the hypothesis indicating if there are differences between the two independent populations explored in this experiment (which would be participants in the TD and BPS condition respectively) (Newbold, 2013). If the results are consistent with the ones found through the regression analysis this would strengthen the findings presented in the previous sections, indicating significant differences between the two groups studied. As Table 16 reports, when considering the difference in bricks variable used to build model 1, the p-value is 0.034, supporting the findings of the regression analysis. When it comes about Model 2 and 3, however, the p-value is higher than what was found in the regression analysis.

Mann-Whitney tests	
	p-value
Model 1	0.034 *
Model 2	0.217
Model 3	0.237

Table 16 - Mann-Whitney tests

When considering the same models described, adding Age as a confounding factor the results appears to be different than the ones found with the first three regression models. It seems reasonable to add Age as a confounding factor, because of the online nature of the experiment designed: as a matter of fact, older participants, who might have not been as familiar as younger participants with the usability of the survey and of the 3D game, may have been more conservative in their estimations in order to allow room for errors in case of misunderstanding on how to use the building tool integrated in the 3D game. On the other hand, younger subjects might have been less conservative in their estimates because of the higher exposure to online means. Table 17, in the next page, reports the results of the 3 models presented above adding as an extra confounding variable Age.

Model 4 looks at the difference between estimated and actual bricks used by participants on average in relation to the treatment they were allocated in. The discrepancy between the subjects in the two conditions is less than the one seen in model 1, corresponding to 31.9% bricks difference on average for subjects in the TD condition. The p-value of the two independent variables is highly significant, with both values under 0.1%, and confidence intervals related to those variables do not cross the zero value, supporting the findings of the p-value. The higher significance emerging from this model, reinforces the results of model 1, by confirming that indeed there is a relationship between the variables under study. The higher significance of model 4 confirms that age might have played a role in the estimation behaviour and capabilities of subjects and, this is the reason why the model is reported in this section.

	Dependent variable:		
	Diff. in bricks (n)	Diff. in time (s)	Diff. with real bricks(n)
	(4)	(5)	(6)
Typical Day	-14.85^{***}	-202.71^{*}	6.01^{**}
	(5.54)	(108.28)	(2.67)
Age	-0.89^{***}	-6.04	0.32^{*}
	(0.36)	(6.36)	(0.17)
Intercept	54.01^{***}	810.55^{***}	-19.44^{***}
	(10.89)	(212.81)	(5.42)
Typical Day 95 C.I.	(−25.89, −3.8)	(−418.61, 13.19)	(−0.69, 11.33)
Age 95 C.I.	(−1.53, −0.23)	(−18.71, 6.63)	(−0.01,− 0.65)
Observations	74	74	74
R^2	0.15	0.05	0.10
Adjusted R^2	0.12	0.03	0.07
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 17 - Regression analyses and C.I. with Age as confounding variable

Model 5 reports the regression analysis performed considering the difference between expected and actual time (in seconds) according to the manipulation participants were allocated in, adding Age as a further confounding variable. Also in this case, the difference between the two treatment groups is lower than the one observed in model 2 at 28.6% less for participants in the TD condition. The significance level of this model is also higher than the one seen in model 2, but still not significant as confirmed by the confidence intervals both crossing zero. This model confirms that there might be a different behavioural pattern associated with different kinds of estimations (resources vs time) and supports also the assumption made for the model before that age might have had an impact on the way estimations were built by the participants. Those different behavioural patterns might be due to estimating resources or activities which, by definition are more “tangible” than the idea of “time” has an impact on how estimation decisions for these aspects are taken, however, to my knowledge, no study in this direction has ever been made.

Model 6 explores the relationship between the difference of expectations in terms of bricks needed in order to complete the task and the “right” number of bricks to complete the task (35) with Age added as a confounding variable together with the treatment type. As in the case of model 3, there is a considerable discrepancy between the two treatments: participants in the TD condition were on average 36.6% closer to the right number of bricks than those in the BPS condition. Unlike model 3, when looking at the p-value, (2.7%) the difference in bricks variable is significant, whereas the p-value for the Age variable, being very close to the 5% threshold of significance is in a grey area in terms of interpretation. However, by considering the confidence intervals of both variables it can be seen that no confidence interval crosses the zero value, therefore, it is reasonable to say that both variables of this model are significant. This model confirms the reasoning behind the structure of model 3, suggesting, once again that participants in the BPS condition were on average less accurate in their estimations.

Since the three models just presented contain more than one confounding variable, all interactions have been explored, which did not yield any significant result. Moreover, all the models were tested for collinearity using VIF (Variance inflation factor), and as can be seen in the results in table 18, no VIF is greater than 5, therefore no collinearity has been found in the models.

Collinearity	
Model	*VIF (reject if VIF > 5)*
4	*Treatment:* 1.05, *Age:* 1.05
5	*Treatment:* 1.05, *Age:* 1.05
6	*Treatment:* 1.04, *Age:* 1.04

Table 18 - Collinearity

Finally, throughout the six models presented, Adjusted R^2 scores were reported; as can be noticed, the values related to this statistic are not particularly high for all models, this is in line with the assumptions and claims of this research, as not all inaccuracies in estimation arise from excessive optimism but only this factor is taken into consideration for the current investigation.

5.5.6. Theoretical implications

From a theoretical perspective, one of the objectives of this experiment, was to explore further some of the constructs described during the first part of the literature review, where the concepts of heuristics and biases were explained. This experiment, focuses on investigating whether different levels of a specific behavioural trait, dispositional optimism, might lead us to be comparatively more positive towards the interpretation, understanding and appraising of future acts, projects or daily tasks. The literature has shown that people are naturally inclined to have an optimistic behaviour when asked to estimate about future events, which in the context of this research I defined with the concept of optimism bias. However, few investigations have been devoted to the understanding of the actual impact that different levels of dispositional optimism can have not only during the estimation phase of a project but also on its final count of resources and time.

This experiment shows that when participants exhibited a higher level of dispositional optimism not only the final count of resources and time was higher, but also the difference between estimated and actual time and resources was higher. Furthermore, data of this experiment reveal that even without manipulating the level of dispositional optimism, subjects were optimistic in their estimations when in the baseline condition (Typical Day) supporting what was found during the literature review and justifying the theoretical approach adopted in this research.

I designed this experiment not only to strenghten the relationship between the theoretical perspective I decided to adopt and my research question, but also to investigate the potential of using an experimental manipulation coming from another field of research, which in this case is the positive psychology one, in the project management field. As a matter of fact, studying the issue of cost and time overruns by adopting a behavioral perspective, showed me that in this field there is a considerable opportunity to develop the subject by capitalising on the sinergies with findings of other related subjects. However, considering the scarcely developed use of experimental methods in the project management field, those interconnections are left behind, leaving unexplored any potential research aimed at unveiling the practical impact of behavioural patterns during the front-end phase of the project.

This experiment shows, that BPS, might help researchers and practitioners to understand, for example, the impact that dispositional optimism has in any kind of estimation-based task. Indeed, to my knowledge, this is the first time that such a manipulation has been used in the cost/time overruns field of research and, more broadly, in the project management and behavioural economics field of research. BPS manipulation, in fact, could be used in other studies in the above-mentioned fields wanting to investigate estimation-based decisions in any circumstance: from personal investment decisions to executive decisions on any type of project entailing a monetary or time decision that could be affected by optimism bias. BPS could be able to show the extent to which different level of optimism might impact those decisions by providing a base from which to develop strategies to mitigate optimism bias, which is also one of the goals of the present study.

All in all, Experiment 1, represents the starting point to better understand the theory behind this research and find a further justification of this through the lenses of the experimental approach.

CHAPTER 6. EXPERIMENT 2 AND 3 RESULTS

6.1 Introduction

Experiment 1 explored if different levels of dispositional optimism have an impact during the estimation process of a task, highlighting the links between optimism bias and schedule/resources underestimation. This was done in order to find a further justification of the theoretical base of this research. I now investigate the two perspectives on project estimation described in the literature review: the internal and the external view on forecasting.

Figure 21 shows the theoretical framework created in the context of this research and highlights Phase A of it because it helps to better understand the aim of experiment 2 and experiment 3. Indeed, the aim is to investigate the strength and weaknesses of the two different perspectives on forecasting by considering one estimation tool per each view. In so doing, positive and negative aspects of both methodologies will be explored not only at a theoretical level but also at a practical level with the experiments.

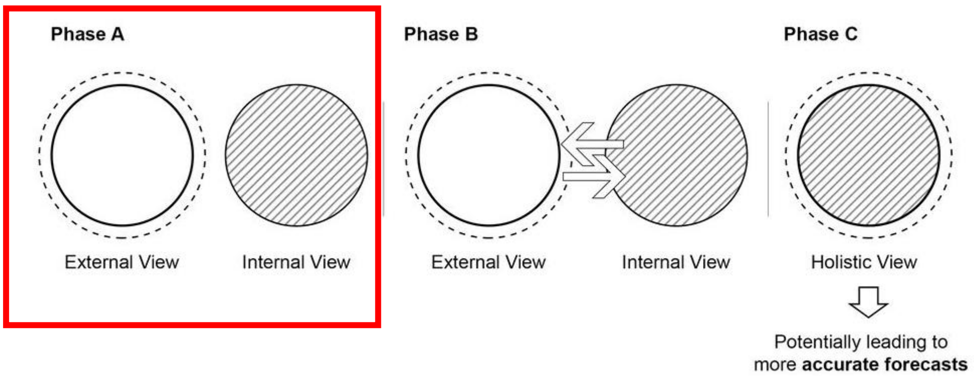


Figure 21 - Integration of the External and Internal view to create a Holistic View on forecasting (by author)

Moreover, the experiments will show which one of the tools used to mitigate optimism bias (optimism uplift and unpacking) gives the best result in terms of forecast precision. In other words, do both or any of these methods reduce the gap between expected versus actual time and/or resources and is there any of the two which is more powerful than the other?

First, I analyse the results of the two experiments separately, following the same structure and analytical approach used for experiment 1. Afterwards, in the theoretical implications section, I carry out a comparative analysis of the two experiments to better understand whether one method is better than the other, what are the reasons for this and if there are instances in which one method is better than the other. This sets the scene for the last experiment looking at Phase B of the theoretical framework (chapter 7).

6.2. Experiment 2

6.2.1. Manipulation

In the second experiment, participants were asked to perform the same building task in the 3D online tool as in the previous experiment. Subjects in the experimental condition were given the average resource (in terms of number of bricks) and time (in terms of seconds) overrun generated by participants in the control group during the first experiment, whose level of dispositional optimism was not artificially altered by a manipulation (i.e. subjects in the Typical Day condition).

Appendix 1, at the end of the chapter, reports in detail the calculations of the so-called “optimism uplifts”, that, as described in more depth in the literature review of this dissertation, is a term ascribed to the infrastructure policy field used in a great number of academic research, practitioners' reports and policy documents. Subjects in the experimental condition, were prompted to carry out their forecasting tasks with the following instructions:

> "*Observing the figure below, representing the structure you will be asked to build, please answer the following questions. Notice, you will have 5 minutes in order to answer the questions.*
>
> *Attention: As specified before, if the form and/or dimensions of the structure you build are different than the one of the figure, the system will make you start again. Once the time given in order to complete the task is over, the system will count the bricks used in every attempt to build the structure.*
>
> ***Your aim is to make estimates that are as precise as possible.*** *Moreover, I would like you to consider the following information before providing your estimations: some participants to the same experiment already did what I am asking you to do.*
>
> ***<u>Their estimations were, on average wrong by 16 bricks and 88 seconds.</u>***
>
> *Please add the additional number of bricks and seconds written above to your estimations. (e.g. if your estimation is 20 bricks then you should add 20 + 16 = 36 bricks and if your time estimation is 6 minutes you should add 6 + 1.28 = 7.28 minutes).*"

The aim of this manipulation is to explore whether providing details on historical data for the same task, in terms of the optimism bias exhibited by previous participants, makes forecasts more precise. If so, it increases the efficiency and reliability of the estimation task, as suggested in the infrastructure project management literature. Moreover, another objective of the experiment is to ascertain if using an optimism uplift has an impact on the final count of resources used. After the estimation task, participants were directed to the 3D building tool page to perform the construction task and once finished, their result in terms of bricks and time were reported.

The control group was asked to conduct the same estimation task, however, they were not given any information on the average overruns of time and resources of previous participants.

6.2.2. Analytical approach

The following variables were considered for all participants in all conditions: the dependent variable was represented by the level of optimism exhibited, measured through the difference between forecasted number of bricks and actual bricks. The independent variable were the two treatment conditions: the Optimism uplift condition (OU) and the simple estimation condition (E1). A further dependent variable was created looking at the difference between the actual number of bricks and the “right” number of bricks, to have more insights on the participants’ forecast precision. When studying the variables related to time, the same independent variable was taken and I followed the same rationale as before for the dependent variable considering it as being the difference in terms of seconds between forecasted time to finish the construction task and actual time it took the participants to complete it. The three variables were studied through a linear regression analysis, creating six models, three of which adjusted for age of the participants. The analysis was performed using the R software, graphs were built using ggplot2 package and regressions summary tables using stargazer package (R Core Team, 2014; Wickham, 2016; Hlavak, 2018).

When looking at the descriptive statistics of the study, expectations in terms of succeeding or not at the task and actual success rate in relation to manipulation received were considered to better understand whether the success rate had been impacted by the manipulation received or not.

In the data presented, one subject was excluded from the analysis as the participant failed to follow the instructions given. In addition to the two instructional manipulation checks present in every experiment, the OU manipulation by asking participants to add 16 bricks and 88 seconds made it easy to detect random or non-attentive entries. In this specific case, the participant estimated a number of bricks which was lower than 16 and for this reason they were excluded from the analysis. I decided to remove this result from the analysis in line with what the literature says about the big effects that random responses might have in the analysis (Credé, 2010). Finally, removing this entry does not create an excessive imbalance between the two groups in terms of number of entries, so it does not impair the quality of the analysis.

6.3. Results

6.3.1. Descriptive statistics

As in the case of the first experiment, Table 19 presents descriptive data. Besides the mean values of the variables under scrutiny, I provide the standard deviation as well, so to look at the dispersion of the data points from the mean and have a preliminary idea about the effectiveness of the manipulation administered in this experiment.

Participants were asked to indicate whether they thought they could complete successfully the task or not at the same time that they were asked for their estimations. Considering that the OU manipulation aimed at making the estimation more precise, our expectation is that participants in the OU condition, on average, would be more optimistic towards successfully completing the task. This is confirmed by the data, showing that, on average, around 14% more participants in the OU condition thought they would successfully complete the task in relation to the other group not receiving the manipulation. However, when we compare this to the actual success rate on the task, it can be seen that, on average, subjects in the OU conditions were more successful than their counterparts in the E1 condition.

Experiment 2, n = 53				
General	m = 35	f = 18	average age = 25.4	
Manipulation	**E1 = 27**		**OU = 26**	
Statistic	***Mean***	***SD***	***Mean***	***SD***
Success data estimate	63%	-	77%	-
Success data actual	33.3%	-	38.5%	-
Average bricks estimate (n)	23	13.6	39	11.4
Actual bricks (n)	40.3	13.2	41.6	14
Average time estimate (s)	411.5	212.8	575.2	294.9
Actual time (s)	1024.6	280.3	935.6	368

Table 19 - Descriptive statistics experiment 2

This gives an early indication of the fact that, when using an Optimism Uplift, forecasters might feel more confident towards their estimations and how those might lead to successfully complete a project. However, this is not always the case, as there are still some behavioural biases related to overconfidence that might come into play when making an estimation. This is consistent with the assumption made in this research that optimism bias cannot be eliminated from the appraisal process, but certain tools might help in the mitigation of it as, for example, the use of optimism uplifts.

Participants in the E1 condition were, on average, more conservative in their estimation by 33.2% in respect to the subjects in the OU condition and took, on average, 9.1% more time to complete the task. In terms of time, when using an optimism uplift the difference between estimated and actual time diminishes and gives an early indication of the fact that optimism uplift helps in getting more accurate estimations. As in the case of the first experiment, actual time values are skewed towards 1200 seconds as, participants who ran out of time are recorded has having taken 1200 seconds to “complete” the task, so this factor, together with the success rate on the task, should be kept in mind when analysing these variables.

Results in terms of the average bricks estimated to complete the task and actual bricks used by the participants are also reported; all entries are reported for both participants succeeding and failing in completing the task in the given timeframe of 20 minutes. Subjects in the E1 condition estimated on average 51.6% fewer bricks to complete the task than those in the OU condition. This high percentage difference is driven by the fact that the optimism uplift asked participants in the relative condition to add 16 bricks. As expected, the standard deviation of participants in the OU condition for this variable, is lower than the one of participants in the E1 condition, giving a preliminary indications that the use of an optimism uplift might be able to improve forecast precision by providing historical information of similar tasks/projects to the one at hand.

It can be observed that the number of bricks of the uplift is the same as the average difference between the two groups. Looking at these data, therefore, it might seem that adding an optimism uplift does not influence the initial level of resources participants in the OU thought they would have used in order to complete the structure. In this sense, the impact of the optimism uplift seems to be merely mechanic rather than aimed at influencing the initial state of mind of the estimator. It could be argued, in fact, that individuals who knew their estimate would get a correction at the end would adapt their behaviour and make initial estimates that were lower, but this is not the case. This might suggest that optimism uplift, unlike the manipulation that was used in the previous experiment, which was specifically aimed at altering the level of dispositional optimism before making the estimate, does not have an impact on the state of mind of the estimator, but rather, it systematically adjusts it. The fact that optimism uplift seems to have a structural impact on the quality of estimations rather than on the estimator, is a preliminary indication of how powerful this tool can be in order to mitigate optimism bias.

Furthermore, since the “right” number of bricks to complete the construction task without any mistake is 35, one can see that subjects in the E1 condition were quite optimistic when estimating the number of bricks, estimating, on average, 41.4% less bricks than what was actually needed to complete the task. Participants in the OU condition however, by estimating on average 10.8% more bricks than what needed to complete successfully the task without mistakes not only were much closer to the “right” number of bricks but also allowed in their estimation some extra resources to be used in case of errors during the task.

When it comes to the actual number of bricks used, on the other hand, subjects in the E1 condition used on average 3.2% less bricks than those in the OU condition, as can be seen graphically in the boxplot in figure 22, reporting the distribution of the actual number of bricks used per manipulation received. This factor shows that even if optimism uplift might help to get more accurate estimations it might also lead to use more resources than needed in order to complete the task.

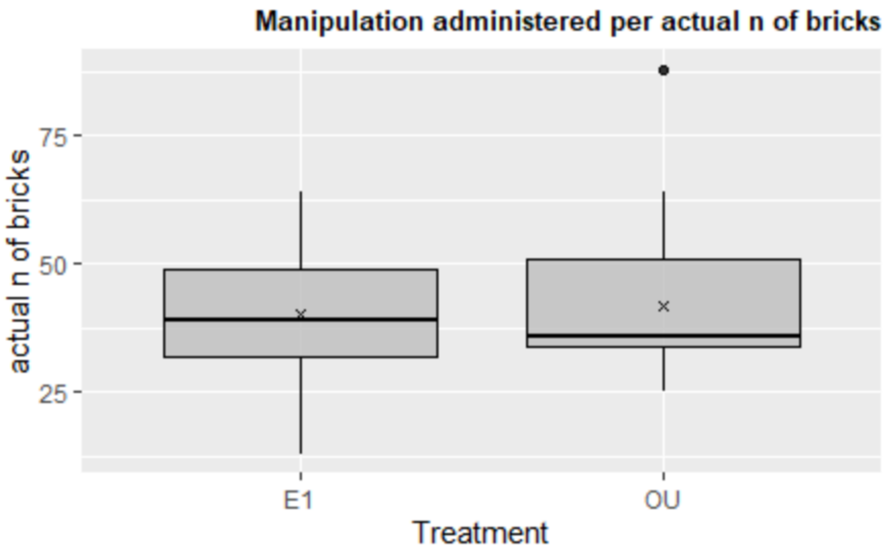


Figure 22 - Manipulation administered per actual number of bricks

6.3.2. Hypothesis and sub hypothesis testing – preliminary analysis.

Hypothesis and sub hypothesis for this experiment are the following:

> *Experimental Hypothesis 1:* Including Optimism Uplift in initial forecast/prediction will make predictions of the final time and resources used to complete a task/project more accurate in absolute value

> *Experimental Sub-hypothesis 1*: Including an Optimism Uplift to a task/project forecast will, in general, increase the final count of resources and overall time to complete a task/project, in respect to a situation where optimism uplift is not included in the initial forecast for the same task/project

To test my hypothesis, I considered all the variables discussed in the previous sections in accordance with the theoretical framework. In the first experiment, I first introduce three models and then, in the robustness check section, I explore three additional models, taking into account one potential confounding variable.

To explore the relationships between forecasted and actual values in terms of time and resources when using an optimism uplift, I created two different variables: one for difference in bricks and one for difference in time. These variables were created because taking the difference in terms of time and resources for the two experimental groups creates a variable that can be directly interpreted as forecast precision.

