universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

"Rational enough to know better, confident enough to ignore anyway"

'The role of overconfidence in group target estimations'

verfasst von / submitted by

Karim Julian Gharbi, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2018 / Vienna 2018

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 914
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Internationale Betriebswirtschaft
Betreut von / Supervisor:	Ass.-Prof. Steffen Keck, PhD

Acknowledgements

In deep gratitude for the guidance, support and patience from my former colleagues of the Strategic Management Department, I'd like to say thank you.

To my beloved ones - this is (finally) for you!

"What a piece of work is man, How noble in reason, how infinite in faculty, In form and moving how express and admirable, In action how like an Angel, In apprehension how like a god, The beauty of the world, The paragon of animals."

William Shakespeare - Hamlet, Act II, scene 2 (1986)

Abstract

Aligned with the principles of behavioural economics, the aim of this thesis is to detect irrationalities in human behaviour that lead to 'non-optimal' decisions. The praised picture of mankind, as portrayed by the excerpt of Hamlet on the previous page, thereby gets gradually disassemble by examining cognitive restrictions in our daily thought processes. Making plausible and realistic predictions about the future events is not only one of the more prominent fallacies, but also represents the cornerstone of this study. The formulation of overly optimistic scenarios, is supposed to be even accentuated through group discussion. The centrepiece of this paper represents an experimental study addressing the issue of collective resource misjudgements under varying task externalities. Participants were asked to predict – individually and collaboratively – the time needed and their level of performance for two subsequent group exercises. Results only partly supported findings from established research on the (group) planning fallacy. Estimation shifts or an accentuated effect of overoptimism was also only evident in a slightly insignificant manner. The third prediction, that task externalities, precisely interdependencies, having a moderating effect on the relation between overconfidence and time misjudgements was not found in this experimental setting. Final discussions went beyond theoretical contributions to the decision making literature and also intended to provide implications within an organizational context.

Table of content

1. Introduction

New York City, March 3rd, 2016. Santiago Calatrava, a globally renowned architect from Spain, comments the inauguration of his latest project: "This is a great moment. This is a gift for all New Yorkers". The train station at ground zero amounted to almost $4 billion in public money and a construction time of eight years. Initially, the transit hub was forecasted to be finished within five years, at a cost of $2.2 billion (NYT, 2016).

Vienna, November 21st, 2017. "Yes, there have been wrong decisions in the past", admits the health councilwoman of the Austrian capital relating to the construction of the Vienna North Hospital. Already in the early stage of construction "We knew, that the costs will exceed the projection, just as the timeline". According to the building supervision, not less than 8.000 defects have been listed. Knowledge shortages and the lack of a general planner were stated as the main reasons that made "a project of this complexity not feasible like this" (translated quotations, derstandard, 2017). Numerically speaking, the total cost will aggregate to almost 1.6 billion Euros, out of an initial estimation of 825 million Euros (derstandard, 2017).

In a study by Ernst and Young (2014), more than 350 megaprojects (capital investments above $1 billion) in the oil and gas industry were examined, of which 64% were facing cost overruns while 73% of the projects reported schedule delays. The immediate effect of project delivery on the company's share price is representative for the significance of this phenomena. Axel Preiss, EY's global oil and gas advisory leader, further comments, "longer-term industry outlooks suggest that project delivery success is actually decreasing, especially in certain segments of the industry, such as deepwater, where complexity is considerably higher" (p.2). According to the report, the failure to meet deadlines and budgets is a combination of several factors at different stages of the project timeline. Inadequate planning, consisting of overly aggressive forecasts as well as optimistic biases, is considered a key driver.

The impact of misplanned projects for individuals, taxpayers, shareholders, organizations, governments or any other affected entities is an undisputed fact, and thus highlights the magnitude of the planning fallacy's phenomena to the overall society. Given the relevance of the topic, research was not long in coming, considering the fact that the entire field of behavioral economics is rather a more recent establishment. Pioneering this concept, Kahneman and Tversky (1979) argued in their masterpiece 'Intuitive prediction' that "errors of judgment are often systematic rather than random" (page i). Subsequent studies were largely concerned to explain the driver of the time component in the misjudgment process (Buehler, Roger; Dale Griffin; Michael Ross 1994; 1995; 2002). The overrun of costs as well as shortfalls of other benefits complement the trinity of misery (Lovallo & Kahneman, 2003). Further manifestations of the dilemma, e.g. the malevolent hiding hand principle by Albert O. Hirschman (1967), give additional insight about the wide applicability of this topic.