When considering sub-hypothesis 1, the graphical analysis of Figure 22, showing the data distribution of the actual number of bricks used by participants in the two conditions, already gives some hints about whether or not including an optimism uplift to a task forecast may increase the final count of resources and overall time to complete a task/project. Indeed, as mentioned above, participants in the OU condition, on average used slightly more bricks in order to complete the task. However, by capitalising on the unique characteristics that the experimental method offers, the exact or "right" number of bricks needed to complete the structure is known to be 35 (see method section for details), I created another variable, considering the difference between the “right” number of bricks and the estimated bricks by participants, to understand the actual effect of adding an optimism uplift during the estimation process. Unfortunately, the same cannot be done for the time variable as there is not a “right” time to finish the task, the 1200 seconds were only set as an arbitrary threshold to not prolong the experiment excessively.

The histogram in Figure 23 shows the distribution of the created variable “difference in bricks” (actual minus estimated bricks) by treatment administered: looking at the Estimation 1 (E1) manipulation column, it is clear how most of the values falls in the range -20/+50, and the distribution curve is skewed to the right, as predicted by the cost overrun literature (Flyvbjerg et al., 2018). OU column, instead, presents less spread values around the average, which is also much lower with 95% of values in the range of -25/+25; the distribution curve is quasi-normal, considering the fact that the graph reports the range values of -40/+60. These distribution curves, give a preliminary indication of the fact that when using an optimism uplift as a de-biasing technique, variability around the true level of resources needed in order to complete a task sharply decreases (in this case a decrease of around 54.5%), making estimations more accurate.

Another consequence of this is the fact that a larger proportion of estimates are above the final number of bricks used to complete the task.

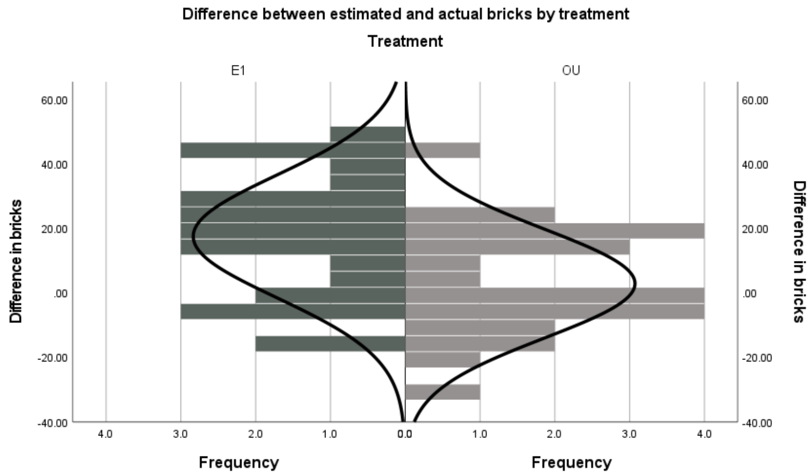


Figure 23 - Difference between estimated and actual bricks by treatment

To better visualise the variable created describing the difference in seconds between the estimated and actual time it took for participants to finish (or fail) the task, Fig. 24, shows two boxplots, one per each treatment. As the failure rate of the task was quite high, values are naturally closer to the maximum amount of time the participants had at their disposal in order to finish the task (1200s). However, when looking at the range of the two boxplots, we notice straight away that the IQR (ranging from 0.25 to 0.75 of the data) of the boxplot representing participants in the OU condition, is much closer to 0 (meaning little to no difference between estimated and actual time), whereas the IQR of the boxplot of participants in the E1 conditions starts in a value that is around 300s higher. Considering that the maximum number of seconds to complete the task was 1200s, 300s difference between the two conditions' IQRs is a considerable discrepancy between the two conditions. This supports using optimism uplift as mitigation of the effect of optimism bias in estimations. Moreover, comparing mean difference in time (represented by the crosses in both boxplots), indicates that participants in the OU condition were on average more accurate than those in the E1 condition.

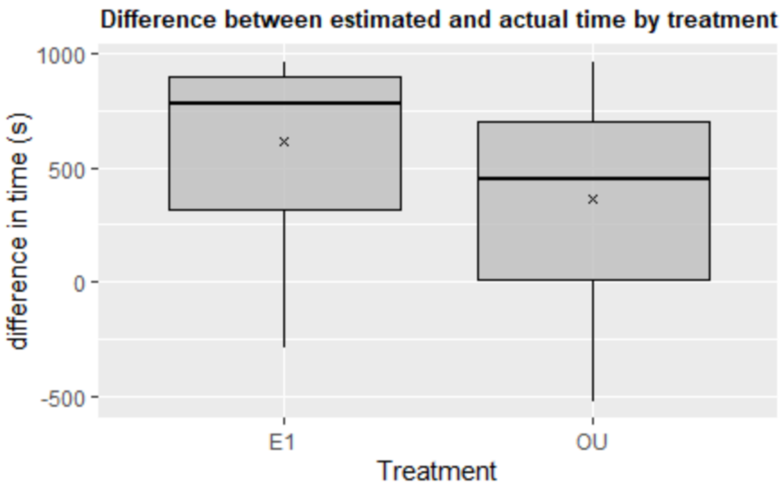


Figure 24 - Difference between estimated and accrual time by treatment

Finally, to better understand the distribution of the third variable created, representing the difference between estimated and “right” number of bricks by treatment, I created two boxplots reporting the average value (with the cross) as well, in fig. 25. Whereas the participants in the OU condition had a distribution directed mostly towards positive values with fewer incidence of negative values, indicating that people were on average conservative in their estimations but with some subjects being optimistic in their predictions. At the same time, those in the E1 condition were mostly optimistic in their estimation with very few values going over 0 and the IQR going from around -5 to -25. This factor represents a further indication that the manipulation had an impact on the estimation bias by making from one side participants more conservative in their estimations and from the other helping them to get closer to the ideal level of resources needed to complete the task. In other words, the IQR reporting the OU manipulation data covers the “true” value of resources needed in order to complete the task, whereas the other boxplot does not. This might be another indication of the fact that optimism uplift has a positive effect on forecast precision.

Patterns unveiled with the preliminary analysis of these three variables might seem to indicate a support towards what is theorised in this research, however, in order to establish if there is a significant relationship between the variables explored and the manipulation received by the participants, in the next section I present three regression models.

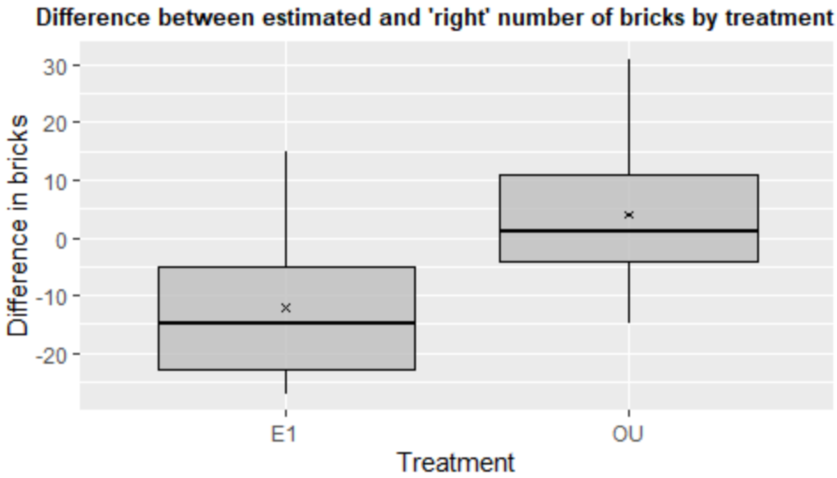


Figure 25 - Difference between estimated and “right” number of bricks by treatment

6.3.3. Hypothesis and sub hypothesis testing – regression analyses

As Table 20 shows, three models are examined, considering only one dependent and one independent variable: the dependent variables for each of the three models are those variables presented in the last section, whereas the independent variable is represented by the treatment received by the participants (Simple estimation task and estimation with Optimism Uplift).

The decision not to include any other confounding variable at this stage of the analysis, is justified by the experimental design that makes it unnecessary to put control as people are randomised in the condition.

Model 1 looks at the difference between estimated and actual bricks used by participants on average in relation to the treatment they were allocated to. Subjects that received the OU manipulation on average had a difference between actual and estimated value of 2.66 bricks (17.30 - 14.64) which corresponds to a difference of 146.7% bricks less than those who received the E1 manipulation. The significance level of the difference in bricks variable in the first linear model is high ($p<0.001$). This model shows that participants in the OU condition were on average more accurate when estimating the number of bricks necessary in order to complete the structure with a much smaller difference between estimated and actual bricks. Significance is also confirmed by looking at the 95% confidence interval, which does not cross zero and consequently shows that the null hypothesis can be rejected.

Model 2 reports the regression analysis performed considering the difference between expected and actual time (in seconds) according to the manipulation participants were allocated in. Subjects that received the OU manipulation on average had a difference between actual and estimated value of 360.39 seconds (613.15 – 252.76) corresponding to 51.9% less time than those that received the E1 manipulation. The p-value of the difference in time coefficient is 0.026. Also in this case, the significance level meets the significance threshold expectation set at 5%. Furthermore, the confidence interval (95%) does not cross the zero value, therefore the null hypothesis of this experiment regarding the time estimations can be rejected.

	Dependent variable:		
	Diff. in bricks (n)	Diff. in time (s)	Diff. with real bricks (n)
	(1)	(2)	(3)
Opt. Uplift	-14.64^{***}	-252.76^{**}	16.00^{***}
	(4.95)	(109.00)	(3.46)
Intercept	17.30^{***}	613.15^{***}	-12.00^{***}
	(3.46)	(77.03)	(2.42)
Opt. Uplift 95 C.I.	$(-24.57, -4.72)$	$(-473.56, -31.97)$	(9.06, 22.94)
Observations	53	53	53
R^2	0.15	0.11	0.30
Adjusted R^2	0.13	0.09	0.28
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 20 - Regression analyses and C.I. experiment 2

Model 3 explores the relationship between the difference of expectations in terms of bricks needed in order to complete the task and the “right” number of bricks to complete the task (35) according to the manipulation subjects were placed in. Participants receiving the OU treatment, were on average 4 (-12 + 16) units wrong from estimating the right number of bricks or, in other words, nearly three times closer to estimate the “right” number of bricks. This result, is in line with the other results seen until now, showing, once again, that OU condition had an impact on estimation capabilities of the subjects, making them more accurate. The p-value is smaller than 0.01%, which confirms that the use of an optimism uplift has a major impact in improving the precision of estimates produced by the subjects.

6.3.4. Hypothesis and sub hypothesis testing – robustness analysis

As in the case of the first experiment, since there could have been other confounding factors that could have hindered the capacity to see the effects investigated in the results reported, such as the age imbalance between the two groups, this variable was plugged into the models explored in Table 20, to see its impact, as showed in Table 22. Moreover, this section offers a few more details on the first three models, including a robustness check using the Mann-Whitney test, in Table 21.

Mann-Whitney tests	
	p-value
Model 1	0.007 **
Model 2	0.02 *
Model 3	5.43e-05 ***

Table 21 - Mann-Whitney tests

Indeed, by considering the distributions of available data as non-normal, in line with what the literature reports, I performed Mann-Whitney tests for the variable under scrutiny in order to check for significance also with this method. Mann-Whitney test, therefore, provides another perspective on the variables explored in the three models described above, producing p-values that recognise the non-normal distribution of the data considered for this analysis and providing a more complete view on the hypothesis explored in this experiment, since it is a test to explore the differences, if any, between two independent populations (Newbold, 2013). As Table 24 reports, when considering the difference in bricks variable used to build regression model 1, the p-value is smaller than 0.01, confirming the findings of the regression analysis. When it comes about Model 2 the p-value observed is at 0.02, slightly lower than the one found with the regression analysis. Model 3, reports a p-value slightly larger than the one observed in the regression analysis, but still very close to 0. These Mann-Whitney tests, therefore, strengthen the results reported with the models in the previous section.

When considering the three models showed beforehand, adding Age as a confounding factor appears to have a slight impact on the result observed. As explained in the context of the first experiment, it seems reasonable to add Age as a confounding factor, because of the online nature of the experiment designed: as a matter of fact, older participants, which might have not been as familiar as younger participants with the usability of the survey and of the 3D game, may have been more conservative in their estimations in order to allow room for errors in case of misunderstanding on how to use the building tool integrated in the 3D game. On the other hand, younger subjects might have been less conservative in their estimates because they spend more time online and they might do more activities in this environment. Table 22 reports the results of the 3 models presented above adding as an extra confounding variable Age.

	Dependent variable:		
	Diff. in bricks (n)	Diff. in time (s)	Diff. with real bricks(n)
	(4)	(5)	(6)
Opt. Uplift	−16.65***	−283.53**	16.66***
	(4.85)	(110.67)	(3.52)
Age	0.73**	11.26	−0.24
	(0.33)	(7.59)	(0.24)
Intercept	−0.33	342.42*	−6.19
	(8.68)	(197.8)	(6.29)
Opt. Uplift 95 C.I.	(−26.46, −6.89)	(−505.83, −61.25)	(9.59, 26.52)
Age 95 C.I.	(0.06, 1.38)	(−3.99, 26.52)	(−0.73, 0.24)
Observations	53	53	53
R^2	0.22	0.13	0.31
Adjusted R^2	0.19	0.10	0.28
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 22 - Regression models and C.I. with Age as a confounding variable

Model 4 looks at the difference between estimated and actual bricks used by participants on average in relation to the treatment they were allocated in. The discrepancy between the subjects in the two conditions is more than the one seen in model 1, corresponding to 192.4% bricks difference on average for subjects in the OU condition. The p-value of the independent variable is still highly significant, however, also the p-value of Age variable is significant with $p<0.001$, and confidence intervals related to those variables do not cross the zero value, supporting the findings of the p-value. The similar level of significance emerging from this model, reinforces the results of model 1, by confirming that indeed there is a relationship between the variables under study. The significance of the Age variable in Model 4 confirms that age might have played a role in the estimation behaviour and capabilities of subjects and, this is the reason why the model is reported in this section.

Model 5 reports the regression analysis performed considering the difference between expected and actual time (in seconds) according to the manipulation participants were allocated in, adding Age as a further confounding variable. Also in this case, the difference between the two treatment groups is higher than the one observed in model 2 at 141.3% more for participants in the E1 condition. The significance level of this model is higher than the one seen in model 2 for the treatment variable, with a p-value of 0.01. However, the Age variable does not seem significant, as confirmed also when looking at the 95% confidence interval. This model even if confirming the results observed in Model 2, shows that Age might have influenced the estimations of the participants, however, this difference is more significant in terms of the treatment variable rather than in the Age variable.

Model 6 explores the relationship between the difference of expectations in terms of bricks needed in order to complete the task and the “right” number of bricks to complete the task (35) with Age added as a confounding variable together with the treatment type. As in the case of model 3, there is a considerable discrepancy between the two treatments: participants in the OU condition were on average 199% closer to the right number of bricks than those in the E1 condition. As seen in Model 3, the p-value is very close to 0, making the difference in bricks variable highly significant, whereas the p-value for the Age variable is quite above the threshold level of significance. This model confirms the reasoning behind the structure of model 3, suggesting, once again that participants in the OU condition were on average more accurate in their estimations.

Since the three models just presented contain one confounding variable, the interactions between treatment and age have been explored, which did not yield any significant result. Moreover, all the models were tested for collinearity using VIF (Variance inflation factor), and as can be seen in the results in table 23, no VIF is greater than 5, therefore no collinearity has been found in the models.

Collinearity	
Model	*VIF (reject if VIF > 5)*
4	*Treatment:* 1.04, *Age:* 1.04
5	*Treatment:* 1.04, *Age:* 1.04
6	Treatment: 1.04, Age: 1.04

Table 23 - Collinearity

Finally, throughout the six models presented, I reported Adjusted R^2; the values related to this statistic are not high for all models, not going above the value of 0.28. This is in line with the assumptions and claims of this research, as not all inaccuracies in estimations arise from excessive optimism but only this factor is taken into consideration for the current investigation.

6.4. Experiment 3

6.4.1 Manipulation

During the third experiment, participants were asked to perform the same building task in the 3D online tool as in the two previous experiments. Subjects in the experimental condition, in this case, before conducting the task, were asked to “unpack” the different actions they would have done in order to complete the task. Unpacking, unlike the optimism uplift manipulation of the experiment beforehand, does not draw from past distributional similarities in order to mitigate optimism bias in estimation tasks, but takes an internal perspective on the task at hand, allowing to analyse it more specifically, emphasizing the different actions to be performed in order to complete the task successfully.

Subjects in the experimental condition, were prompted to carry out their forecasting tasks with the following instructions:

> *“Before starting to build the structure, you will have 7 minutes to analyse and write down the actions, in chronological order, that you will need to perform in order to complete it. An example of action could be “which bricks are part of the structure and which are not”. Don’t worry about grammar or sentence structure, just separate the actions with a point or numbered list. The actions you will write down will be reviewed by the researcher,* ***please make sure you write at least five actions for your results to be considered valid.****”*

The aim of this manipulation is to explore whether giving the chance to the subjects to think further about the task they were going to perform would have made forecasts more precise and therefore increase the reliability on the estimation task as suggested in the body of literature ascribed to the internal view on forecasting, discussed in the literature review. After the unpacking task, participants were directed to the estimation page, where they were asked if they thought they would have successfully completed the task or not, with how many bricks and how long it would have taken). 3D building tool page to perform the construction task and once finished, their result in terms of bricks and time were reported.

6.4.1 Analytical approach

The following variables were considered for all participants in all conditions: the dependent variable was represented by the level of error in forecasting measured through the difference between forecasted number of bricks and actual bricks. The independent variable were the two treatment conditions: the Unpacking condition (U) and the simple estimation condition (E2). Subjects in the E2 condition received the same instructions as participants in the control group of experiment two (E1), however, to differentiate data of one experiment from the others, subjects' control condition for experiment 3 is named E2. A further dependent variable was created looking at the difference between the actual number of bricks and the "right" number of bricks, to have more insights on the participants' forecast precision. When studying the variables related to time, the same independent variable was taken and a similar process for the dependent variable was followed considering the dependent variable as being the difference in terms of seconds between forecasted time to finish the construction task and actual time it took the participants to complete it. The three variables were studied through a linear regression analysis, creating six models, three of which adjusted for age of the participants. The analysis was performed using the R software, graphs were built using ggplot2 package and regressions summary tables using stargazer package (R Core Team, 2014; Wickham, 2016; Hlavak, 2018).

When studying the descriptive statistics of the study, expectations in terms of succeeding or not to the task and actual success rate in relation to manipulation received were considered to better understand whether the success rate had been impacted by the manipulation received or not.

6.5 Results

6.5.1 Descriptive statistics

Table 24 presents descriptive statistics.

Experiment 3, n = 52				
General	m = 31	f = 21	average age = 28.1	
Manipulation	E2 = 26		U = 26	
Statistic	*Mean*	*SD*	*Mean*	*SD*
Success data estimate	69%	-	73%	-
Success data actual	38.5%	-	42.3%	-
Average bricks estimate (n)	19.6	10.6	25.5	13.9
Actual bricks (n)	62.9	24.3	49.7	22.6
Average time estimate (s)	391.5	226.5	445.6	234.6
Actual time (s)	985.7	312.2	970.7	318.3

Table 24 - Descriptive statistic experiment 3

Participants, together with their estimations, were asked to indicate whether they thought they could complete the task successfully. Considering the nature of the Unpacking (U) manipulation, aimed at making the estimation more precise, the expectation on this would be that participants in the U condition, on average, would tend to be more optimistic towards successfully completing the task. This is confirmed by the data, showing that, on average, more participants in the U condition thought they would have successfully completed the task in relation to the other group not receiving the manipulation. When those data are compared to the ones of actual success rate of the task, on average, subjects in the U conditions were more successful than their counterparts in the E2 condition.

This gives an early indication that, when participants unpack their task into different actions, they feel slightly more confident towards their estimations and this might lead to successfully complete a project. This is not always the case, however, as there are still some behavioural biases related to overconfidence that might come into play when making an estimate. This is consistent with the assumption made in this research that optimism bias cannot be eliminated from the appraisal process, but certain tools might help in the mitigation of it as, for example, the use of unpacking techniques.

Data related to estimated and actual time to finish the task are also presented in the table of the descriptive statistics for the experiment. Participants in the E2 condition were, on average, more conservative in their time estimation by 12.9% in respect to the subjects in the U condition and took, on average, 15 seconds more to complete the task. This indicates that, in terms of time, when unpacking a task before elaborating estimates the difference between estimated and actual time slightly diminishes. As in the case of the other experiments, actual time values are skewed towards 1200 seconds as, participants who ran out of time are recorded has having taken 1200 seconds to “complete” the task, so this factor, together with the success rate on the task, should be kept in mind when analysing these variables.

Further to these data, results in terms of the average bricks estimated to complete the task and actual bricks used by the participants are reported; for these entries both participants succeeding and failing in completing the task in the given timeframe of 20 minutes were recorded. Subjects in the E2 condition estimated on average 26.2% fewer bricks to complete the task than those in the U condition. Once again, subjects in the E2 condition were optimistic when estimating the number of bricks, estimating, on average, 56.4% fever bricks than what needed to complete the task (at least 35 bricks). Participants in the U condition were also optimistic in their estimations, however, they estimated, on average 31.4% bricks fever than what needed in order to complete the task successfully without mistakes. Those data seem to indicate that Unpacking had an impact on the precision of the estimates of the participants, however, the power of this optimism bias mitigation tool, does not seem as powerful as the one explored in the previous experiment.

When it comes about the actual number of bricks used, on the other hand, subjects in the U condition used on average 23.4% less bricks than those in the E2 condition, as can be seen graphically in the boxplot in figure 26, reporting the distribution of the actual number of bricks used per manipulation received. This factor shows that unpacking a task into sub-actions is helpful in order to reduce the total level of resources used in order to carry out a task or a project when compared to a situation where unpacking did not take place.

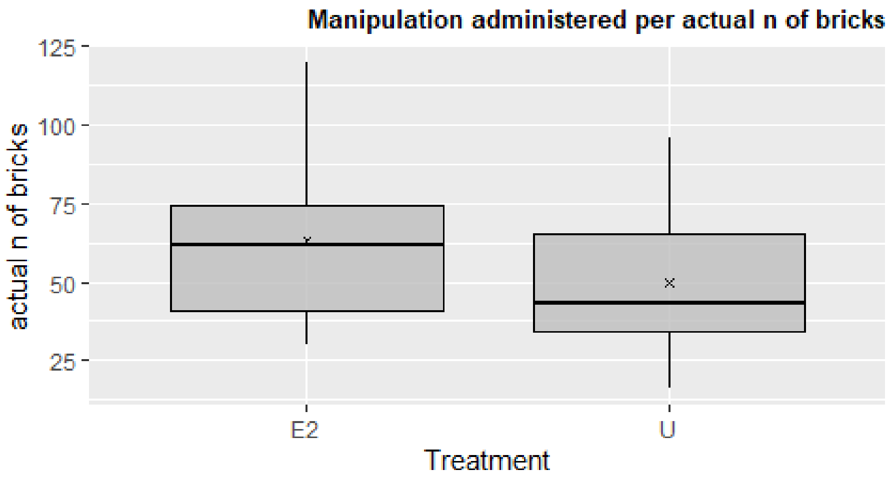


Figure 26 - Manipulation administered per actual number of bricks

6.5.2 Hypothesis and sub hypothesis testing – preliminary analysis.

Hypothesis for this experiment is the following:

Experimental Hypothesis 1: Unpacking a determined task gives better result in terms of forecast precision *(where forecast precision is defined as the minimisation of difference between expected value when making a forecast and actual value once the task/project is completed)*

As in the other experiments, to focus the analysis mainly on the variables under investigation, I introduce three models and then, in the robustness check section, three more models are explored, considering one confounding variable.

To explore the relationships between forecasted and actual values in terms of time and resources when doing an unpacking task opposite to a situation where this is not used, after having considered the existing variables, I created three different variables, one for difference in bricks, one for difference in time and one for difference between the “right” number of bricks and actual bricks. I have built these variables considering that, by taking the difference in term of time and resources for the two manipulation groups, information in terms of forecast precision would have been easier to unveil through statistical testing, in line with the definition of “forecast precision” given in the hypothesis of this experiment.

Figure 27, with a histogram, shows the distribution of the created variable “difference in bricks” (actual minus estimated bricks) by treatment administered: looking at the Estimation 2 (E2) manipulation column, it is clear how most of the values falls in the range -10/+120. U column, instead, presents slightly less spread values around the average which is also lower of around 20 units in the positive part of the curve. These distribution curves, give a preliminary indication of the fact that when using unpacking as a de-biasing technique, variability around the true level of resources needed in order to complete a task decreases (in this case a decrease of around 12%), making estimations more accurate.

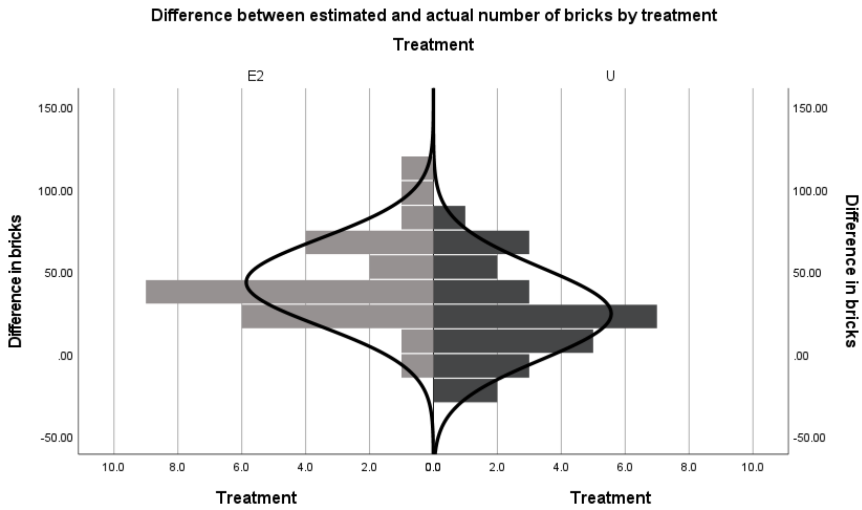


Figure 27 - Difference between estimated and actual number of bricks by treatment

To visualise the variable created to describe the difference in seconds between the estimated and actual time it took for participants to finish (or fail) the task, Fig. 28, shows two boxplots, one per each treatment. As the failure rate of the task was high, values are closer to the theoretical maximum amount of time the participants had at their disposal in order to finish the task (1200s). However, when looking at the range of the two boxplots, we notice that the IQR of the boxplot representing participants in the U condition, is closer to 0 (meaning little to no difference between estimated and actual time), whereas the IQR of the boxplot related to the subjects in the E2 condition is more distant to 0. As showed also with the crosses in both boxplots, this gives an indication of the fact that participants in the U condition were on average more accurate than those in the E2 condition.

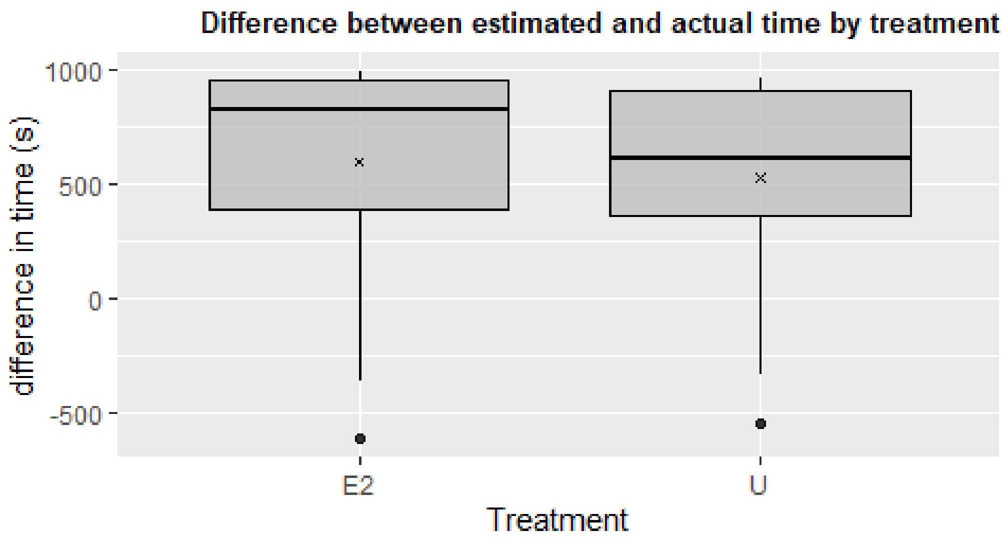


Figure 28 - Difference between estimated and actual time by treatment

Finally, to better understand the distribution of the third variable created, representing the difference between estimated and "right" number of bricks by treatment, I created two boxplots reporting also the average value (with the cross), in fig. 29. Participants in the U condition had a distribution, once again, closer to the 0 value as they were on average more conservative in their estimations in relation to subjects in the E2 condition, which were more optimists in their estimation with very few values going over 0 and the IQR going from around -10 to -25. This factor represents a further indication that the manipulation had an impact on the estimation bias by making from one side participants more conservative in their estimations and from the other helping them to get closer to the ideal level of resources needed to complete the task.

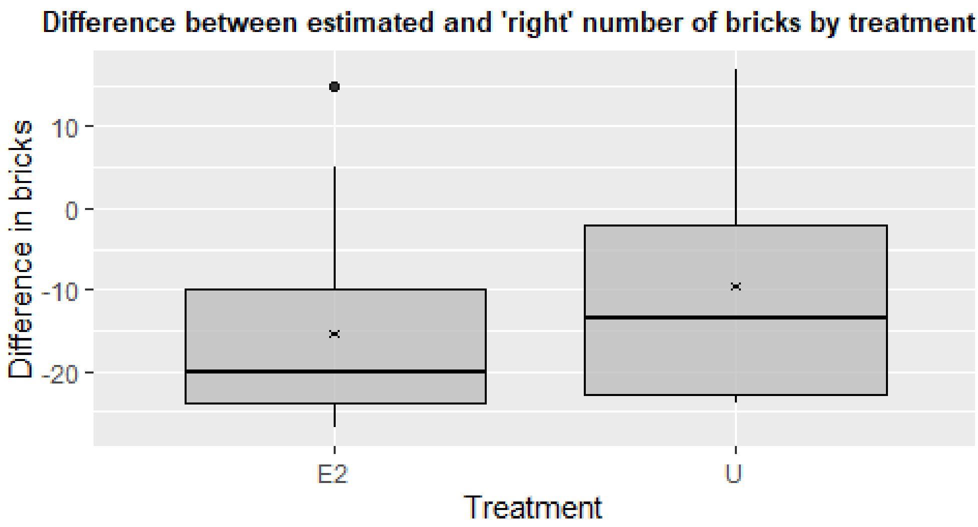


Figure 29 - Difference between estimated and “right number of bricks by treatment

6.5.3 Hypothesis and sub hypothesis testing – regression analyses

As Table 25 shows, three models are examined, considering only one dependent and one independent variable: the dependent variables for each of the three models are those variables presented in the last section, whereas the independent variable is represented by the treatment received by the participants (Simple estimation task and estimation with Unpacking task).
The decision to not include any other possible confounding variable at this stage of the analysis, is justified by the experimental design which aims to look at how the inclusion of an unpacking task impacts the accuracy in the estimations of the subjects, net of any other factor that could alter their perception of the estimation and task in general.

Model 1 looks at the difference between estimated and actual bricks used by participants on average in relation to the treatment they were allocated to.

Subjects that received the U manipulation on average had a difference between actual and estimated value of 24.23 bricks (43.27 - 19.04) which corresponds to a difference of 56.4% bricks less than those who received the E2 manipulation. The significance level of this first linear model is 0.015. This model supports the data observed through the descriptive statistics reported in the previous section, showing that participants in the U condition were on average more accurate when estimating the number of bricks necessary in order to complete the structure with a smaller difference between estimated and actual bricks. Significance is also confirmed by looking at the 95% confidence interval, which does not cross zero and as a consequence shows that the null hypothesis can be rejected when it comes about estimations in terms of resources.

Model 2, reports the regression analysis performed considering the difference between expected and actual time (in seconds) according to the manipulation participants were allocated in. Subjects that received the U manipulation on average had a difference between actual and estimated value of 525.11 seconds (594.19 – 69.08) corresponding to 12.3% seconds less than those that received the E2 manipulation. The p-value of this regression model, however, is far above the 0.05 significance threshold as confirmed also by the confidence interval (95%) which crosses the zero value; as a consequence, the null hypothesis of this experiment regarding the time estimations cannot be rejected.

The first two models, therefore, show that when it comes to testing how an unpacking task has an impact on forecast precision in terms of resources and time and in turn if it is able to mitigate optimism bias, there might be different behavioural factors and/or processes that come into play. Therefore, when studying optimism bias, it should not be assumed that the impact in terms of magnitude of optimism bias in estimations of resources and time is similar. The fact that this assumption is often incorrect should lead to the re-examination of prior research. However, the different patterns observed in the two models, might also be related to the fact that the difference in time variable, unlike the difference in bricks variable, is anchored to the maximum amount of time allowed to participants to finish the task (1200 seconds), therefore, the sensitivity of the model might be altered by this factor.

Model 3, explores the relationship between the difference of expectations in terms of bricks needed in order to complete the task and the "right" number of bricks to complete the task (35) according to the manipulation subjects were placed in. Participants receiving the U treatment, were on average 9.53 units (-15.38 + 5.85) wrong from estimating the right number of bricks or, in other words, 47% closer to estimate the "right" number of bricks. This result, is in line with the other results seen until now, showing, once again, that U condition had an impact on estimation capabilities of the subjects, making them more accurate. When looking at the significance of the model, however, a p-value of 0.09 is observed, which, when compared with the 95% confidence interval, crossing the 0 value, shows that the difference highlighted are not statistically significant.

	Dependent variable:		
	Diff. in bricks (n)	Diff. in time (s)	Diff. with real bricks (n)
	(1)	(2)	(3)
Unpacking	-19.04^{**}	-69.08	5.85^{*}
	(7.56)	(122.96)	(3.43)
Intercept	43.27^{***}	594.19^{***}	-15.38^{***}
	(5.35)	(86.95)	(2.42)
Unpacking 95 C.I.	$(-34.23, -3.85)$	$(-316.06, 177.93)$	$(-1.05, 12.74)$
Observations	52	52	52
R^2	0.11	0.01	0.05
Adjusted R^2	0.09	-0.01	0.04
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 25 - Regression models and C.I. Experiment 3

6.5.4 Hypothesis and sub hypothesis testing – robustness analysis

As in the case of the other experiments, since there could have been other confounding factors that could have hindered the capacity to see the effects investigated in the results reported, such as the age imbalance between the two groups, this variable was plugged into the models explored in Table 25, to see its impact, as showed in Table 27. Moreover, this section offers a few more details on the first three models, including a robustness check using the Mann-Whitney test, in Table 26.

Mann-Whitney tests	
	p-value
Model 1	0.01*
Model 2	0.25
Model 3	0.10

Table 26 - Mann-Whitney tests

Indeed, by considering the distributions of available data as non-normal, in line with what the literature reports, I performed Mann-Whitney tests for the variables under scrutiny in order to check for significance also with this method and to explore whether there are some considerable differences in terms of p-value between the two methods, so to have a more complete view on the hypothesis explored. Mann-Whitney test, therefore, provides another perspective on the variables explored in the three models described above, producing p-values that recognise the non-normal distribution of the data considered for this analysis and providing a more complete view on the hypothesis explored in this experiment, since it is a test to explore the differences, if any, between two populations (Newbold, 2013). As Table 26 reports, when considering the difference in bricks variable used to build model 1, the p-value is 0.01, confirming the findings of the regression analysis. When it comes about Model 2 the p-value observed is at 0.25, lower than the one found with the regression analysis, but still far from the significance level. These Mann-Whitney tests, therefore, confirms the results of the regression analyses presented in the previous section.

When considering the three models showed beforehand, adding Age as a confounding factor appears to have a slight impact on the result observed. As explained in the context of the two experiments before, it seems reasonable to add Age as a confounding factor, because of the online nature of the experiment designed: as a matter of fact, older participants, which might have not been as familiar as younger participants with the usability of the survey and of the 3D game, may have been more conservative in their estimations in order to allow room for errors in case of misunderstanding on how to use the building tool integrated in the 3D game. On the other hand, younger subjects might have been less conservative in their estimates because of the higher exposure to online means. Table 27 reports the results of the 3 models presented above adding as an extra confounding variable Age.

Model 4 looks at the difference between estimated and actual bricks used by participants on average in relation to the treatment they were allocated in. The discrepancy between the subjects in the two conditions is less than the one seen in model 1, corresponding to 33.4% bricks difference on average for subjects in the U condition. The p-value of the independent variable is still significant, whereas the p-value of Age variable is not significant with a value of 0.11, and confidence intervals related to this variable crossing the zero value, supporting the finding of the p-value. The similar level of significance emerging from this model, reinforces the results of model 1, by confirming that indeed there is a relationship between the variables under study. The non-significance of the Age variable in Model 4 shows that even if age might have played a role in the estimation behaviour and capabilities of subjects, this is not statistically significant.

	Dependent variable:		
	Diff. in bricks (n)	Diff. in time (s)	Diff. with real bricks(n)
	(4)	(5)	(6)
Unpacking	−16.56**	−124.06	4.70
	(7.60)	(120.71)	(3.45)
Age	−0.56	12.43**	0.26
	(0.35)	(5.51)	(0.16)
Intercept	57.87***	270.49	−22.14***
	(10.45)	(166.04)	(4.74)
Unpacking 95 C.I.	(−31.453, −1.664)	(−360.648, 112.523)	(−2.058, 11.455)
Age 95 C.I.	(−1.241, 0.119)	(1.633, 23.230)	(−0.049, 0.568)
Observations	52	52	52
R^2	0.16	0.10	0.10
Adjusted R^2	0.12	0.08	0.07
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 27 - Regression models and C.I. with Age as confounding variable

Model 5 reports the regression analysis performed considering the difference between expected and actual time (in seconds) according to the manipulation participants were allocated in, adding Age as a further confounding variable. In this case, the difference between the two treatment groups is higher than the one observed in model 2 at 59.5% more for participants in the E2 condition. For the difference in time variable, the model is still not significant. However, the Age variable is significant, as confirmed also when looking at the 95% confidence interval. This model even if confirming the non-significance of the difference in time variable observed in Model 2, shows that Age had a statistically significant effect on the estimations of the participants.

Model 6 explores the relationship between the difference of expectations in terms of bricks needed in order to complete the task and the “right” number of bricks to complete the task (35) with Age added as a confounding variable together with the treatment type. As in the case of model 3, there is a discrepancy between the two treatments: participants in the U condition were on average 23.8% closer to the right number of bricks than those in the E2 condition. As seen in Model 3, the p-value is not significant as in the case of the p-value for the Age variable that, even if lower, is still above the threshold for the level of significance, making it a variable that can be discarded from the current analysis.

Since the three models just presented contain one confounding variable, all interactions have been explored, which did not yield any significant result. Moreover, all the models were tested for collinearity using VIF (Variance inflation factor), and as can be seen in the results in table 10, no VIF is greater than 5, therefore no collinearity has been found in the models.

Collinearity	
Model	*VIF (reject if VIF > 5)*
4	*Treatment:* 1.04, *Age:* 1.04
5	*Treatment:* 1.04, *Age:* 1.04
6	Treatment: 1.04, Age: 1.04

Table 28 - Collinearity

Finally, throughout the six models presented, Adjusted R^2 scores were reported; as can be noticed, the values related to this statistic are not particularly high for all models, not going over the value of 0.12. This is in line with the assumptions and claims of this research, as not all inaccuracies in estimation arise from excessive optimism but only this factor is taken into consideration for the current investigation.

6.6. Theoretical implications

The two experiments just presented, take as a starting point for the analysis two mitigation techniques used both in the literature and in practice ascribed respectively to the external and the internal view on forecasting. In the case of the external view, optimism uplift is used to reduce the impact of optimism bias in estimations; this technique considers historical data from past similar projects in order to elaborate an average value representing the average inaccuracy in estimations of all similar past endeavours. The experimental method, with its controlled environment, gives the chance to look at the net impact of these mitigation methods. With observational data, inaccuracies could arise, for example, from the data of past similar projects, as the task carried out by participants in order to build the optimism uplift is the same than the one carried out by participants in the OU condition.

Furthermore, in the literature review I introduced the concept of the fudge factor, which has some similar characteristics with the optimism bias uplift and is studied and used in the corporate finance field. Indeed, the fudge factor, as in the case of the optimism bias uplift, is a tool used in order to decrease uncertainty related to forecast imprecisions. As Brealey et al. (2012) report, the use of a fudge factor may lead to create a distorted image of the future cash flows for a given capital investment project and may hinder some of the characteristics of this investment that may be pivotal for the go or no-go decision to invest in the project in the first place. Therefore, using a fudge factor or an optimism uplift, in certain occasions, might lead to an overcorrection of the estimates, which might be detrimental for the overall project performance.

The literature I analysed, however, also reports arguments in support of the use of optimism uplifts, substantiated with relevant case studies in the infrastructure management field, however, no studies, to my knowledge, looked at the actual impact that an optimism uplift might have on the final count of resources used in order to complete a task. Once again, with the help of the experimental method, I was able to do this in experiment 2. In this experiment, the use of an optimism uplift led participants to use a greater amount of resources to complete their task (whilst also improving forecasting accuracy).

This factor is a very important one when considering which strategies to use in order to mitigate optimism bias in estimations, especially when appraising big projects, where the potential waste of resources could end up costing millions to project promoters. Therefore, accounting only for techniques coming from an external view might be detrimental not only as an estimation strategy of a project but also for its deliverables of it in terms of time and resources.

Experiment 3, on the other hand, presents a technique that belongs to the internal view on forecasting, unpacking, which aims to look at the estimation task from another perspective focusing on the actual steps that participants had to do in order to complete their task. Supporters of the internal view, in fact, claim that using as a base for estimation past similar projects or tasks might lead to biased estimation as they do not account for the unique nature of every endeavour. Results of my experiment show that even if on average participants in the unpacking condition were more accurate in their estimations when compared to the control condition, they were considerably less accurate than participants that received the optimism uplift condition. This indicates that unpacking might be regarded as a less powerful optimism bias mitigation tool than optimism uplift as confirmed also by the regression models presented for both experiments.