Taking a look back at the above mentioned real world examples from various industries and regions, the two variables project complexity and overconfidence seem to be predominant. Referring to the latter one and trying to explain, where in the human nature these optimistic tendencies are rooted in and how they unveil in our daily lives, the forthcoming part discusses several theories from psychological background. Towards the end of that literature review, findings will be reexamined for multi agent interactions. The early and largely covered polarization effect in groups (Myers and Lamm, 1975) is going to shed light on more extreme decisions and direct the analysis to the main research gap. Collaborative failure in time and efficiency estimations will be subject to the experimental part of this paper, which will hopefully contribute to motivate further research and someday find applicability in practice.

2. Literature review

The overall topic of overconfidence finds vast research in its breath and depth, which made it beneficial to subcategorization and niche studies. As Moore and Healy (2008) discussed, this topic was related to entrepreneurial entry, explain the high rates of corporate merger and acquisition despite frequent failure, high stock trade volumes, labor strikes and even causes of war. Even going a step further, Plous (1993) named the nuclear catastrophe at Chernobyl and to the explosion of the space shuttle Challenger outcomes of overconfidence and added, "no problem in judgment and decision making is more prevalent and more potentially catastrophic than overconfidence" (p. 217). Given the wide applicability and interest, scholars attempted to structure the field by reconciling differing definitions. A framework of three distinct variables, established itself as research standard (Moore and Healy, 2008).

The most prominent component "overestimation", represents roughly 64% of empirical papers and is described as the following: "overestimation of one's actual ability, performance, level of control, or chance of success" (p.3). Within this group, the planning fallacy bias is the most relevant manifestation and thus is going to be backed up with detailed insights shortly. Second, the individuals' believe to be better than others is given the title "overplacement" and amounts to not more than 5% of the literature. Svenson's (1981) findings on car drivers' self-perception is probably the much-noticed example of the better than average effect. Last, the habit of giving estimations in a way too precise manner, is generally known as "overprecision". The reaming thirty-one percent of literature dealing with this phenomenon, found too narrow prediction intervals as a key driver. As we are going to detect, individual's answers are more than half of the time wrong, even within very high intervals of confidence.

2.1. Overplacement

Given the literature's structure on overplacement, which largely consists of the better than average effect, the focus in this subsection will also reside in that interest. Contrary to traditional social comparison theories, where points of reference are commonly other people, the better than average effect defines misjudgments of a person's characteristics with regard to a given norm or standard (Alicke & Govorun, 2005). This self enhancement phenomenon is frequently an interplay between several factors like inferior comparison targets, extensive focalism or egocentrism in terms of evaluating one's own characteristics and behaviors. However, there is contradicting evidence within research that this effect is nothing but an automatically applied heuristic in social judgment processes. Following this argument, any positively distorted self-perception does not necessarily have to be underlined by profound trait comparison, but rather might be a consequence of a general positivity bias (Alicke & Govorun, 2005).

Nonetheless, it is worth mentioning that the better than average is a bias that promotes and sustains harmony within personal relationships and fosters goal striving in professional realms. According to a study from Zenger (1992), almost 40% of one firm's professional engineers placed themselves among the top 5% of performers at the firm. Another statistical paradox, and probably the most cited study from this field, was introduced by Ola Svenson (1981). When asking American / Swedish drivers to rate their abilities on the streets, 93% / 69% respectively indicated that they were more skillful and less risky than the median driver in their own country. Neglecting the above mentioned statistical doubts concerning the result, one could observe that believing in one's superiority may lead to greater risk taking. This might be counterproductive when it comes to traffic safety, but can actually be beneficial in business or administration environments to reach or stay on top positions.