However, participants in the unpacking condition, used fewer bricks on average than participants in the control condition, suggesting that this mitigation technique did not lead to the use of a greater amount of resources. The first implication of this results is that it seems that even if both techniques have shortcomings, some of them are complementary. Optimism uplift is likely to increase the amount of resources used to complete a task, while unpacking does not. However, unpacking is weaker in increasing forecasting precision due to optimism bias when compared to optimism uplift.

This can also be seen by looking at the analysis of the variable describing the difference between the “real” number of bricks and the actual number of bricks. The IQR of the observations pertaining to the Unpacking technique did not cover the correct number of bricks needed to complete the task (35 bricks) whereas the IQR of the observations for the optimism uplift technique not only covered this value, but also accounted for mistakes that might happen during the execution phase of the task. The same can be said looking at estimations in terms of time: on average participants in the optimism uplift condition were more accurate than those in the unpacking condition.

The second implication from this reasoning is that, if some of the weaknesses of the two techniques are complementary then it does not make sense to look at those two perspectives as completely different and opposite, but we should approach them as two faces of the same medal both theoretically and practically.

Another factor arising from the analysis of experiment 2 is that optimism uplift seems to have a structural impact on the quality of estimations rather than on the behaviour of the estimator. When it comes to experiment 3, with unpacking, this effect is not observed, probably because subjects, by focusing only on the specific characteristics of the task at hand and not having any reference point regarding estimation inaccuracies on a similar task, are more likely to formulate their estimations based on the specific circumstance and therefore more influenced by their personal inclinations and capabilities towards the task. For this reason, unpacking might not be a strong enough debiasing technique if used alone, as confirmed also by the regression models analysed in the previous sections.

Furthermore, in both experiment 2 and 3, participants in the experimental groups, were, on average, more optimistic in assessing the likelihood of successfully completing the task than subjects in the control condition. This was also reflected in the actual success rate of the participants as, subjects in the experimental condition where on average more successful than those in the control condition in finishing the task. The consequence of this is that the use of both debiasing techniques lead to a greater success rate in completing the task, proving the usefulness of both techniques in this perspective. If both techniques lead to a greater success rate, than it is reasonable to assume that when using them together this should lead to an even higher success rate and not impair their usefulness respectively. This is true also when considering the highlighted complementarity of the two approaches, one looking at the specific nature of the task at hand and the other one looking at the past distributional similarities with the same task.

All in all, with the analysis of experiment 2 and 3 I gathered initial perspectives pertaining to the debate of internal versus external view on forecasting in the literature and operationalised them through the practical task part of those experiments. The results I presented, highlight the usefulness of both approaches in tackling certain challenges that represent an issue whenever an estimation about a task or project needs to be provided. Those results also indicate the strength and weaknesses of those approaches, which, in more than one occasion were shown to be complementary, opening the discussion for a cross-fertilisation of the two areas. This discussion has the potential to improve the current state of project behaviour when considering the forecasting process as I explore in more detail in the next chapter.

Appendix A: Optimism uplift calculation

Estimated bricks	Estimated time (s)	success or not	actual bricks	actual t (s)	actual result	Undos
21	260	yes	42	454	TRUE	1
8	240	yes	77	1200	TRUE	7
28	600	no	70	992	TRUE	8
25	300	yes	56	1108	TRUE	5
29	320	yes	34	364	TRUE	0
9	240	yes	53	951	TRUE	6
31	480	yes	37	544	TRUE	3
42	840	yes	36	506	TRUE	0
36	570	yes	39	665	TRUE	1
38	960	yes	40	877	TRUE	0
41	870	yes	34	381	TRUE	2
29	435	yes	43	506	TRUE	0
20	300	yes	74	878	TRUE	4
42	510	yes	34	310	TRUE	1
30	570	yes	37	866	TRUE	1
48	810	yes	42	384	TRUE	0
30	900	yes	71	551	TRUE	2
38	990	yes	34	246	TRUE	1
37	810	yes	42	885	TRUE	0
AVERAGE						
31	**579**		**47**	**667**		
Difference between estimated and actual time						
number of bricks			**number of seconds**			
16			**88**			

CHAPTER 7. EXPERIMENT 4

7.1. Introduction

After having explored the two perspectives on project estimation described in the literature review through experiment 2 and 3, the internal and the external view on forecasting and having highlighted strengths and weaknesses of both approaches and how some of those might be complementary between each other's, I now move on to discuss the phase B of the theoretical framework presented in Chapter 2.

Figure 30 shows the above-mentioned theoretical framework and highlights Phase B to better understand the aim of experiment 4. In fact, with this experiment, I investigate if the combination of the two different perspectives on forecasting considered in the previous experiments are able to yield better results in terms of optimism bias. In so doing, the integration process advised in the "Holistic view" framework will be explored not only at a theoretical level but also at a practical level with this experiment.

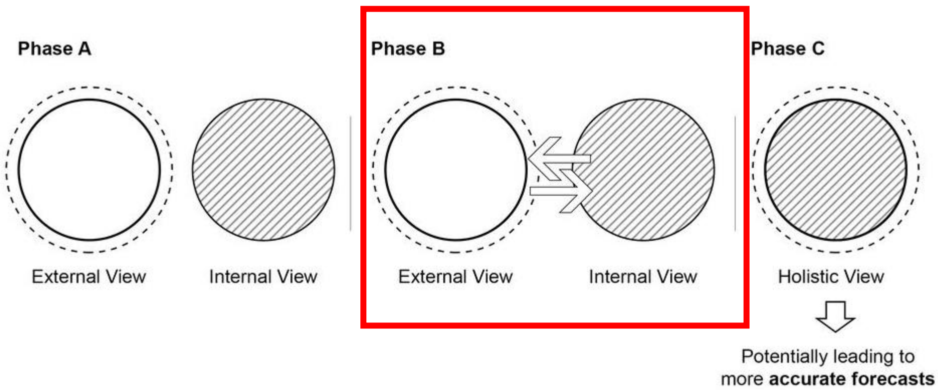


Figure 30 - Integration of the External and Internal view to create a Holistic View on forecasting (by author)

Moreover, through the analysis of the experiment's results, I can show if the concurrent use of the tools explored in the previous two experiments reduces or increases the gap between expected versus actual time and/or resources. This factor will be contextually analysed with the findings of the previous experiments, which highlighted that one of the tools was "stronger" than the other in terms of optimism bias mitigation. In case there is an effect of mitigation on optimism bias, experiment 4 will give us insights on whether this effect is cumulative or compounded.

First, I analyse the results of the experiment, following the same structure and analytical approach used for the previous experiments. Afterwards, in the theoretical implications section, I carry out an analysis of the experiment to better understand what an integration process between the two discussed perspectives implicates. In this way, I show whether the claim of phase C of the Holistic view framework is realistic or not and what are the consequences both at the practical and theoretical level.

7.2. Manipulation

For the last experiment, participants were asked to perform the same building task in the 3D online tool as the previous experiments. Subjects in the experimental condition, in this case, before conducting the task, were asked to "unpack" the different actions they would have done in order to complete the task. After that, in the estimation page, participants were given the average resource (in terms of number of bricks) and time (in terms of seconds) overrun generated by participants in the control group during the first experiment, whose level of dispositional optimism was not artificially altered by a manipulation (those subjects in the Typical Day condition).

The aim of this treatment is to explore whether the combination of the two previous manipulations is able to increase forecast precision, in respect to a situation where only one of these is used. Indeed, the two manipulations take as theoretical foundation two different perspectives on how to mitigate optimism bias one deriving from the internal view (unpacking) and one from the external one (optimism bias uplift). Considering the theoretical framework of this research, suggesting the importance to adopt an holistic view which looks at aspects both from the inside and outside view, putting them together by capitalising on their differences, rather than treating them as two mutually exclusive perspectives as most of the literature currently does, designing and running an experiment which would give more information on the validity of this seemed to be the natural conclusion for this research's experimental design.

The control group was asked to conduct the same estimation task, however, they were not given any information on the average overruns of time and resources of previous participants neither were prompted to perform an unpacking task before the said estimation task, de facto having the same experimental flow as treatment E1 and E2 discussed for the two previous experiments.

7.3. Analytical approach

The following variables were considered for all participants in all conditions: the independent variable, was represented by the level of optimism exhibited measured through the difference between forecasted number of bricks and actual bricks. The dependent variable were the two treatment conditions: the Optimism Uplift + Unpacking condition (OUP) and the simple estimation condition (E3). A further dependent variable was created looking at the difference between the actual number of bricks and the “right” number of bricks, to have more insights on the participants’ forecast precision. When studying the variables related to time, the same dependent variable was taken and a similar process for the independent variable was followed considering the independent variable as being the difference in terms of seconds between forecasted time to finish the construction task and actual time it took the participants to complete it. The three variables were studied through a linear regression analysis, creating six models, three of which adjusted for age of the participants. The analysis was performed using the R software, graphs were built using ggplot2 package and regressions summary tables using stargazer package (R Core Team, 2014; Wickham, 2016; Hlavak, 2018).

When studying the descriptive statistics of the study, expectations in terms of succeeding or not to the task and actual success rate in relation to manipulation received were considered to better understand whether the success rate had been impacted by the manipulation received or not.

7.4. Results

7.4.1. Descriptive statistics

Table 29 presents descriptive statistics. Besides the mean values of the variables under scrutiny, I provide the standard deviation as well, so to look at the dispersion of the data points from the mean and have a preliminary idea about the effectiveness of the manipulation administered in this experiment.

Experiment 4, n = 52				
General	m = 20	f = 31	average age = 31.1	
Manipulation	E3 = 24		OUP = 28	
Statistic	*Mean*	*SD*	*Mean*	*SD*
Success data estimate	64%	-	77%	-
Success data actual	36.4%	-	57.7%	-
Average bricks estimate (n)	23.1	13.5	36.4	12.9
Actual bricks (n)	54.5	30.5	41	11.9
Average time estimate (s)	349.6	185.5	446.25	171.5
Actual time (s)	1093.3	300.1	1010.5	284.5

Table 29 - Descriptive statistics experiment 4

Participants, together with their estimations, were asked to indicate whether they thought they could complete successfully the task or not. Considering the nature of the OUP manipulation, aimed at making the estimation more precise, the expectation on this would be that participants in the OUP condition, on average, would tend to be more optimistic towards successfully completing the task. This is confirmed by the data, showing that, on average, more participants in the OUP condition thought they would have successfully completed the task in relation to the other group not receiving the manipulation. When those data are compared to the ones of actual success rate of the task, on average, subjects in the OUP conditions were considerably more successful (around 21% more) than their counterparts in the E3 condition.

This sets of data, gives an early indication of the fact that, when participants unpack their task into different actions, and are given an optimism uplift they might feel more confident towards their estimations and how those might lead to successfully complete a project.

Data related to estimated and actual time to finish the task are also presented in the table of the descriptive statistics for the experiment. Participants in the E3 condition were, on average, more conservative in their time estimation by 24.3% in respect to the subjects in the OUP condition and took, on average, 7.9% more time to complete the task. This indicates that, in terms of time, when elaborating estimates after unpacking a task and using an optimism uplift the difference between estimated and actual time diminishes. As in the case of the other experiments, actual time values are skewed towards 1200 seconds as, participants who ran out of time are recorded has having taken 1200 seconds to "complete" the task, so this factor, together with the success rate on the task, should be kept in mind when analysing these variables.

Results in terms of the average bricks estimated to complete the task and actual bricks used by the participants are also reported; for these entries both participants succeeding and not succeeding in completing the task in the given timeframe of 20 minutes were recorded. Subjects in the E3 condition estimated on average 44.7% fewer bricks to complete the task than those in the OUP condition. Subjects in the E3 condition, moreover, were quite optimistic when estimating the number of bricks compared to the "right" number of bricks it would have taken to complete the structure without any mistake, estimating, on average, 41% less bricks. On the other hand, subjects in the OUP condition were not optimistic in their estimations, as they estimated on average 1.4 units (or 3.9%) more than what needed in order to complete the task successfully without mistakes. Interestingly, subjects in the OUP condition, considering the average value, are the ones whose estimations were closer to the "right" number of bricks. Those data seem to indicate that Unpacking plus Optimism uplift had an impact on the precision of the estimates of the participants, however, in order to have more insights about this, I analysed the distribution of the variable, presented in the next section, before plugging it in a regression model.

When it comes about the actual number of bricks used, on the other hand, subjects in the OUP condition used on average 28.3% less bricks than those in the E3 condition, as can be seen graphically in the boxplot in figure 31, reporting the distribution of the actual number of bricks used per manipulation received. This factor shows that UOP is helpful in order to reduce the total level of resources used in order to carry out a task or a project when compared to a situation where these actions did not take place.

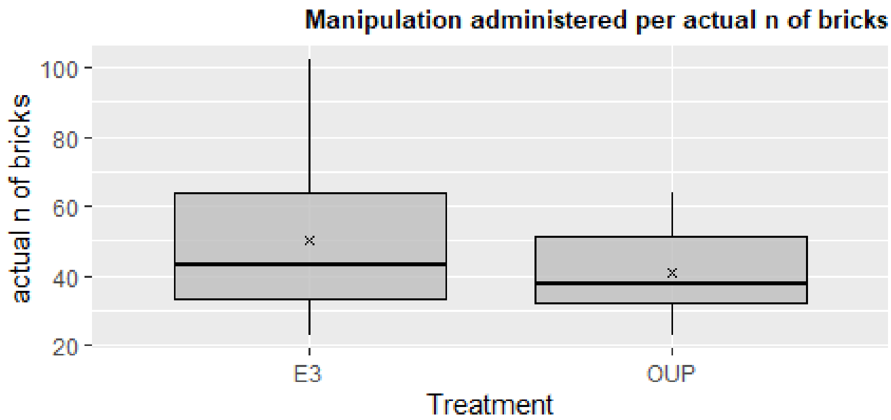


Figure 31 - Manipulation administered per actual number of bricks

7.4.2 Hypothesis and sub hypothesis testing – preliminary analysis

Hypothesis for this experiment is the following:

> *Experimental Hypothesis 1:* Combining unpacking with optimism uplift gives better results in terms of forecast precision than using the two approaches separately *(where forecast precision is defined as the minimisation of difference between expected value when making a forecast and actual value once the task/project is completed)*

To reject the null hypothesis, all the variables discussed in the descriptive statistics section were considered in accordance with the theoretical framework developed in the literature review of this dissertation and I built six statistical models. As in the case of the other experiments, in order to focus the analysis mainly in the variables directly studied and under investigations, I will first introduce three models and afterwards, in the robustness check section, three more models will be explored, taking into account more than one confounding variable.

To explore the relationships between forecasted and actual values in terms of time and resources when doing an unpacking task and providing information through a resource and time optimism bias uplift, opposite to a situation where those are not used, after having considered the existing variables, three different variables were created, one for difference in bricks, one for difference in time and one for difference between the “right” number of bricks and the actual bricks. These variables were created by considering that, by taking the difference in term of time and resources for the two manipulation groups, considerations in terms of forecast precision would have been easier to unveil through statistical testing, in line with the definition of “forecast precision” given in the hypothesis of this experiment.

Figure 32 shows the distribution of the created variable “difference in bricks” (actual minus estimated bricks) by treatment administered: looking at the Estimation 3 (E3) manipulation column, it is clear how most of the values falls in the range -25/+80. OUP column, instead, presents much less spread values around the average with most values falling in the range of -10/+30. These boxplots, give a preliminary indication of the fact that when using an unpacking plus optimism uplift as a de-biasing technique, variability around the level of resources needed in order to complete a task decreases (in this case a decrease of around 89.7%), making estimations considerably more accurate.

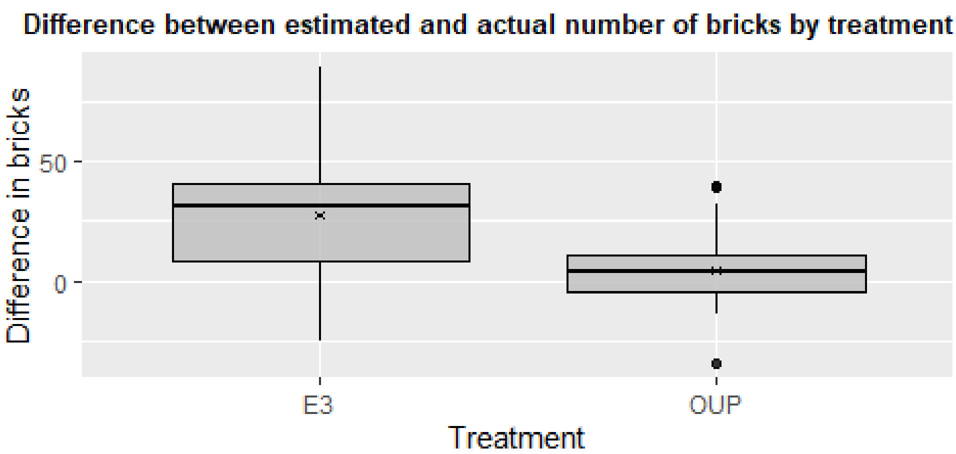


Figure 32 - Difference between estimated and actual number of bricks by treatment

To visualise the variable created to describe the difference in seconds between the estimated and actual time it took for participants to finish (or fail) the task, Fig. 33, shows two boxplots, one for each treatment. As the failure rate of the task was rather high, values are naturally closer to the maximum amount of time the participants had at their disposal in order to finish the task (1200s). However, when looking at the range of the two boxplots, we notice that the IQR of the boxplot representing participants in the OUP condition, is considerably closer to 0 (meaning little to no difference between estimated and actual time), whereas the IQR of the boxplot related to the subjects in the E3 condition is more distant to 0. As showed as well with the crosses in the two boxplots, representing the average value of the distributions, participants in the OUP condition were on average more accurate than those in the E3 condition, a difference that stands at around 250 seconds, which, when compared to the total time participants had in order to complete the task, accounts for more than one fifth of it.

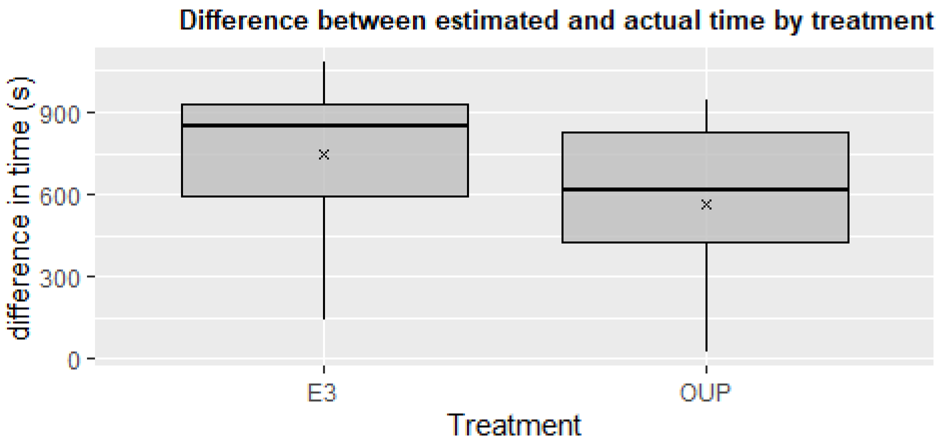


Figure 33 - Difference between estimated and actual time by treatment

Finally, to better understand the distribution of the third variable created, representing the difference between estimated and “right” number of bricks by treatment, I created two boxplots reporting also the average value (with the cross), in fig. 34. Participants in the OUP condition had a distribution, closer to the 0 value as they were on average more conservative in their estimations in relation to subjects in the E3 condition, which were more optimists in their estimations with few values going over 0 and the IQR going from around -4 to -22, with the average value being around -12. This factor, represents a further indication that the manipulation had an impact on the estimation bias by making from one side participants more conservative in their estimations and from the other helping them to get closer to the ideal level of resources needed to complete the task.

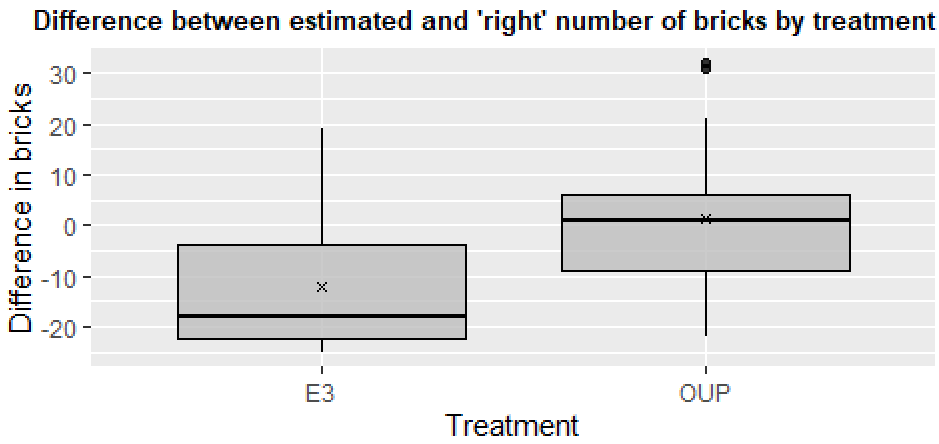


Figure 34 - Difference between estimated and "right" number of bricks by treatment

7.4.3 Hypothesis and sub hypothesis testing – regression analyses

As Table 30 shows, three models are examined, considering only one dependent and one independent variable: the dependent variables for each of the three models are those variables presented in the last section, whereas the independent variable is represented by the treatment received by the participants (Simple estimation task and Unpacking task plus estimation with Optimism Uplift).

The decision to not include any other possible confounding variable at this stage of the analysis, is justified by the experimental design which aims to look at how the inclusion of an unpacking task before the estimation plus a suggestion in terms of optimism uplift impacts the accuracy in the estimations of the subjects, net of any other factor that could alter their perception of the estimation and task in general.

Model 1 looks at the difference between estimated and actual bricks used by participants on average in relation to the treatment they were allocated to. Subjects that received the OUP manipulation on average had a difference between actual and estimated value of 1.57 bricks (31.42 – 29.85) which corresponds to a difference of 181% bricks less than those who received the E3 manipulation. The significance level of this first linear model is quite high being less than 0.0001. This model supports the data observed through the descriptive statistics reported in the previous section, showing that participants in the OUP condition were on average considerably more accurate when estimating the number of bricks necessary in order to complete the structure with the smallest difference between estimated and actual bricks observed until now. Significance is also confirmed by looking at the 95% confidence interval, which does not cross zero and as a consequence shows that the null hypothesis can be rejected when it comes about estimations in terms of resources.

Model 2, reports the regression analysis performed considering the difference between expected and actual time (in seconds) according to the manipulation participants were allocated in. Subjects that received the OUP manipulation on average had a difference between actual and estimated value of 564.25 seconds (743.71 – 179.46) corresponding to 27.4% seconds less than those that received the E3 manipulation. The p-value of this regression model is 0.023. Also in this case, the significance level meets the significance general threshold expectation set at 5%, even if the significance level is higher than the one observed in the previous case when analysing resources estimations. Furthermore, the confidence interval (95%) does not cross the zero value, therefore the null hypothesis of this experiment regarding the time estimations can be rejected.

Model 3, explores the relationship between the difference of expectations in terms of bricks needed in order to complete the task and the “right” number of bricks to complete the task (35) according to the manipulation subjects were placed in. Participants receiving the OUP treatment, were on average 1.43 unit (-11.96 + 13.39) wrong from estimating the right number of bricks or, in other words, 157.2% closer to estimate the “right” number of bricks. This result, is in line with the other results seen until now, showing, that OUP condition had an impact on estimation capabilities of the subjects, making them more accurate. When looking at the significance of the model, a p-value smaller than 0.001 is observed, which confirms the pattern that has been observed in the first two models indicating that the use of an optimism uplift in combination with performing an unpacking task before the estimation task has a major impact in getting more precise estimates.

	Dependent variable:		
	Diff. in bricks (n)	Diff. in time (s)	Diff. with real bricks (n)
	(1)	(2)	(3)
OU + Unpack.	-29.85^{***}	-179.46^{**}	13.39^{***}
	(7.22)	(76.43)	(3.65)
Intercept	31.42^{***}	743.71^{***}	-11.96^{***}
	(5.30)	(56.09)	(2.68)
OU + Unpack. 95 C.I.	(−41.34, −12.35)	(−332.98, −25.94)	(6.01, 20.71)
Observations	52	52	52
R^2	0.22	0.01	0.21
Adjusted R^2	0.20	0.08	0.20
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 30 - Regression models and C.I. experiment 4

7.4.4 Hypothesis and sub hypothesis testing – robustness analysis

As in the case of the previous experiments, considering the fact that there could have been other confounding factors creating some noise in the results reported, such as the age imbalance between the two groups, this variable was plugged into the models explored in Table 30, to see its impact, as showed in Table 32. Moreover, this section offers a few more details on the first three models, including a robustness check using the Mann-Whitney test, in Table 31.

Mann-Whitney tests	
	p-value
Model 1	0.0009 ***
Model 2	0.02 *
Model 3	0.0007 ***

Table 31 - Mann-Whitney tests

Indeed, by considering the distributions of available data as non-normal, I performed Mann-Whitney tests for the variables under scrutiny in order to check for significance also with this method. The reason for this is that as it has been showed by Flyvbjerg et al. 2018, cost overruns, calculated by looking at the difference between estimated and actual resources used in order to complete a project, as in the case of the variables under scrutiny in this research, rarely follow a normal distribution. Indeed, they usually have a fat tailed distribution which indicates that “extreme” values are more likely to appear in the distribution in comparison to a normal distribution. In this specific study, we have seen that extreme values related to over optimism in estimations are much more likely than the opposite, as can be seen also graphically with the figures showing the distributions of difference in time and difference in bricks variables. Therefore, statistically testing those variables following the assumptions of a normal distribution would not yield results useful in order to describe the phenomenon analysed in this dissertation. For this reason, I performed the Mann-Whitney test, that not only is able to account for the non-normality of the distributions under investigation, but is also able to test the hypothesis indicating if there are differences between the two independent populations explored in this experiment (which would be participants in the E3 and OUP condition respectively) (Newbold, 2013), so to support the findings of the regression analysis showed in the previous sections indicating significant differences between the two groups studied. As Table 31 reports, when considering the difference in bricks variable used to build model 1, the p-value is 0.0009, confirming that the two populations are not identical and supporting the findings of the regression analysis. When it comes about Model 2 the p-value observed is at 0.02, once again showing that the two populations are not identical. Model 3 reports a p-value >0.001. These Mann-Whitney tests, therefore, strengthen the results reported with the models in the previous section.

When considering the three models showed beforehand, adding Age as a confounding factor appears to have a slight impact on the result observed. As explained in the context of the first experiment, it seems reasonable to add Age as a confounding factor, because of the online nature of the experiment designed: as a matter of fact, older participants, which might have not been as familiar as younger participants with the usability of the survey and of the 3D game, may have been more conservative in their estimations in order to allow room for errors in case of misunderstanding on how to use the building tool integrated in the 3D game. Younger subjects might have been less conservative in their estimates because of their greater familiarity with online games. Table 32 reports the results of the 3 models presented above adding as an extra confounding variable Age.

	Dependent variable:		
	Diff. in bricks (n)	Diff. in time (s)	Diff. with real bricks(n)
	(4)	(5)	(6)
OU + Unpack.	-29.18^{***}	-203.30^{**}	13.02^{***}
	(7.94)	(84.14)	(4.03)
Age	0.31	3.13	0.05
	(0.42)	(4.51)	(0.22)
Intercept	23.06^{*}	658.49^{***}	-13.28^{**}
	(12.74)	(134.93)	(6.46)
OU + Unpack. 95 C.I.	(−45.14, −13.22)	(−372.39, −34.22)	(4.92, 21.11)
Age 95 C.I.	(−0.55, 1.16)	(−5.92, 12.19)	(−0.38, 0.48)
Observations	52	52	52
R^2	0.22	0.11	0.21
Adjusted R^2	0.19	0.07	0.18
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 32 - Regression models and C.I. with Age as a confounding variable

A detailed discussion of every model will not be done in this occasion, as, no model reports Age as a significant variable. Interestingly, all models for the variable OU & Unpacking are still significant, therefore, they strengthen the findings reported for the first three models on the efficacy of using an optimism uplift and performing an unpacking action before starting out with a project or task.

Since the three models just presented contain more than one confounding variable, all interactions have been explored, which did not yield any significant result. Moreover, all the models were tested for collinearity using VIF (Variance inflation factor), and as can be seen in the results in table 33, no VIF is greater than 5, therefore no collinearity has been found in the models.

Collinearity	
Model	*VIF (reject if VIF > 5)*
4	*Treatment:* 1.04, *Age:* 1.04
5	*Treatment:* 1.04, *Age:* 1.04
6	Treatment: 1.04, Age: 1.04

Table 33 - Collinearity

7.5. Theoretical implications

Experiment 4 explores the potential benefits of integrating two techniques coming from two different and oftentimes deemed incompatible perspectives on how to improve the quality of estimations in terms of resources and time needed to complete a task or a project. As highlighted in this and in chapter 6, strengths and weaknesses of the two approaches might seem to be in general complementary. The first supporting element of this claim I provide with this research, is the high significance of all the coefficients studied through regressions in this chapter. Indeed, they indicate that the null hypothesis of the current experiment can be rejected, hence, combining unpacking with optimism uplift gives better results in terms of forecast precision in respect to a situation where the two approaches are used separately. As a consequence, the use of an “holistic” view on forecasting is likely to have a greater impact in terms of optimism bias mitigation than using separately tools belonging to the inside or the outside view.

Other supporting elements are represented by looking at some of the results’ variables of this experiment: in the case of the total amount of bricks used in order to complete the task of subjects in the OUP compared to those in the E3 condition, for example, we observe that subjects in the OUP condition used

28% less bricks than those in the E3. This is a different result than in the second experiment, where subjects in the OU condition ended up using more resources than those in the E2 control condition. It seems, therefore, that the concurrent use of unpacking with an optimism uplift helped in correcting one of the challenging elements that I highlighted during the analysis of the second experiment. Namely, the fact that using an optimism uplift to mitigate biases arising from estimations might have the same effect that the use of a fudge factor has been showed to have in the context of capital expenditure estimations, studied in the corporate finance field, increasing the final account of resources used in order to complete the project.

Even if this research has a stronger focus on how to improve the quality of estimations in terms of time and resources through the adoption of an holistic view, the corollary of this is the relation that the appraisal process of a project has with its success. This is the reason why, throughout the experiments, I gathered data regarding the expectation of participants towards succeeding or not in completing the task given and if they actually succeeded or not into completing it. In this experiment, the success rate of participants in the experimental condition (OUP) was of 57%, hence, more than half of the participants undertaking the task succeeded when using an optimism uplift and unpacking the task before performing it. Comparing this result with the one from subjects in the control condition, we see how the success rate was considerably lower at 36.4%. Therefore, the difference in success rate between experimental and control condition is of around 21%.

Analysing the difference of success rate between experimental and control condition also for the previous two experiments, we have seen that, in both cases, the mitigation tools used in order to decrease optimism bias during the appraisal process yielded a positive effect in reducing the detrimental effect that it has in terms of project success. Indeed, in both cases a difference of around 5% was observed, with both experimental groups being slightly more successful than their counterparts in the control group. These results, suggest that the concurrent use of the two debiasing techniques had a stronger effect than the use of them singularly.

It seems that the integration of the two techniques lead to a greater rate of success in more than a cumulatively manner. In fact, even if adding up the differences seen in experiments 2 and 3 in success rate of the two experiments between experimental and control group, the result is only half of what we have seen in experiment 4. This result highlights the importance that the estimation process has in increasing the likelihood of a task/project success.

Moreover, it shows how the use of different methods and of an “holistic view” rather than only internal or external view has the potential to enrich the initial knowledge that the decision maker has regarding the task or project at hand, which in turns will impact the quality of its decisions and will influence the overall execution phase, increasing the chances of success. As the emphasis of this research is more on the relationship between optimism bias mitigation tools and forecast precisions, other variables related to project success were not studied, therefore, in order to strengthen these results both from a practical and theoretical perspectives, it is necessary to investigate further the matter, using as preliminary results the one presented in this experiment.

When analysing the variables of this experiment related to forecast precision i.e. difference between estimated and actual bricks by treatment and difference between estimated and actual time by treatment of the subjects in the OUP condition, we notice straight away that those differences are the smallest observed throughout all the conditions of all the experiments. As graphically showed in the boxplot in Figure 32, indeed the IQR of the observation for the variable related to the difference in bricks for participants in the OUP condition is very narrow and the average value is very close to zero, indicating that forecast precision in this group was considerably high.

When comparing this result with the results observed in the previous two experiments, we observe that in terms of forecast precision, subjects in the OU condition were on average more accurate than subjects in the U condition, perhaps for the structural rather than only descriptive nature of the external view technique. For this reason, the expectation in terms of forecasting precision of integrating the two models was that the forecast precision would not have been particularly impacted by adding Unpacking to an Optimism Uplift.

When comparing the IQR of observation of subjects in the OU condition in experiment 2 with those in the OUP condition in the experiment 4, however, we see that forecast precision improved by around 20%. This finding is quite important because, in real terms, an improvement of 20% of forecast precision, would mean that one fifth of the total budget of a project could be virtually saved and this would improve the whole cost efficiency of the project. If this would be achieved by improving the appraisal process during the front-end phase of the project (which is the less cost intensive phase of it), with the integration of two tools analysing different perspectives and features of the project at hand, it would, in turn, have the potential to increase significantly the project performance.

In terms of time estimations versus actual time to complete the task, on the other hand, even if we observe the same pattern highlighted in the previous paragraph, with participants in the OUP condition having the narrowest IQR of observations in relation to any other group of any other experiment, the IQR does not cover the “0s” value. In fact, the average value, is quite distant from the 0s value which would indicate a perfect forecast precision, and definitely more distant than the average value observed in the group of subjects receiving the OU manipulation only. Therefore, even if the range of estimations was minimised with the use of both techniques, forecast precision was somewhat greater when using optimism uplift technique only. This finding, supports the idea presented during the discussion of the previous experiments that there might be different behavioural patterns associated to the elaboration of an estimate in terms of time rather than in resources, hence, more studies in this direction should be done. Indeed, in the current literature, no differentiation between estimations in terms of time and in terms of resources exists, as it is assumed that cognitive biases have the same impact on any type of estimation task. Recognising, studying and defining those differences, might be a further step towards the achievement of greater estimations' efficiency and precision.

Another variable created in order to investigate further forecast precision, is the variable describing the difference between estimated bricks and the “real” number of bricks needed in order to complete the task. Even if such a variable would be impossible to measure for a real-life project, it gives the opportunity to look, net of other factors, the precision of the different techniques used in the experiment in helping the decision maker to understand the ideal level of resources needed in order to complete the task. In other words, this variable could be used to investigate the reliability of the appraisal technique used in order to elaborate the estimates.

In this experiment, the IQR of the observations of participants in the OUP condition, not only is the narrowest encountered in relation to all other conditions and experiments, but its average value is very close to 0. This indicates that, on average, participants in the OUP conditions were able to get the closest estimate in terms of resources to the ideal level needed to complete the task. This finding suggests, once again, that the two techniques together performed better than what they did singularly, supporting the perspective put forward by the theoretical framework of this research i.e. the “holistic view” on forecasting.

As in the case of the difference between estimated and actual bricks variable, we can see that the IQR of participants in the OUP condition covers the 0 value, meaning no discrepancy between estimated and actual resources, as seen for participants in the OU condition but not for participants in the U condition. In this sense, we observe that the integration of the two techniques not only minimised the difference between estimated and actual resources, but it also helped the decision maker to have a clearer idea of what the ideal level of resources to be used in order to complete the task are.

All in all, phase C of the theoretical framework presented in this research, suggests that using an “holistic view” on forecasting, by adopting debiasing techniques both from the internal and external views might lead to more accurate forecasts and this is supported by the findings of experiment 4.

CHAPTER 8. DISCUSSION

8.1. Introduction

In this chapter I summarize the key findings of the experiments presented in the last three chapters, to highlight the connections between them and with the last phase of the theoretical framework presented in Chapter 2. After having discussed through the analysis of all the experiments how these are aimed at addressing the underlying assumptions of Phase A and Phase B, considering also the current context of the project management literature on the matter and the integration theory presented in this research, I now move on to formalize the discussion of Phase C of the theoretical framework (Figure 35).

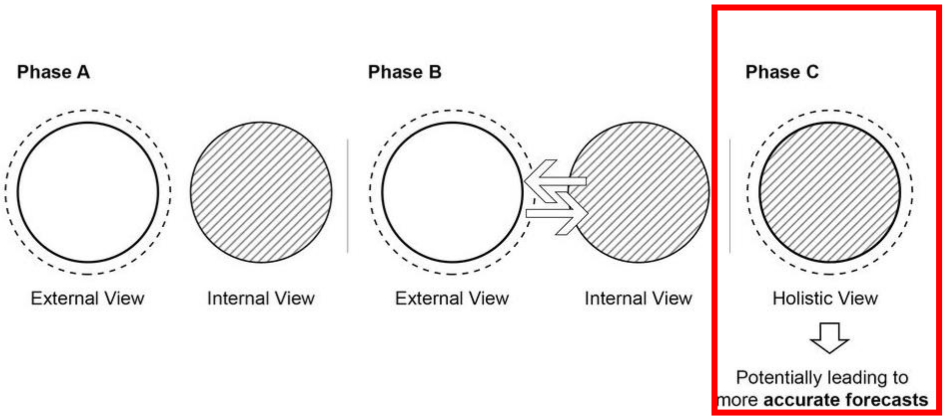


Figure 35 - Integration of the External and Internal view to create a Holistic View on forecasting (by author)

After this, I present power analyses as calculated before doing the experiments but with the means and standard deviations found with the analysis of the experiments run for the sake of this study. Through this, I show if the experiments were overpowered or underpowered, and I provide a useful baseline for researchers wanting to approach this subject in the future with an experimental perspective.

The discussion on the adjusted power calculations, leads to underline some of the issues of the experimental design of this research, that I will use in order to highlight the limitations of this work. The final section is devoted to outlining the future avenues of research for this topic so to highlight once more the importance of studying optimism bias, the planning fallacy and cost/resources/time overruns, with the ultimate aim to improve the current state of project performance and delivery across the Globe.

8.2. Results' discussion

8.2.1. Experiment 1

Experiment 1, was aimed at investigating the relationship between different levels of dispositional optimism and resources' underestimation or overestimation, hypothesizing a positive relationship between higher level of dispositional optimism and resources/schedule underestimation. The importance of this experiment, even if not directly related to any of the phases of the theoretical framework is to analyse the underlying assumptions behind the theory presented in this research: optimism has an impact whenever we need to make an estimation and, most of times, the impact is detrimental. This leads in most cases, to the use of more resources than initially planned which, in turn, is very likely to result in failing to complete the task/project at hand.

This experiment showed that, higher level of dispositional optimism lead participants not only to be less accurate in their estimations but also to use more resources to complete the task with an higher failure rate. These findings are particularly relevant because from one side they validate the theory behind this research and from the other one they offer a deeper understanding of why researching on this matter is so important when applying this reasoning to big infrastructure projects. Indeed, the bigger the project, the bigger is the monetary impact that overoptimism has in the economy of it and the bigger is the impact on the society this project might have, since, in most of the cases, those projects are funded by our taxes. This experiment, therefore, not only shows how overoptimism has an impact during the preliminary planning phase of a task or project but also how this factor reverberates on all deliverables of it affecting all the stages of the project life cycle.

The experiment, moreover, showed that participants in the BPS condition (the manipulation adapted from the positive psychology field used to alter the level of dispositional optimism exhibited by the subjects) were not the only ones being overoptimistic in their estimations, in fact, also subjects allocated to the other condition were inaccurate in their appraisal task. This factor suggests that, as reported by the literature taken as a ground for this research, optimism bias is something present in each and every one of us. Indeed, not considering it whenever investigating issues related to a task or project estimation would mean overlooking at one of the founding behaviours human beings exhibit whenever asked to make future forecasts. In this sense, experiment 1 wants to underline the importance of tackling the subject under investigation from a behavioural perspective.

Finally, this experiment, by using the BPS manipulation, coming from the positive psychology field, informs on the applicability of this manipulation on other contexts than the one of the use of thinking about positive future events as mood improver for the patient. This study, in fact, uses BPS to artificially manipulate the level of dispositional optimism so to appreciate the impact that different levels of optimism have on decision-making, opening the potential use of this manipulation in other experimental fields of social sciences, in line with recent studies published on the topic (Carrillo et al.,2019; Haakerens et al., 2020; Loveday et al. 2018).

8.2.2. Experiment 2

As mentioned in the literature review, few studies have been devoted into understanding the impact of optimism uplifts on estimates' precision. However, Jennings (2012), highlights that uplifts may have a negative effect on the final outcome of the project (in terms of time and cost), as it could legitimize changes during the implementation phase of the project given the "extra" budget applied to the initial forecast. On the other side, a considerable number of governments (among which the UK), following the research output of scholars belonging to the outside view on forecasting have made compulsory to include in infrastructure project forecasts calculations including an optimism uplift based on historical performance of similar projects (HM Treasury, 2013).

In this context, experiment 2, looked at the impact that providing details on historical data for the same task has, in terms of the optimism bias exhibited by previous participants, and if this makes forecasts more precise in relation to actual resources used to complete the task. Findings of the experiment showed that the impact of the optimism uplift seems to be structural rather than aimed at influencing the initial state of mind of the estimator. It could be argued, in fact, that individuals who knew their estimate would get a correction at the end, would adapt their behaviour and make initial estimates that were lower, but this was not the case. This suggests that optimism uplift, unlike the manipulation that was used in experiment 1, does not have an impact on the state of mind of the estimator, but rather, it systematically levels out and mitigates (not eliminates) differences that may arise from different levels of optimism. The fact that optimism uplift seems to have a structural impact on the quality of estimations rather than on the estimator, is an indication of how powerful this tool can be in order to mitigate optimism bias.

If from one side the structural nature of optimism uplift has been proved to have a positive impact in making predictions of the final time and resources used to complete the task more accurate in absolute value, on the other hand, when subjects included an optimism uplift in their estimations, they ended up using, on average, slightly more resources (bricks) than their counterparts not using it. This factor shows that even if optimism uplift might help to get more accurate estimations it might also lead to use more resources than needed in order to complete the task, in line with the experimental sub-hypothesis.