A contradictory approach, however, was formulated by Kruger (1999), who displayed that the overplacement effect is limited to "easy" tasks that participants feel competent in or success is

bound to occur. Contrary, when confronted with more complex or difficult tasks, people estimate their chance to succeed or perform well to be below average.

To be understood as a complement to the above discussed effects of overconfidence, Kruger and Dunning (1999) proposed that lower skilled individuals are just unaware of or unable to recognize their own incompetence, which thus leads to an "inflated self-assessment". The so called 'Dunning Kruger Effect' explains the cognitive bias of illusory superiority as the following: "the miscalibration of the incompetent stems from an error about the self, whereas the miscalibration of the highly competent stems from an error about others." (p. 1127)

2.2. Overprecision

The phenomenon of an individual's excessive certainty regarding the accuracy of its beliefs, is discussed under the general term of overprecision. Within literature, the most frequent application or testing of this fallacy is done with confidence intervals. Asked to report their confidence, participants indicate the probability that their answer to a given problem is right. The majority of these item-confidence paradigms give proof that the level of confidence systematically exceeds the actual accuracy. Further findings from Mannes and Moore (2013), describe a negative correlation between confidence in judgements and adjustments of estimations after receiving feedback about accuracy and the cost of being so. They further claim that typical lab experiments with numerical estimates, like the length of the Nile river, don't really picture judgements of our everyday life. Time estimations, however, are relevant for each and every one of us and also every day. The cost of cutting things too close, arriving late, missing planes or bouncing checks are just a few of the daily pitfalls though. So opposite to what people should do, in cases where they are not as right as they think they are, they actually stick under resistance to any advice to their initial judgements.

However, the results from the studies just mentioned have to be treated with caution. As a matter of fact, the effect of overprecision on its own is rarely discussed within popular research paradigms. Instead it is mostly confounded with the concept of overestimation. Thus, making it impossible to distinguish the relative influence of each (Moore and Healy, 2008). Their example of the case where being excessively sure having a certain item right, illustrates the dilemma of distinguishing an overestimation of one's own performance and an overconfidence in the precision of knowledge. There is though a solution to this problem, namely instructing a setting where perceptions of performance are measured over a set of items. These frequentistic judgments (e.g. number of correct items), are indeed less prone to overconfidence than probabilistic (single item-level) judgements. As Moore and Healy summed up (pp. 7,8), "(…) overconfidence in item-confidence judgments is attributable to overprecision. (…), whereas set-level (frequentistic) judgments will not necessarily be biased upward by overprecision."

2.3. Overestimation

Getting closer to the centerpiece of this thesis, this subsection deals with the largest representative of the overconfidence literature, accounting to almost two thirds of all publications. The excessive certainty into the own ability, performance, level of control, or chance of success is summarized by the three effects: illusion of control, contrary evidence and planning fallacy. This paper will discuss the latter one in detail, however, since all of the three seem to be attributed to irrationality, the examination will depart from failures of human cognitive abilities. The behavioural economist and professor at Duke University, Dan Ariely, defines is his field of expertise, three relevant manifestations of irrationality that correlate with the overall idea of overestimation. Those include estimations of our own ideas, of things we create and also about our own behaviour (Ariely, 2010)

2.3.1. Not invented here bias – ‘the toothbrush theory’

The question how people relate to their own ideas versus to the ones of others’ is perfectly illustrated by an essay of Mark Twain (1923, p. 175) called ‘some national stupidities’:

„The slowness of one section of the world about adopting the valuable ideas of another section of it is a curious thing and unaccountable. This form of stupidity is confined to no community, to no nation; it is universal. The fact is the human race is not only slow about borrowing valuable ideas—it sometimes persists in not borrowing them at all.“

In his further elaboration, Twain gave the example of the German stove that was superior in every department, compared to the American one, but still wasn’t accepted by the American industry since the design was foreign.

Almost a hundred years later, breaking this concept down to company level, Ariely (2010), examines the failure of Sony. After a series of highly successful inventions, like the transistor radio, the Walkman or the PlayStation, inventions that weren’t developed in-house were completely dismissed by the engineers. A fact that was later also admitted by Sony’s CEO at that time, Sir Howard Stringer. The missed opportunities on products such as MP3 players, flat-screens or smartphones are well known witnesses of failure due to the not-invented-here bias.