However, the experiment showed that when using an optimism uplift, variability around the true level of resources needed in order to complete the task sharply decreased (more than 50%), making estimations much more accurate, hence, justifying the use of this tool to mitigate optimism bias. This factor represents also a further indication that the manipulation had an impact on the estimation bias by making from one side participants more conservative in their estimations and from the other helping them to get closer to the ideal level of resources needed to complete the task.

This experiment showed how using an optimism uplift has positive and negative sides and that, even if this tool is able to increase forecast precision might have unintended consequences on increasing the resource level used in order to compete the task/project. These factors, therefore, should be considered by governments whenever implementing policies that regulate the use of optimism uplifts as, together with the use of optimism uplifts, they should put in place corrective measures that aim at solving the criticalities of this tool, so to maximise the positive impact that this mitigation tool has.

8.2.3. Experiment 3

Experiment 3 was focused on analysing the impact that an optimism bias mitigating tool coming from the internal view has on forecast precision: unpacking a task in subcategories. Unpacking, unlike the optimism uplift manipulation of the experiment beforehand, does not draw from past distributional similarities in order to mitigate optimism bias in estimation tasks. Unpacking, in fact, takes an internal perspective on the task at hand allowing to analyse it more specifically, emphasizing the different actions to be performed in order to complete the task successfully. Findings of this experiment showed that, in contrast with what was found in the previous experiment, the level of resources used by participants unpacking the task, were considerably lower than their counterpart not doing it. This factor shows that unpacking a task into sub-actions is helpful in order to reduce the final level of resources used in order to complete a task or a project. However, results showed that even if on average participants in the unpacking condition were more accurate in their estimations when compared to the control condition, they were considerably less accurate than participants that received the optimism uplift condition. The results, therefore, suggest that the descriptive nature of the unpacking tool does not result in the same structural effect that optimism uplift has when aiming to mitigate optimism bias, even if unpacking is able to partially "solve" some of the flaws that the use of optimism uplift has.

I argue that even if both techniques have shortcomings, in some cases they are complementary and that they can improve the quality of forecasting. Indeed, optimism uplift is likely to increase the amount of resources used to complete a task, while unpacking does not. However, unpacking is a weaker tool than optimism uplift in addressing the impact of optimism bias in forecast tasks.

8.2.4. Experiment 4

Experiment 4, by following the assumptions behind the "Holistic view" on forecasting presented in the literature review had the aim to practically investigate if the combination of the two different tools, belonging to two different perspectives on forecasting considered in the previous experiments, can yield better results in terms of tackling optimism bias to improve forecast precision.

The results of this experiment show that, participants in the OUP condition (i.e. unpacking the task in subcategories and applying an optimism uplift to their estimations) were on average considerably more accurate when estimating the number of bricks necessary in order to complete the structure with the smallest difference between estimated and actual bricks observed in any other condition of any other experiment in this study. Moreover, with this experiment I showed that when participants unpack their task into different actions and are also given an optimism uplift, not only they feel more confident towards their estimations but they are also more likely to successfully complete a project. The concurrent use of the two tools, moreover, seems to have had not only an impact into improving the efficacy of both techniques into getting more reliable estimates, but also into adjusting some other factors that might have negatively influenced forecast precision. An example of this is given by the fact that unlike previous experiments none of the regression models considering also Age as a confounding factor resulted to be significant. This is another factor that supports the argument of an “holistic view” in forecasting as the concurrent use of unpacking and optimism uplift was able to adjust for the effect that age had in the previous experiments. In this sense with this experiment, I argue that an holistic view fosters the mitigation power of optimism bias in estimations.

Another finding of this experiment is represented by the fact that when comparing measurements of forecast precision between experiments of this study, experiment 4 showed an improvement of around 20% compared to experiment 2, and even more with experiment 3. We have seen, furthermore, that adding the effect of the two tools used separately in experiment 2 and 3, does not give the same result in terms of forecast precision achieved in experiment 4, showing that the concurrent use of the two mitigation tools does not follow an additive property but a compounded one, which represents a strong argument in support of the holistic view.

In this sense, phase C of the theoretical framework presented in this research, suggests that using an “holistic view” on forecasting, by adopting debiasing techniques both from the internal and external views might lead to more accurate forecasts and this is supported by the findings of experiment 4. Analysing the results of experiment 4 however, allowed us to take the reasoning behind this framework even further, by suggesting that not only precisions in forecasts might be more accurate but also that the improved efficiency in forecasting can have a positive impact on the success rate of the task/project at hand, opening new avenues of research on this topic.

Overall, Experiment 4 not only represents the operationalisation of the “holistic view” in forecasting when considering the common task every participant to this study had to complete, but it also gives a general indication on the importance of integrating different techniques coming from different perspectives in forecasting, so to improve not only forecast accuracy but also project performance.

8.3. Adjusted power analysis

The rationale I followed in order to calculate the adjusted power analysis for the experiments I run is the same as the one described in section 4.7 of this dissertation. For this reason, I will not describe it again. In this case, however, instead of the anticipated means and standard deviations for the groups part of the experiment (i.e. control and experimental group) I will use the actual mean and standard deviation found during my experiments.

8.3.1. Experiment 1

Table 34a and 34b report the adjusted power analysis for experiment 1. When calculating sample size using means and standard deviations from the results of the first experiment, we can see that the sample size suggested in order to have statistically significant results is higher than the one assumed in Chapter 4.

Study Parameters	
mean group 1	14.76
σ_1 group 1	21.05
σ_2 group 2	15.14
mean group 2	26.35
α	0.05
β	0.05
Power	0.95
Enrolment ratio	1

Sample size	
Group 1	65
Group 2	65
TOTAL	**130**

Table 34a and 34b - Power analysis experiment 1

We had an early indication that this might have been the case, during the analysis of the results of this experiment, especially when looking at the different observations recorded between time and resource estimations. In fact, when I analysed the first two regression models of the experiment, I argued that when it comes to testing optimism in terms of resources and time there might be different behavioural factors and/or processes that come into play. However, another argument I made was related to the fact that the different patterns observed in the two models, might also have something to do with the fact that the difference in time variable, unlike the difference in bricks variable, is anchored to the maximum amount of time allowed to participants to finish the task (1200 seconds). Therefore, the sensitivity of the model might be altered by this factor, suggesting that reproducing these results using a larger sample might be warranted, as confirmed by the sample size of 130 provided in Table 34b.

Furthermore, when analysing the third regression model for the experiment, looking at the difference between estimated and “right” number of bricks we observed a p-value very close to the 0.05 acceptance threshold but still higher. Therefore, my recommendation in this case was to consider the possibility that the sample for the analysis was too small, and that more studies in this direction should be done, recommendation that is confirmed also in light of the adjusted power analysis provided in this section.

8.3.2. Experiment 2

Table 35a and 35b report the adjusted power calculations for experiment 2. In this case, considering also the results presented in the previous chapters, the expectation was that the preliminary assumptions relating to the power analysis for this experiment were not very far from the actual number needed. This, together with the fact that I showed how powerful optimism uplift is in mitigating optimism uplift, also formed the above-mentioned expectation. Indeed, we can see that, also in this case, when adding means and standard deviations found in this experiment, sample size is slightly larger than the one used in this study.

Study Parameters	
mean group 1	2.65
σ_1 group 1	17.30
σ_2 group 2	16.88
mean group 2	19.01
α	0.05
β	0.05
Power	0.95
Enrolment ratio	1

Sample size	
Group 1	40
Group 2	40
TOTAL	**80**

Table 35a and 35b - Power analysis experiment 2

8.3.2. Experiment 3

Table 36a and 36b report the adjusted power analysis for experiment 3. First thing that should be mentioned is that, unlike the assumptions made before running the experiments, experiment 2 and experiment 3 do not have same sample size. I made the assumption of similar sample size for the two experiments because I found no evidence in the literature that any of the two tools was stronger than the other in improving forecast precision. For this reason, it would have not made sense to test the hypotheses related to the two different tools using a different number of participants, as there would have been no ground to justify that decision.

However, we had an early indication of the fact that this experiment might have been underpowered when I showed that optimism uplift is a stronger tool than unpacking in improving forecast precision and that they also have different natures: one is more structural and the other one is more descriptive, one operates at the estimate level and the other one at the estimator level. Hence, to capture these differences observed through the experiments, it becomes clear that in order to have stronger results using the unpacking manipulation a larger sample size than the one calculated for the optimism uplift manipulation is needed. Those considerations are confirmed by the adjusted calculation with the means and standard deviations found in experiment 3, that report an estimated sample size of 106 participants, which is around 28% more than the one for experiment 2.

Study Parameters	
mean group 1	24.23
σ_1 group 1	27.95
σ_2 group 2	26.56
mean group 2	43.27
α	0.05
β	0.05
Power	0.95
Enrolment ratio	1

Sample size	
Group 1	53
Group 2	53
TOTAL	**106**

Table 36a and 36b - Power analysis experiment 3

8.3.4. Experiment 4

Table 37a and 37b report the adjusted power calculations for experiment 4. Given the results of this experiment, the expectation is that the assumed sample size is not very different from the adjusted sample size. This is also supported by other results and findings of the experiment such as the high significance of the regression models, the complementarity of the two optimism mitigation tools analysed and the fact that this experiment showed that there is a compounded effect in the improvement of forecast precision rather than only an additive effect when using the two tools together. Indeed, adjusted power analysis for this experiment suggests a sample size of 62 participants, which is not very far from the actual sample size used to run experiment 4 in this study.

Study Parameters	
mean group 1	4.6
σ_1 group 1	29
σ_2 group 2	16.37
mean group 2	33.74
α	0.05
β	0.05
Power	0.95
Enrolment ratio	1

Sample size	
Group 1	31
Group 2	31
TOTAL	**62**

Table 37a and 37b - Power analysis experiment 4

8.3.5. Further considerations

According to the adjusted power calculations, the ideal number of participants to the experiments of this research would be of 378. The sample I used to run the experiments was of 231; the reasons behind this difference have been explained during this section, however, the fact that there is a mismatch between the numbers assumed at the beginning of the study and the number resulting for running the experiments is quite normal and does not undermine the results of the experiments, neither poses as a limitation. By running the experiments, we were able to access information that before were not available, an example of this is the different nature of the two forecasting tools and how where they operate in the forecasting context. That new information, have given insights not only on the discussion of the findings in relation to the theory presented in this work but have also provided the basis to continue studying this topic through experiments, as the power calculations in this section can be used in future studies. Since experimental studies in construction project management, as discussed during the literature review, are very scarce, this study might also represent the starting point to begin making research in the field adopting a different perspective.

8.4. Limitations

In this section, I highlight some of the limitations of this research. First, I talk about some general limitations that are connected with the theoretical lenses I chose to analyse the issue of cost/time overruns with. Afterwards, I look at the limitations of my experimental design and what consequences this had in the interpretation of the results.

In Chapter 2, I proposed a conceptual framework based on behavioural economics constructs, which, acknowledging the limits of human rationality also make a point on the difficulty to study patterns related to human behaviour: every forecaster and every decision-maker will adopt different behaviours in different circumstances. In this sense, the researcher approaching this topic following the behavioural perspective, accepts the fundamental limitation that human behaviour is hardly quantifiable and classifiable and that studying behavioural biases may lead to an overgeneralization of behavioural patterns that in reality are only present in a limited number of cases. There is no way to overcome this limitation, however, through the isolation of specific behavioural patterns, those may lead to the formulation of techniques that even if not able to solve completely the behavioural bias might reduce it. The same can be applied to the issue of optimism bias, which is impossible to eliminate but through studies like this one might be able to help in mitigating it in a more efficient way.

To move forward from this limitation, the suggestion is to devote more research efforts in exploring appropriate and innovative laboratory and out of laboratory simulations able to mimic in the best possible way what happens in the real world as I tried to do in more than one occasion with my experimental design. This consideration leads me to the discussion of an intrinsic limitation of experiments: considering the closed environment in which experiments take place, many scholars are worried about the external validity of the results achieved through controlled trials. As in the previous case, this limitation is hard, if not impossible, to overcome; however, as I discussed in Chapter 4, even if acknowledging this limitation, I believe that an experimental approach is the best way to proceed in order to explore the issues of optimism bias in forecasting, not only because no experimental studies on the topic have been made before in the project management field, but also because the controlled environment of the experiment gave me the opportunity to unveil connections between variables that in an observational study would have been impossible to gather.

Besides the structural limitations of the experimental approach, there are also some limitations of my experimental design I would like to discuss.

I would like to highlight that the power calculations are performed only using as a basis projected/actual number for mean and standard deviation relative to the difference in bricks variable and not for the variable related to difference in time. The main reason why I made this choice is because, I gave a time constraint of 20 minutes to participants for them to complete the task, so to make sure that participant would not distract while doing the task. Indeed, if subjects would have had more time to complete the task this could have impacted on their attention level and this could have potentially led to poor results. Moreover, a time cap was introduced, in an effort to mimic what happens in real life infrastructure project: there is a maximum delay (even if very generous, as in the case of the experimental task of this research) that is acceptable for any project, after which the project is declared failed, meaning that, de facto, every project has a time cap, besides the one initially estimated or modified as operations go. Having a time cap, however, had a consequence on the way preliminary calculations and results related to this variable are presented, as people failing in the task all had a task time equal to 1200 seconds, which skewed data towards this value. For this reason, calculating a sample size using the time difference variable would have been potentially misleading as it would not have accounted for the part of the population that would have taken more time than the allowed 1200 seconds in order to complete or fail the task. In this sense, the difference in bricks variable, not presenting a limitation on maximum number of bricks to be used in order to complete the structure, represented a better choice in order to capture most of the population’s values and calculate a more reliable sample size.

This limitation of the difference in time variable can be also seen on the results presented: regression models, even if significant, in some cases, are not as effective as the models related to resource measurements in unveiling the links between the variable under study. Again, this is mainly due to the fact that data are skewed towards a single value; therefore, when looking at time measurement, this factor, together with the ones highlighted during the discussion and analysis of experiments' results should be considered.

Furthermore, because of the insurgence of Covid-19 I had to re-design my experiments from a laboratory setting to an online one; this had an impact on the way experimental task was carried out by the participants. Indeed, initially, subjects should have performed the experimental task with physical building bricks. As most people are used, since a very young age, in the use of building bricks, this would have made the mechanics behind the task not only more familiar to participants but also easier. The introduction of the 3D building game in the experimental design presented the challenge of being less intuitive to use than physical building blocks, with the necessity to provide much more instructions to participants. Moreover, in the case of the online setting, certain participants might have been more comfortable with the use of online tools whereas some others not (I made the case of younger versus older participants), and this had an impact on the result of the experiments, which is hardly quantifiable. One way in which I tried to mitigate this limitation, for example, was choosing to add "Age" as a confounding variable in the regression analyses presented in the robustness checks sections of every experiment. Indeed, not only it was interesting to see what the relationship between age and optimism bias is and if it was significant but controlling the models for this variable was also instrumental in checking for the potential impact that participants' age had in completing or not the experimental task.

Finally, when considering the optimism uplift manipulation administered in experiments 2 and 4, the way it was computed represents an oversimplification of what happens in reality. In fact, in order to calculate the uplifts for projects, different past projects are considered, whereas in my experiment I calculated it considering the same task. This factor influenced the experiments results by altering the effectiveness of using optimism uplifts. On the other side, using the same task to calculate the uplifts gave the opportunity to analyse the impact of those net of any other effect, providing an assessment on the "net-efficacy" of the tool which is a perspective none explored in the literature until this moment.

In line with this last observation, in the next section, I will discuss what are the opportunities in terms of future research that my investigation unveiled, with the aim to provide a research agenda for the field and increase the interdisciplinarity of it with the use of novel methods and research tools.

8.5. Future research

I mentioned in more than one occasion throughout this dissertation, the behavioural perspective in the project management field is gaining more and more ground, because both academics and practitioners started to emphasise the centrality of human behaviour in any kind of endeavour. However, unlike other fields of research, the behavioural account is still heavily criticised, perhaps given the mechanical view of project management that has been put forward since the very first studies in the field, where project management tools were considered more important than phenomena such as how behaviour of stakeholders impacts the overall project life cycle. In this sense, the first recommendation I would like to make in terms of future research, is for academics in the field to start considering more the behavioural perspective and how this impacts all stages of construction projects, especially when looking at complex and stakeholders-rich projects as the infrastructure ones. These studies might bring not only new knowledge, but also important savings that, especially for infrastructure projects would benefit the whole society.

Furthermore, with this study, I introduced a conceptual framework based on the current research in the project management field of cost and resource forecasting, that I called the "holistic view". With the Holistic view, I aim to reconcile the ongoing debate of two perspectives, the internal and external views, that are deemed incompatible, by posing that they represent two faces of the same medal. In this sense, I suggest developing further studies that look at the efficacy of the holistic view over the external versus internal account. These studies have not only the potential to further strengthen the results coming from the current research, but they also give the opportunity to re-assess the position of leading academics in the field emphasising one perspective over the other.

Studies supporting the holistic view account should also be aimed at analysing how to operationalise this perspective in terms of policies implementation. In fact, even if with this work I analysed the current policy landscape of those countries including in their regulations for public infrastructure projects innovative forecasting tools, those are mainly the fruit of research belonging to the external view, so I did not look at what specific sets of policies might be implemented in order to capitalize on the strengths deriving from the adoption of a holistic view approach. Also, I suggest starting a dialogue with governments and infrastructure projects' practitioners in order to understand what the impact of the implementation of certain policies would be to mitigate optimism bias using the holistic view as a framework. This would provide the opportunity from one side to improve the knowledge on the field of research and from the other to start a journey towards the formalisation and implementation of those policies, which might eventually lead to important savings both in terms of money and of time. In order to do this, studies on holistic view should involve also the adoption of other tools besides the ones I used in this study, also to see if the findings of this study are applicable to other tools belonging to the external or internal perspective. Finally, developing more research dedicated to further understand the complementarity of those tools, which I stressed when discussing the findings of this study, is another opportunity to further strengthen the theoretical and practical basis for the adoption of a holistic view.

Complementarity, however, was not the only property I found when investigating the integration of unpacking and optimism uplift. In fact, even if adding up the differences seen in experiments 2 and 3 in success rate of the two experiments between experimental and control group, the result is only half of what we have seen in experiment 4. It seems, therefore, that the integration of the two techniques lead to a greater rate of success in more than a cumulatively manner. As the emphasis of this research is more on the relationship between optimism bias mitigation tools and forecast precisions, other variables related to project success were not studied, therefore, in order to strengthen these results both from a practical and theoretical perspectives, it is necessary to investigate further the matter, using as preliminary results the one I presented in this research.

When considering the experimental method I adopted in the context of this research, unveiling connections and properties of the forecasting tools that before were not explored because deemed incompatible rather than complementary has given me the opportunity to pave the way for future research in this field, where experimental studies are very scarce. As I also specified in the literature review, experimental studies in the project management field are almost non-existent. With this research, I wanted to highlight the importance of using this method also in this field, so to find causal connections between phenomena that until this moment are hardly explained in our discipline. Experimental research in project management, has the potential to offer a different perspective on the current challenges the field is facing. The use of experiments would also give the chance to researchers to use innovative techniques in order to gather data, that might provide unprecedented insights on the subject. In this research, for example, I used a brick-building game in order to gain insights on the phenomenon of optimism bias and cost and time overruns, that would have been much harder (and longer) to gather if real-life complex projects would have been analysed instead. Also, there would have been the risk to pollute my analysis with many other factors that would have been hardly controllable considering the extreme complexity of infrastructure projects. The use of the experimental method gave me to opportunity to specifically analyse the behavioural trait I was interested in and make a connection with the infrastructure management context. In the same way, other studies focused at exploring a multitude of stakeholder's decisions or behaviours at any stage of the project life cycle might be done and this would provide substantial new contribution to the field.

Furthermore, in my study I run the experiments online. Developing comparative research reproducing the same experiments in a laboratory setting would give the opportunity to assess the behavioural differences of participants performing the same task in two different environments, which may also contribute to informing about the effectiveness of both online and laboratory experiments in this field.

When looking at the variables studied during my experiments, in more than one occasion I expressed how there might be different behavioural patterns associated with different kinds of estimations (I made the case of resources versus time estimations). I gave a possible interpretation, suggesting that those patterns might be due to the fact that the estimation of resources are by definition more “tangible” than the idea of “time”. In the experiments, furthermore, I found that when altering the level of dispositional optimism in experiment 1 the treatment variable related to time was not significant whereas when using optimism uplift in experiment 2 it was. This might mean that when subjects are guided towards the adoption of a structural change rather than being influenced at being more or less precise, they are more likely to get time estimates right. In this factor might lie another differentiation between estimation of time and estimation of resources. However, to my knowledge, no study in this direction has ever been made.

Devoting more studies into the understanding of the different behaviours affecting people when asked to estimate for different things, might be an instrumental step to implement specific tools to mitigate the impact of optimism bias in a more effective way. Indeed, in the current literature, no differentiation between estimations in terms of time and in terms of resources exists, assuming that optimism bias, and more generally, cognitive biases, have the same impact on any type of estimation task. Recognising, studying and defining those differences, might be a further step towards the achievement of greater estimations' efficiency and precision.