Additional proof of that phenomenon on individual level can be found in a quite prominent historical example. The Serbian inventor Nikolai Tesla came up with the idea, while working for Thomas Edison, an outspoken direct current proponent, that the illumination of lightbulbs over greater distance and power supply of large industrial machines is only feasible with alternating current. Even though AC was the solution to meet the large and growing demand for electricity of the modern world, Edison was not only dismissing the technology, but also ran a whole campaign against it, propagating the danger and harmfulness.

The upside of fixation, to highly commit oneself and invest in a certain concept, is seldom overweighing the cost of dismissing others' ideas that are just better than one's own.

The toothbrush theory, an ideogram within the scientific world, perfectly sums up the concept of the NIH bias. "Everyone wants a toothbrush, everyone needs one, everyone has one, but no one wants to use anyone else's" (Ariely, 2010, p.185).

2.3.2. Even the monkey, in his mother's eyes, is an antelope

Similar to the principle of commitment and consistency in Cialdini's theory of influence (2009), the answer to the question 'why do we overvalue what we make' is most commonly the following. Much of what really matters to us is our creation.

Literature has come up with a concept called the egg theory that was mainly characterized by the revolutionary work of the Austrian-American social scientist Ernest Dietrich. In the 1950s, instant baking and cooking mixes gained large popularity in the American society. However, sales numbers of cake mixes could never catch up, which lead to various incorporations of business psychologists and behavioural marketers in sales teams. Within his research Dietrich found that cakes often carry great emotional significance or symbolize special occasion: "The most fertile moment occurs when the woman pulls the finished cake, bread loaf, or other baked product out of the oven. In a sense it is like giving birth" (Dichter, 1964, p. 21). Therefore, just adding water and bake it, marginalises the role of the women (of the cook) and holds the risk of disappointing the guests by not serving them something special. Confirming Dichter's theory, to leave out the eggs and make people at home add them along with fresh milk and oil, sales took off. Successfully understanding the crossover between pride of ownership and time efficiency, the TV-chef Sandra Lee has formulated the „70/30 Semi-Homemade® Philosophy" that combines off-the-shelf products with just the right amount of personalization.

Another manifestation of personal involvement leading to overvaluation, is denoted by Norton and colleagues (2011) as "the IKEA effect". The meaning of labour, seems to be deeply rooted in our genetics. Even animals prefer sources of food that require effort to obtain. When it comes to furniture, the Swedish pieces might not be the most expensive ones or the hardest to get for an average income household. Moreover, contrary to their allegedly simplicity, the assembly usually requires much more time and effort. In the end, the finished object (completion is crucial, according to Norton et al.) causes a feeling of pride and an increased valuation, which the creators falsely also expect others to share. In their experimental design, the Norton et al. (2011) made Harvard students to bid in an auction for their self-made origami creatures. On average participants valued their objects five times higher than non-creating bidders.

2.3.3. Adam & Eve problems

In the future, we are all wonderful human beings. We save more, we eat more healthy, we procrastinate work less etc. Meanwhile in the present, we fall for temptations over and over again. This lack of self-control, just as Eve eating the fruit from the forbidden tree, is studied under the term of hyperbolic time discounting. According to Ariely and Wertenbroch (2002) "immediately available rewards have a disproportionate effect on preferences relative to more delayed rewards, causing a time-inconsistent taste for immediate gratification" (p.219).

Apparently, inconsistencies of preferences of a single decision maker within two different points of time, legitimate the act of questioning whether humans are even able to make correct estimations about themselves. As Thaler and Shefrin (1981) put it, "the individual at a point in time is assumed to be both a farsighted planner and a myopic doer" (p.39).

Whereas theoretical validation of that effect is anything but scarce, controlled empirical or experimental evidence is more like a rarity. Nevertheless, in a study by Read et al. (1999),

participants were asked to pick a series of three different movies to watch on three different days, ranging from highbrow to lowbrow. Results confirm that subjects that had to make a simultaneous choice (select all three movies on the first day), were much more likely to take highbrow movies for the second and third day, than subjects with a sequential choice (each day). In other words, long term benefits such as cultural enrichment or education got traded off for immediate pleasure, mainly due to an illusional self-perception.