During experiment 1, I looked at how a higher level of dispositional optimism leads people to be more optimist in their forecasts. However, few investigations have been devoted to the understanding of the actual impact that different levels of dispositional optimism can have not only during the estimation phase of a project but also on its final count of resources and time. Therefore, studies providing results in terms of forecast accuracy considering different levels of optimism might be a good opportunity to investigate further in the matter and understand if it is possible to adjust mitigation techniques based on the level of dispositional optimism of the forecaster, so to have a tailored approach for each occasion.

When analysing the experiments, moreover, I decided to use Age as a confounding variable in my robusteness check not only for the reasons explained in the limitations section but also because considering other variables specifically related to the estimator might give interesting insights in how those are releted with forecast accuracy. In more than one occasion, I showed how age was a significant factor in my models when looking at forecast accuracy, therefore, more studies should be made in this direction. Further studies should not only be dedicated to the impact that age has on forecast accuracy, but also considering other variables, such as sex, occupation or ethnicity.

Overall, this study provides a rich basis for developing further research that are not only focused on the specific subject of optimism bias but that can also give direction to studies belonging to other areas of decision-making considering the behavioural perspective and to the project management field on the use of innovative methods, such as the one I presented with this research, so to foster contributions with relevant theoretical, practical and policy implications.

CHAPTER 9. CONCLUSIONS AND POLICY IMPLICATIONS

With this research I looked at the following research question:

> "*To what extent the creation of a holistic model embedding the inside and outside views in forecasting can improve current policies and practices aimed at mitigating optimism bias in infrastructure projects?"*

In order to answer this question, I introduced the concepts related to the decision-making process according to behavioural economics to analyse the idea of cognitive biases and understand their impact on decision-making processes. After that, I introduced the concept of the planning fallacy and optimism bias, following a behavioural economics perspective and I looked at two theories, prospect and support theory that aim at mitigating the effect of the planning fallacy and the optimism bias arising from it, but emphasise two different sides of the problem, the inside and the outside view. After discussing the two theories I proposed a new conceptual framework, to be used as a ground for further analyses and improvement of current forecasting methods that I named the "Holistic view".

Thereafter, I defined optimism bias when looking it under an infrastructure project management's perspective as the deceptive formulation of appraisals, given by the delusional optimism in regard to the attributes of the iron triangle (cost, quality and time) from one side and an excessive optimism in terms of stakeholders' capabilities during the project life cycle, from the other. I also discussed how planners and project promoters, tend to overvalue positive outcomes coming from the envisioned success of the project, to oversimplify project's activities and not to focus on potential risks. Consequently, promoters will tend to undertake projects that are unlikely to have the benefits planned at the appraisal stage that in many cases will lead to a situation of cost overruns and/or delays.

After having established the links between the theory discussed and the infrastructure project management field, I analysed in more detail the current perspectives and lenses with which the issue of optimism bias is studied in the literature, focusing on infrastructure projects and the forecasting methods that informed some of the policies implemented by countries in order to mitigate the impact of optimism bias in the appraisal of large projects.

Subsequently, I introduced the chosen method, online experiments, and I explained the design of the four experiments I created, as well as describing the online platform and 3D web game I developed in order to administer the experiments to the over 230 participants to the study.

Experiments were all designed for a specific purpose: Experiment 1 looked at the relationship between level of dispositional optimism and forecast precision; the main finding of this experiment is that a higher level of dispositional optimism negatively influences forecast precision. This has helped to support and validate the behavioural lenses I decided to adopt in order to analyse the issue of optimism bias.

Experiment 2 and experiment 3 each looked at what is the impact of the use of optimism uplift (optimism bias mitigation tool coming from the external perspective) and unpacking (mitigation tool coming from the internal view) on forecast precision. The results I presented, highlighted the usefulness of both approaches in tackling certain challenges that represent an issue whenever an estimation about a task or project needs to be elaborated. These results, moreover, by indicating strength and weaknesses of those approaches, emphasised the different nature of the two tools: optimism uplift being a more structural tool operating directly on the estimation and unpacking being a more descriptive tool operating on the estimator. In addition, this analysis indicated that, under some conditions, those tools can be considered complementary. The experiments, therefore, indicated how we should approach internal and external view as two faces of the same medal both theoretically and practically, in line with what I suggest with my research question.

Experiment 4 looked at the impact on forecast precision of using the tools used in the previous two experiments together. Results showed that forecast precision, when looking at resources estimations, improved of around 20% in respect to a situation when only optimism uplift is used. This finding, translated in real terms, means that one fifth of the total budget of a project could be virtually saved, improving the whole cost efficiency of the project. If this would be achieved by improving the appraisal process during the front-end phase of the project (which is the least cost intensive phase of it), with the integration of two tools analysing different perspectives and features of the project at hand, it would, in turn, have the potential to increase significantly the project performance. The experiment revealed that using both techniques at the same time improved the quality of the forecast in a compounded rather than only cumulative manner, pointing at yet another important benefit of using a holistic view that integrates both the internal and external perspective. All those factors are pivotal when considering which strategies to use in order to mitigate optimism bias in estimations, especially when appraising mega projects, where the potential waste of resources could end up costing millions to project promoters and taxpayers (in the case of public infrastructure projects). Therefore, the experiments showed that using only techniques coming from the external or internal view might be detrimental not only as an estimation strategy for a project but also for its deliverables in terms of time and resources. Moreover, results revealed how the use of different methods and of a “holistic view” has the potential to enrich the initial knowledge that the decision-maker has regarding the task or project at hand, which in turns will impact the quality of its decisions and will influence the overall execution phase, increasing the chances of success.

Finally, taking into account all the considerations above, with this work, I showed that academia, practitioners and governments should consider the adoption of a holistic view in order to improve the current state of forecasting research, regulations and policies, as the benefits arising from the operationalisation of this perspective would be numerous and not only related to a monetary cost/time efficiency, as it is currently believed.

REFERENCES

Ahiaga-Dagbui, D.D. and Smith, S.D., 2014. Rethinking construction cost overruns: cognition, learning and estimation. *Journal of Financial Management of Property and Construction*, 19(1), pp.38-54.

Aibinu, A.A. and Pasco, T., 2008. The accuracy of pre-tender building cost estimates in Australia. *Construction Management and Economics*, 26(12), pp.1257-1269.

Alacevich, M., 2014. Visualizing uncertainties, or how Albert Hirschman and the World Bank disagreed on project appraisal and what this says about the end of “high development theory”. *Journal of the History of Economic Thought*, 36(2), pp.137-168.

Allais, M., 1953. Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica: Journal of the Econometric Society*, pp.503-546.

Amadi, A.I. and Higham, A., 2017. Latent geotechnical pathogens inducing cost overruns in highway projects. *Journal of Financial Management of Property and Construction*, *22*(3), pp.269-285.

Anderson, S.D. and Tucker, R.L., 1994. Improving project management of design. *Journal of Management in Engineering*, 10(4), pp.35-44.

Annema, J.A., Koopmans, C. and Van Wee, B., 2007. Evaluating transport infrastructure investments: The Dutch experience with a standardized approach. *Transport Reviews*, *27*(2), pp.125-150.

Annual report on Major projects 2020 to 2021. (16/02/2022): https://www.gov.uk/government/publications/infrastructure-and-projects-authority-annual-report-2021/annual-report-on-major-projects-2020-to-21

Ansar, A., Flyvbjerg, B., Budzier, A. and Lunn, D., 2014. Should we build more large dams? The actual costs of hydropower megaproject development. *Energy Policy*, 69, pp.43-56.

Babbie, E.R., 2015. The practice of social research. Nelson Education.

Baddeley, M., 2013. *Behavioural Economics and Finance*. London: Routledge Ltd

Belyanin, A., 2003. Daniel Kahneman and Vernon Smith: Nobel Prize for the Feeling of Reality (Economic Analysis of Human Behaviour). Economic Issues, (1), p.4.

Bentham, J., 1823 'An Introduction to the Principles of Morals and Legislation,' Reprinted in: N.A. Page (ed.), *Utility Theory. A Book of Readings*, New York.

Betselier, J. and Vanhoucke, M., 2016. Practical application and empirical evaluation of reference class forecasting for project management. *Project Management Journal*, 47(5), pp.36-51.

Beukers, E., Bertolini, L. and Te Brömmelstroet, M., 2012. Why Cost Benefit Analysis is perceived as a problematic tool for assessment of transport plans: A process perspective. *Transportation Research Part A: Policy and Practice*, 46(1), pp.68-78.

Blaikie, N. and Priest, J., 2017. Social research: Paradigms in action. John Wiley & Sons.

Blaikie, N., 2007. Approaches to social enquiry: Advancing knowledge. Polity.

Bordat, C., McCullouch, B.G., Labi, S. and Sinha, K.C., 2004. An analysis of cost overruns and time delays of INDOT projects.

Bordley, R.F., 2014. Reference class forecasting: resolving its challenge to statistical modeling. *The American Statistician*, 68(4), pp.221-229.

Borgatta, E.F. and Bohrnstedt, G.W., 1974. Some limitations on generalizability from social psychological experiments. Sociological Methods & Research, 3(1), pp.111-120.

Brealey, R.A., Myers, S.C., Allen, F. and Mohanty, P., 2012. *Principles of corporate finance.* Tata McGraw-Hill Education.

Bruzelius, N., B. Flyvbjerg, W. Rothengatter. (2002). Big decision, big risks. Improving accountability in mega projects. *Transport Policy,* Vol. 9, no. 2, pp. 143–154.

Bryman, A. (2012). Social research methods. Oxford university press. 3rd edn. New York.

Buehler, C., Anthony, C., Krishnakumar, A., Stone, G., Gerard, J. and Pemberton, S., 1997. Interparental conflict and youth problem behaviors: A meta-analysis. *Journal of Child and family studies*, *6*(2), pp.233-247.

Buehler, R., Griffin, D. and Peetz, J., 2010. The planning fallacy: Cognitive, motivational, and social origins. In *Advances in experimental social psychology* (Vol. 43, pp. 1-62). Academic Press.

Buehler, R., Griffin, D. and Ross, M., 1994. Exploring the" planning fallacy": Why people underestimate their task completion times. *Journal of personality and social psychology*, *67*(3), p.366.

Burrell, G., 1979. Morgan. G.(1979) Sociological paradigms and organizational analysis. London: I-Ieinemann.

Candy, P.C. (1989). ‘Alternative paradigms in educational research’. The Australian Educational Researcher, 16(3), 1-11.

Cantarelli, C.C., Flybjerg, B., Molin, E.J. and Van Wee, B., 2010. Cost overruns in large-scale transportation infrastructure projects: Explanations and their theoretical embeddedness. *European Journal Transportation Infrastructure Research*, 10 (1), pp. 5-18

Cantarelli, C.C., Molin, E.J., van Wee, B. and Flyvbjerg, B., 2012. Characteristics of cost overruns for Dutch transport infrastructure projects and the importance of the decision to build and project phases. *Transport Policy,* 22, pp.49-56.

Carrillo, A., Rubio-Aparicio, M., Molinari, G., Enrique, Á., Sánchez-Meca, J. and Baños, R.M., 2019. Effects of the Best Possible Self intervention: A systematic review and meta-analysis. *PloS one*, *14*(9).

Cassar, G., 2010. Are individuals entering self-employment overly optimistic? An empirical test of plans and projections on nascent entrepreneur expectations. *Strategic Management Journal*, *31*(8), pp.822-840.

Collis, J. and Hussey, R., 2013. Business research: A practical guide for undergraduate and postgraduate students. Macmillan International Higher Education.

Credé, M. (2010). Random responding as a threat to the validity of effect size estimates in correlational research. Educational and Psychological Measurement, 70, 596–612.

Creswell, J. (2009). Research design: Qualitative, quantitative, and mixed methods approaches (3rd ed.). Thousand Oaks, CA: Sage.

Croson, R. and Gächter, S., 2010. The science of experimental economics. Journal of Economic Behavior & Organization, 73(1), pp.122-131.

Denicol, J., Davies, A. and Krystallis, I., 2020. What are the causes and cures of poor megaproject performance? A systematic literature review and research agenda. *Project Management Journal*, *51*(3), pp.328-345.

Dewey, J., 2008[1920]. Reconstruction in philosophy. In J. Boydston & R. Ross (Eds.), The middle works of John Dewey, 1899-1994 (Vol. 12, pp. 77-202). Carbondale: Southern Illinois University Press. (Original work published 1920)

Easterby-Smith, M., Thorpe, R. and Jackson, P.R., 2012. Management research. Sage.

Elkjaer, B. and Simpson, B., 2011. Pragmatism: A lived and living philosophy. What can it offer to contemporary organization theory?. In Philosophy and organization theory (pp. 55-84). Emerald Group Publishing Limited.

Falk, A. and Heckman, J.J., 2009. Lab experiments are a major source of knowledge in the social sciences. science, 326(5952), pp.535-538.

Fischoff, B. and Lichtenstein, S., 1978. Don't attribute this to Reverend Bayes. *Psychological Bulletin*, *85*(2), p.239.

Flyvbjerg, B. and Bester W., 2021, "The Cost-Benefit Fallacy: Why Cost-Benefit Analysis Is Broken and How to Fix It," *Journal of Benefit-Cost Analysis*, October, pp. 1-25.

Flyvbjerg, B. and Cowi, 2004 (c). *Procedures for dealing with optimism bias in transport planning: Guidance document.*

Flyvbjerg, B. and Sunstein, C.R., 2016. The principle of the malevolent hiding hand; or, the planning fallacy writ large. *Social Research: An International Quarterly*, *83*(4), pp.979-1004.

Flyvbjerg, B., 2005. Design by deception: The politics of megaproject approval. *Harvard Design Magazine, Spring/Summer,* (22), pp.50-59.

Flyvbjerg, B., 2006. From Nobel Prize to project management: Getting risks right. *Project management journal*, 37(3), pp.5-15.

Flyvbjerg, B., 2008. Curbing optimism bias and strategic misrepresentation in planning: Reference class forecasting in practice. *European Planning Studies*, 16(1), pp. 3-21

Flyvbjerg, B., 2009. Survival of the unfittest: why the worst infrastructure gets built—and what we can do about it. *Oxford review of economic policy*, 25(3), pp.344-367.

Flyvbjerg, B., 2013 (a). How planners deal with uncomfortable knowledge: The dubious ethics of the American Planning Association. *Cities*, 32, pp.157-163.

Flyvbjerg, B., 2013 (b). Quality control and due diligence in project management: Getting decisions right by taking the outside view. *International Journal of Project Management*, 31(5), pp.760-774.

Flyvbjerg, B., 2016. The Fallacy of Beneficial Ignorance: A Test of Hirschman's Hiding Hand. *World Development*, *84*, pp.176-189.

Flyvbjerg, B., Ansar, A., Budzier, A., Buhl, S., Cantarelli, C., Garbuio, M., Glenting, C.,

Holm, M.S., Lovallo, D., Lunn, D. and Molin, E., 2018. Five things you should know about cost overrun. *Transportation Research Part A: Policy and Practice*, *118*, pp.174-190.

Flyvbjerg, B., Cantarelli, C., Molin, E. J., & van Wee, B., 2010. Cost Overruns in Large-scale Transportation Infrastructure Projects: Explanations and their Theoretical Embeddedness. European Journal of Transport Infrastructure Research, 10(1), pp. 5-18

Flyvbjerg, B., Garbuio, M. and Lovallo, D., 2009. Delusion and deception in large infrastructure projects: two models for explaining and preventing executive disaster. *California management review*, 51(2), pp.170-194.

Flyvbjerg, B., Glenting, C. and Rønnest, A.K., 2004 (a). Procedures for dealing with optimism bias in transport planning. *London: The British Department for Transport, Guidance Document*

Flyvbjerg, B., Holm, M.S. and Buhl, S., 2002. Underestimating costs in public works projects: Error or lie?. *Journal of the American planning association*, 68(3), pp.279-295.

Flyvbjerg, B., Holm, M.S. and Buhl, S.L., 2003. How common and how large are cost overruns in transport infrastructure projects?. *Transport reviews*, *23*(1), pp.71-88

Flyvbjerg, B., Skamris Holm, M.K. and Buhl, S.L., 2004(b). What causes cost overrun in transport infrastructure projects?. *Transport reviews*, 24(1), pp.3-18.

Flyvbjerg, Bent, Holm M.S. , and Søren L. Buhl, 2005. How (In)accurate are Demand Forecasts in Public Works Projects? The Case of Transportation. *Journal of the American Planning Association*, vol. 71(2), Spring, pp. 131-146

Fosnaugh, J., Geers, A. L., & Wellman, J. A., 2009. Giving off a rosy glow: The manipulation of an optimistic orientation. The Journal of Social Psychology, 149(3), 349-364.

Frederiksen, L. and Davies, A., 2008. Vanguards and ventures: Projects as vehicles for corporate entrepreneurship. *International Journal of Project Management*, *26*(5), pp.487-496.

Gigerenzer, G. and Selten, R., 2002. *Bounded rationality: The adaptive toolbox*. MIT press

Gigerenzer, G. and Todd, P.M.,2007 . Environments that make us smart: Ecological rationality. *Current directions in psychological science*, *16*(3), pp.167-171.

Gil, N. and Lundrigan, C., 2012. The leadership and governance of megaprojects. *Manchester: Manchester University*.

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

Global Infrastructure Outlook, 2016. Infrastructure investment needs 50 countries, 7 sectors to 2040. *Oxford Economics*.

Grant-Muller, S.M., Mackie, P., Nellthorp, J. and Pearman, A., 2001. Economic appraisal of European transport projects: the state-of-the-art revisited. *Transport Reviews*, *21*(2), pp.237-261.

Greene, J.C. and Hall, J.N., 2010. Dialectics and pragmatism: Being of consequence. Handbook of mixed methods in social and behavioral research, pp.119-144.

Greisdorf H., "Relevance thresholds: a multi-stage predictive model of how users evaluate information," Information Processing & Management, vol. 39, pp. 403, 05. 2003.

Guba, E.G. and Lincoln, Y.S., 1994. Competing paradigms in qualitative research. Handbook of qualitative research, 2(163-194), p.105.

Hansson, B., 1975. The appropriateness of the expected utility model. *Erkenntnis*, *9*(2), pp.175-193.

Heekerens, J.B. and Eid, M., 2021. Inducing positive affect and positive future expectations using the best-possible-self intervention: A systematic review and meta-analysis. *The Journal of Positive Psychology*, *16*(3), pp.322-347.

Heekerens, J.B., Eid, M. and Heinitz, K., 2020. Dealing with conflict: Reducing goal ambivalence using the best-possible-self intervention. *The Journal of Positive Psychology*, *15*(3), pp.325-337.

Helson, H., 1964. Adaptation-level theory: an experimental and systematic approach to behaviour, New York: Harper & Row

Hinze, Jimmie, Selstead, Gregory A., 1991. Analysis of WSDOT Construction Cost Overruns. Washington State Department of Transportation, Olympia, WA.

Hirschman, A.O., 1967. The principle of the hiding hand. *The public interest*, 6, p.10.

Hirschman, Albert O. [1967] 2015. Development Projects Observed, 3rd ed. (Brookings Classic), with new foreword by Cass R. Sunstein and new afterword by Michele Alacevich. Washington, DC: Brookings Institution.

Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.

HM Treasury, 2013. *Supplementary Green Book Guidance: Optimism Bias.* London.(16/02/2022): https://www.gov.uk/government/publications/green-book-supplementary-guidance-optimism-bias.

H.M. Treasury, 2016. National Infrastructure Delivery Plan 2016-2021. *Funding and Finance Supplement.* (16/02/2022): https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/520086/2904569_nidp_deliveryplan.pdf

Holden, M.T. and Lynch, P., 2004. Choosing the appropriate methodology: Understanding research philosophy. The marketing review, 4(4), pp.397-409.

Holm, M.S. and Buhl, S., 2002. Underestimating costs in public works projects: Error or lie?. *Journal of the American planning association*, *68*(3), pp. 279-295.

Honig, D., 2018. *Navigation by judgment: Why and when top down management of foreign aid doesn't work*. Oxford University Press.

Howe, K.R., 1988. Against the quantitative-qualitative incompatibility thesis or dogmas die hard. Educational researcher, 17(8), pp.10-16.

Hutton, J., 2019. Cost Overruns of major government projects. *TaxPayers' Alliance*.16/02/2022: https://d3n8a8pro7vhmx.cloudfront.net/taxpayersalliance/pages/16679/attachments/original/1566302500/Cost_overruns_of_major_government_projects.pdf?1566302500

Invernizzi, D.C., Locatelli, G. and Brookes, N., 2018. Cost overruns–helping to define what they really mean. *Proceeding of Civil Engineering – Civil Engineering*.

Jacquemet N. and L'Haridon O., 2018. EXPERIMENTAL ECONOMICS: Method and Applications. Cambridge University Press.

Jadhav, P., Desai, D. and Gupta, A., 2016. Analysis of construction cost overrun causes-contractor's view. *Imperial Journal of Interdisciplinary Research*, *2*(8), pp.908-910.

Jennings, W., 2012. Why costs overrun: risk, optimism and uncertainty in budgeting for the London 2012 Olympic Games. Construction Management and Economics, 30(6), pp.455-462.

Ji, S.H., Park, M. and Lee, H.S., 2011. Cost estimation model for building projects using case-based reasoning. *Canadian Journal of Civil Engineering*, 38(5), pp. 570-581.

Kahneman, D. and Tversky, A., 1972. Subjective probability: a judgment of representativeness. *Cognitive Psychology* 3,430-54**.**

Kahneman, D. and Tversky, A., 1979. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, pp. 263-291.

Kahneman, D. and Tversky, A., 1996. On the reality of cognitive illusions.

Kahneman, D. and Tversky, A., 2000. *Choices, values, and frames*. Cambridge University Press.

Kahneman, D., 2011. *Thinking, fast and slow*. Macmillan.

Kahneman, D., Lovallo, D. and Sibony, O., 2011. Before you make that big decision. *Harvard business review*, 89(6), pp.50-60.

Kapteyn, A., 1985. Utility and economics. *De Economist*, *133*(1), pp.1-20.

Kazdin, A.E., 1978. Methodological and interpretive problems of single-case exper

Kelemen, M.L. and Rumens, N., 2008. An introduction to critical management research. Sage.

Kim, B.C. and Reinschmidt, K.F., 2011. Combination of project cost forecasts in earned value management. *Journal of Construction Engineering and Management*, 137(11), pp.958-966.

King, L. A., 2001. The health benefits of writing about life goals. Personality and Social Psychology Bulletin, 27, 798–807.

Koch, C., 2012. Contested overruns and performance of offshore wind power plants. *Construction Management and Economics*, 30(8), pp.609-622.

Kruger, J. and Evans, M., 2004. If you don't want to be late, enumerate: Unpacking reduces the planning fallacy. *Journal of Experimental Social Psychology*, *40*(5), pp.586-598.

Kutsch, E. and Hall, M., 2010. Deliberate ignorance in project risk management. *International journal of project management*, 28(3), pp.245-255.

Leavitt, D., Ennis, S. and McGovern, P., 1993. The cost escalation of rail projects: Using previous experience to re-evaluate the calspeed estimates.

Lee, J.K., 2008. Cost overrun and cause in Korean social overhead capital projects: Roads, rails, airports, and ports. *Journal of Urban Planning and Development*, 134(2), pp.59-62.

Lefley, F., 2013. The appraisal of ICT and non-ICT capital projects: a study of the current practices of large UK organisations. *International Journal of Managing Projects in Business*, 6(3), pp.505-533.

Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M. and Combs, B., 1978. Judged frequency of lethal events. *Journal of experimental psychology: Human learning and memory*, *4*(6), p.551.

Lincoln, Y.S., Lynham, S.A. and Guba, E.G. (2011) Paradigmatic Controversies, Contradictions, and Emerging Confluences, Revisited, in Denzin N.K. and Lincoln Y.S. (eds.) Sage Handbook of Qualitative Research. London: Sage Publications.

Liu, L. and Napier, Z., 2010. The accuracy of risk-based cost estimation for water infrastructure projects: preliminary evidence from Australian projects. *Construction Management and Economics*, 28(1), pp.89-100.

Liu, X., Stoutenborough, J. and Vedlitz, A., 2017. Bureaucratic expertise, overconfidence, and policy choice. *Governance*, 30(4), pp.705-725.

Locatelli, G., Invernizzi, D.C. and Brookes, N.J., 2017. Project characteristics and performance in Europe: An empirical analysis for large transport infrastructure projects. *Transportation research part A: policy and practice*, 98, pp.108-122.

Lovallo, D. and Kahneman, D., 2003. Delusions of success. *Harvard business review*, *81*(7), pp.56-63.

Lovallo, D. and Sibony, O., 2010. The case for behavioral strategy. *McKinsey Quarterly*, 2(1), pp.30-43.

Lovallo, D., Clarke, C. and Camerer, C., 2012. Robust analogizing and the outside view: two empirical tests of case-based decision making. *Strategic Management Journal*, *33*(5), pp.496-512.

Love, P., Davis, P., Ellis, J. and On Cheung, S., 2010. Dispute causation: identification of pathogenic influences in construction. *Engineering, Construction and Architectural Management*, *17*(4), pp.404-423.

Love, P.E. and Ahiaga-Dagbui, D.D., 2018. Debunking fake news in a post-truth era: The plausible untruths of cost underestimation in transport infrastructure projects. *Transportation Research Part A: Policy and Practice*, *113*, pp.357-368.

Love, P.E., Ahiaga-Dagbui, D.D. and Irani, Z., 2016. Cost overruns in transportation infrastructure projects: Sowing the seeds for a probabilistic theory of causation. *Transportation research part A: policy and practice*, 92, pp.184-194.

Love, P.E., Edwards, D.J. and Irani, Z., 2011. Moving beyond optimism bias and strategic misrepresentation: An explanation for social infrastructure project cost overruns. *IEEE transactions on engineering management*, *59*(4), pp.560-571.

Love, P.E., Sing, C.P., Wang, X., Irani, Z. and Thwala, D.W., 2014. Overruns in transportation infrastructure projects. *Structure and Infrastructure Engineering*, *10*(2), pp.141-159.

Love, P.E., Sing, M.C., Ika, L.A. and Newton, S., 2019. The cost performance of transportation projects: The fallacy of the Planning Fallacy account. *Transportation Research Part A: Policy and Practice*, *122*, pp.1-20.

Love, P.E., Smith, J., Simpson, I., Regan, M. and Olatunji, O., 2015. Understanding the landscape of overruns in transport infrastructure projects. *Environment and Planning B: Planning and Design*, 42(3), pp.490-509.

Loveday, P.M., Lovell, G.P. and Jones, C.M., 2018. The best possible selves intervention: A review of the literature to evaluate efficacy and guide future research. *Journal of Happiness Studies*, *19*(2), pp.607-628.

Lynch, B.D., 2019. What Hirschman's hiding hand hid in San Lorenzo and Chixoy. *Water,* 11(3), p.415.

MacDonald, M., 2002. Review of large public procurement in the UK. Study commissioned by HM treasury. Retrivied from internet the 16/02/2022: http://elibrary.lt/resursai/ES/Leidiniai/green_books/Archive/greenbook_mott.pdf.

Marczyk, G., DeMatteo, D. and Festinger, D., 2005. Essentials of research design and methodology. John Wiley & Sons Inc.

Markowitz, H., 1952. Portfolio selection. *The journal of finance*, 7(1), pp.77-91.

Merrow, E.W., 2011. Industrial Megaprojects: Concepts, Strategies and Practices for Success 1st ed. John Wiley & sons, ed., Cambridge University Press.

Meyer, W.G. (2014), “The effect of optimism bias on the decision to terminate failing projects”, *Project Management Journal*, Vol. 45 No. 4, pp. 7-20.

Mills, J., Shilson, S., Woodley, Q. and Woodwark, A., 2011. *Keeping Britain moving-The United Kingdom’s transport infrastructure needs*, Mckinsey and Company, 2011.

Min, K.S. and Arkes, H.R., 2012. When is difficult planning good planning? The effects of scenario-based planning on optimistic prediction bias. *Journal of Applied Social Psychology*, 42(11), pp.2701-2729.

Ministry of Transport and Building (2015). *Evaluering af ny anlægsbudgettering (Evaluation of new project budgeting).*

Morgan, D.L., 2014. Pragmatism as a paradigm for social research. Qualitative Inquiry, 20(8), pp.1045-1053.

Mouter, N., 2014. Cost-benefit analysis in practice: *A study of the way Cost-Benefit Analysis is perceived by key individuals in the Dutch CBA practice for spatial-infrastructure projects. TRAIL Research School*.

MTO Qualification Committee (2006). *Consultant Performance and Selection System Consultant Appraisal Reviews Consultant Infraction Reports Procedures Guide (Revised)*. Ontario, Ontario Ministry of Transportation

MTO Qualification Committee (2017). *Consultant Performance and Selection System Consultant (CPSS), Procedures Guide*. Ontario, Ontario Ministry of Transportation

Muller, R. ed., 2016. *Governance and Governmentality for Projects: Enablers, Practices, and Consequences*. Taylor & Francis.

Newbold, P., 2013. *Statistics for business and economics*. Pearson.

Odeck, J. and Kjerkreit, A., 2019. The accuracy of benefit-cost analyses (BCAs) in transportation: An ex-post evaluation of road projects. *Transportation Research Part A: Policy and Practice*, 120, pp.277-294.

Odeck, J., 2004. Cost overruns in road construction—what are their sizes and determinants?. *Transport policy*, 11(1), pp.43-53.

Odeck, J., 2014. Do reforms reduce the magnitudes of cost overruns in road projects? Statistical evidence from Norway. *Transportation Research Part A: Policy and Practice*, 65, pp.68-79.

Odeyinka, H. and Perera, S., 2009, September. An evaluation of the budgetary reliability of bills of quantities in building procurement. *In RICS COBRA Research Conference* (pp. 435-446).

Odeyinka, H., Larkin, K., Cunningham, G., Weatherup, R. and McKane, M., 2010, September. Assessing risk impacts on the variability between tender sum and final account. In *RICS Construction and Building Research Conference*. Royal Institution of Chartered Surveyors (RICS).

Olatunji, O.A., 2008. A comparative analysis of tender sums and final costs of public construction and supply projects in Nigeria. *Journal of Financial Management of Property and Construction*, 13(1), pp.60-79.

Olsson, N.O., Nystrom, J. and Pyddoke, R., 2019. Governance regimes for large transport infrastructure investment projects: *Comparative analysis of Norway and Sweden. Case Studies on Transport Policy*.

Onwuegbuzie, A.J. and Johnson, R.B., 2006. The validity issue in mixed research. Research in the Schools, 13(1), pp.48-63.

Park, Y.I. and Papadopoulou, T.C., 2012. Causes of cost overruns in transport infrastructure projects in Asia: their significance and relationship with project size. *Built Environment Project and Asset Management*, 2(2), pp.195-216.

Patton, M.Q., 1988. Paradigms and pragmatism. Qualitative approaches to evaluation in education: The silent scientific revolution, pp.116-137.

Pearce, L.D., 2012. Mixed methods inquiry in sociology. American Behavioral Scientist, 56(6), pp.829-848.

Peters, M.L., Flink, I.K., Boersma, K. and Linton, S.J., 2010. Manipulating optimism: Can imagining a best possible self be used to increase positive future expectancies?. The Journal of Positive Psychology, 5(3), pp.204-211.

Peters, M.L., Vieler, J.S. and Lautenbacher, S., 2016. Dispositional and induced optimism lead to attentional preference for faces displaying positive emotions: An eye-tracker study. The Journal of Positive Psychology, 11(3), pp.258-269.

Pezzo, S.P., Pezzo, M.V. and Stone, E.R., 2006. The social implications of planning: How public predictions bias future plans. *Journal of Experimental Social Psychology*, *42*(2), pp.221-227.

Picciotto, R., 1994b. Marseille Discussion: Learning via the Hiding Hand. *Appendix A of Lloyd Rodwin and Donald A. Schön, eds., Rethinking the Development Experience: Essays Provoked by the Work of Albert O. Hirschman*, pp.301-304.

Picciotto, Robert. 1994a. "Visibility and Disappointment: The New Role of Development Evaluation." *In Rodwin and Schon 1994, 210–30 and 341–42*.

Pickrell, D.H., 1990. Urban Rail Transit Projects: Forecast Versus Actual Ridership and Costs [October 1989].

Plott, C.R., 1991. Will economics become an experimental science?. Southern Economic Journal, pp.901-919.

Pollack, J., 2007. The changing paradigms of project management. International journal of project management, 25(3), pp.266-274.

Potts, K. and Ankrah, N., 2008. *Construction cost management: learning from case studies*. Routledge.

Prater, J., Kirytopoulos, K. and Ma, T., 2017. Optimism bias within the project management context: a systematic quantitative literature review. International Journal of Managing Projects in Business, 10(2), pp.370-385.

Pryke, S. and Smyth, H., 2012. *The management of complex projects: A relationship approach*. John Wiley & Sons

Pychyl, T.A., Morin, R.W. and Salmon, B.R., 2000. Procrastination and the planning fallacy: An examination of the study habits of university students. *Journal of Social Behavior and Personality*, *15*(5), p.135.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Remenyi, D., Williams, B., Money, A. and Swartz, E., 1998. Doing research in business and management: an introduction to process and method. Sage.

Robson, C., & McCartan, K., 2016. Real world research. John Wiley & Sons.

Rodwin, L. and Schön, D.A., 1994. *Rethinking the Development Experience: Essays provoked by the work of Albert O. Hirschman*. Washington, DC: Brookings Institution.

Rotaris, L., Danielis, R., Marcucci, E. and Massiani, J., 2010. The urban road pricing scheme to curb pollution in Milan, Italy: Description, impacts and preliminary cost–benefit analysis assessment. *Transportation Research Part A: Policy and Practice*, 44(5), pp.359-375.

Roth, A.E., 1988. Laboratory experimentation in economics: A methodological overview. The Economic Journal, 98(393), pp.974-1031.

Salling, K.B. and Banister, D., 2009. Assessment of large transport infrastructure projects: the CBA-DK model. *Transportation Research Part A: Policy and Practice*, *43*(9), pp.800-813.

Salling, K.B., 2008. *Assessment of transport projects: risk analysis and decision support* (Doctoral dissertation, Ph. D. thesis, Department of Transport, Technical University of Denmark).

Samset, K.F. and Volden, G.H., 2013. Investing for Impact. Lessons with the Norwegian State Project Model and the first ivvestment projects that have been subjected to external quality assurance. *Concept rapport*.

Samuelson, P.A. and Nordhaus, W.D., (1985). Economics. 12th International Edition, New York: McGraw-Hill

Samuelson, P.A. and Nordhaus, W.D., (1992). Economics. 14th International Edition, New York: McGraw-Hill

Saunders A., M. Thornhill, A., and Lewis, P., 2009. Research methods for business students. Prentice Hall: London.

Scazzieri, R., 2003. Experiments, heuristics and social diversity: a comment on Reinhard Selten. In Observation and Experiment in the Natural and Social Sciences (pp. 85-98). Springer, Dordrecht.

Scheier, M. F., Carver, C. S., & Bridges, M. W., 1994. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the life orientation test. Journal of Personality and Social Psychology, 67, 1063–1078.

Scheier, M.F. and Carver, C.S., 1985. Optimism, coping, and health: assessment and implications of generalized outcome expectancies. Health psychology, 4(3), p.219.

Selten, R., 1998. Features of experimentally observed bounded rationality. European Economic Review, 42(3-5), pp.413-436.

Selten, R., 2003. Emergence and future of experimental economics. In Observation and experiment in the natural and social sciences (pp. 63-70). Springer, Dordrecht.

Sheldon, K.M. and Lyubomirsky, S., 2006. How to increase and sustain positive emotion: The effects of expressing gratitude and visualizing best possible selves. *The journal of positive psychology*, *1*(2), pp.73-82.

Shmueli, O., Pliskin, N. and Fink, L., 2016. Can the outside-view approach improve planning decisions in software development projects?. *Information Systems Journal*, *26*(4), pp.395-418.

Shook, J.R., 2000. Dewey's empirical theory of knowledge and reality. Vanderbilt University Press.

Siemiatycki, M. (2010). Managing Optimism Biases in the Delivery of Large-Infrastructure Projects: A Corporate Performance Benchmarking Approach. *European Journal Of Transport And Infrastructure Research, 10(1)*, pp. 30-41.

Siemiatycki, M., 2009. Academics and auditors: Comparing perspectives on transportation project cost overruns. *Journal of Planning Education and Research*, *29*(2), pp.142-156.

Simon, H.A., 1955. A behavioral model of rational choice. *The quarterly journal of economics*, pp.99-118.

Simon, H.A., 1979. Information Processing Models of Cognition. *Annual Review of Psychology,* 30, pp. 363–96.

Skitmore, R.M., Stradling, S.G. and Tuohy, A.P., 1989. Project management under uncertainty. *Construction Management and Economics*, *7*(2), pp.103-113.

Sloan, B., Tokede, O., Wamuziri, S. and Brown, A., 2014. Cost analysis error? Exploring issues relating to whole-life cost estimation in sustainable housing. *Journal of Financial Management of Property and Constructio*n, 19(1), pp.4-23.

Slovic, P., Finucane, M.L., Peters, E. and MacGregor, D.G., 2002. Risk as analysis and risk as feelings. *Decision Research*.

Sovacool, B.K., Gilbert, A. and Nugent, D., 2014. An international comparative assessment of construction cost overruns for electricity infrastructure. *Energy Research & Social Science*, 3, pp.152-160.

Streeten, Paul P. 1984. “Comment.” In *Pioneers in Development*, edited by Gerald M. Meier and Dudley Seers, 115–8. New York: World Bank and Oxford University Press.

Tam, C.M., Deng, Z.M., Zeng, S.X. and Ho, C.S., 2000. Performance assessment scoring system of public housing construction for quality improvement in Hong Kong. *International Journal of Quality & Reliability Management*, 17(4/5), pp.467-478.

Taleb, Nassim Nicholas, 2014. Standard Deviation. In: Brockman, J. (Ed.), This Idea Must Die: Scientific Theories That Are Blocking Progress. Harper Perennial, New York, pp. 535–537.

Teddlie, C. and Tashakkori, A., 2010. Overview of contemporary issues in mixed methods research. Handbook of mixed methods in social and behavioral research, 2, pp.1-41.

Terrill, M., Coates, B. and Danks, L., 2016, November. Cost overruns in Australian

Transport Infrastructure Projects. *In Proceedings of the Australasian Transport Research Forum* (pp. 16-18).

Tetlock, P., 2005. *Expert political judgment: How good is it? How can we know?*. Princeton University Press.

Thomas, A.B., 2004. Research skills for management studies. Psychology Press.

Thurairajah, N., Xiao, H., Boyd, D. and Reed, A., 2018. Challenges of Early Estimation of Infrastructure Projects within the UK: an Information Perspective. In: 34th Annual ARCOM Conference: A Productive Relationship: Balancing Fragmentation and Integration, 3-5 September 2018, Belfast

Trafikverket. (2014) *Planläggning av vägar och järnvägar*. Version 1.0. Borlänge, Sweden: Trafikverket.

Tversky, A. and Kahneman, D., 1974. Judgment under uncertainty: Heuristics and biases. *science*, *185*(4157), pp.1124-1131.

Tversky, A. and Kahneman, D., 1983. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological review*, *90*(4), p.293.

Tversky, A. and Kahneman, D., 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, *5*(4), pp.297-323.

Tversky, A. and Koehler, D.J., 1994. Support theory: A nonextensional representation of subjective probability. *Psychological review*, *101*(4), p.547.

Varian, H.R., 2010. Microeconomics. A Modern Approach. University of California at Berkeley.

Volden, G.H. and Andersen, B., 2018. The hierarchy of public project governance frameworks: An empirical study of principles and practices in Norwegian ministries and agencies. *International Journal of Managing Projects in Business*, 11(1), pp.174-197.

Volden, G.H. and Samset, K., 2017. Governance of major public investment projects: principles and practices in six countries. *Project Management Journal*, 48(3), pp.90-108.

Wachs, M., 1986. Technique vs. advocacy in forecasting: A study of rail rapid transit. *Urban Resources*, 4(1), pp.23-30.

Wachs, M., 1989. When planners lie with numbers. American Planning Association. *Journal of the American Planning Association*, 55(4), p.476.

Wachs, M., 1990. Ethics and advocacy in forecasting for public policy. *Business and Professional Ethics Journal*, 9(1/2), pp.141-157.

Weber R.A. and Camerer C.F., 2003. Cultural conflict and merger failure: An experimental approach. Management Sci. 49(4):400–415.

Webster, M. and Sell, J. eds., 2014. Laboratory experiments in the social sciences. Elsevier.

Weyer, B., 2011. *Perspectives on optimism within the context of project management: A call for multilevel research* (No. 59). Working Papers of the

Institute of Management Berlin at the Berlin School of Economics and Law (HWR Berlin).

Wickham, H., 2016. *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag New York.

Wicks, A.C. and Freeman, R.E., 1998. Organization studies and the new pragmatism: Positivism, anti-positivism, and the search for ethics. Organization science, 9(2), pp.123-140.

Williams, M.J., 2017. The political economy of unfinished development projects: Corruption, clientelism, or collective choice?. *American Political Science Review*, 111(4), pp.705-723.

Williams, T., Vo, H., Samset, K. and Edkins, A., 2019. The front-end of projects: a systematic literature review and structuring. *Production Planning & Control*, pp.1-31.

Willoughby, C., 2003. The first experiments in operations evaluation: Roots, hopes, and gaps. *World Bank Operations Evaluation Department: The First*, 30, pp.3-15.

Windmann, S., Kirsch, P., Mier, D., Stark, R., Walter, B., Güntürkün, O. and Vaitl, D., 2006. On framing effects in decision making: linking lateral versus medial orbitofrontal cortex activation to choice outcome processing. *Journal of Cognitive Neuroscience*, *18*(7), pp.1198-1211.

Wójtowicz, A. and Winkowski, J., 2018. Heuristics: Daniel Kahneman vs Gerd Gigerenzer. In *Rationality and Decision Making* (pp. 253-277). Brill.

Woolcock, M., 2019. Why Does Hirschmanian *Development Remain Mired on the Margins? Because Implementation (and Reform) Really is a Long Voyage of Discovery'* (No. 347). Center for International Development at Harvard University.