2.4. The Planning Fallacy*

Similar to the just mentioned drawbacks of lacking self-control, the effect of the planning fallacy (hereinafter referred to as PF) makes deviations from an initial intent countable. The interplay of overestimating one's own rate of work with an irrational level of ignorance, attributed this bias a significant relevance and has drawn the attention of several renowned scholars. In their pioneering contribution, Kahneman and Tversky (1979, p. i) stated: "Errors of judgment are often systematic rather than random, manifesting bias rather than confusion. Thus, man suffers from mental astigmatism as well as from myopia, and any corrective prescription should fit this diagnosis."

In a later interview with Gallup (2012), Kahneman described the PF as, "… that you make a plan, which is usually a best-case scenario. Then you assume that the outcome will follow your plan, even when you should know better." 'Knowing better' from past experience or being 'rational enough', does not prevent individuals from systematically underestimating the time and resources required to complete projects. According to Buehler, Griffin and Ross (1994), people holding two contradictory believes, namely thinking that their current estimations are

* Even though the planning fallacy is considered a subsection of the overestimation effect, it is yet treated in a separate unit, due to the substantial significance of the insights.

accurate or realistic, even though knowing that their previous predictions have been overly optimistic. Kahneman and Tversky (1979) explained this behaviour with the differences of singular and distributional information. They argue that people who rely on singular information, which relates to aspects of the specific target task, adopt to an internal perspective and work out an independent strategy. On the other hand, focusing on distributional information embraces an external perspective and thus people compare the actual task with similar past projects. Feeding the existence of the PF, individuals are prone to derive forecasts based on the internal perspective and neglect to learn from their experience. Buehler, Griffin and Ross (1994) suggested that making use of both systems would help to overcome the optimism bias.

A slightly different approach was formulated by Roy et al. (2005), who argued that people do not remember correctly the amount of time it took them to finish tasks in the past. Hence, positively biased memory leads to a biased prediction about future event duration.

However, Buehler and colleagues (1994) contradict the theory by Roy, by discussing the role of deadlines. In their study, they examined whether students' time estimation to hand in their final thesis was overly optimistic. The participants had to predict when (number of days) they realistically were going to submit, as well as when they'd submit in case "everything went as poorly as it possibly could." Only 30% of the students in the "realistic case" and fewer than half (49%) in the "worst case", managed to finish by the predicted time, indicating a strong general tendency of over optimism. On average, the completion time of 55 days was missed in the realistic estimate by 22 days and even in the pessimistic case by 7 days, representing an error of 40%, 13% respectively. Against the argument of Roy et al. (2005), Buehler and colleagues found that the individuals' predicted completion times were highly correlated with the actual completion time. In other words, students that predicted to take longer, actually took longer. Therefore, highlighting the informative content of predictions, even if they are optimistically biased.

"Each game is unique, and this one is no different than any other" (Phil De Glanville)

Despite the evidential value of using relevant distributional information, like past time experience, Buehler et al. (2002) found that people rarely incorporate them to make predictions about future events. Next to the very nature of predictions, drawing attention on upcoming scenarios rather than bygones, people commonly attribute little relevance of their past experience to their current task. In particular when its apparent implications (e.g. longer completion time) would cause a feeling of unease and challenge their optimistic plans. The often mentioned intent "learning from mistakes of others" seems even more unrealistic in this context. The lack of background knowledge makes people commonly doubt the generalizability of others' experiences and even tend to attribute other peoples' mistakes, unlike one's own, to their personal disposition.

In the presence of having the relevant and sufficient experience, a reason why people still take on risky projects, is the overestimation of related benefits while neglecting the costs (Kahneman, 2011). Risky, in that sense, refers to high levels of complexity due to new technology or hidden costs and obstacles like prior required political approval. Examples are mostly found in large construction projects. However, ventures like the Sydney Opera House, which was inaugurated 10 years later and at a cost almost 15 times higher as expected, are most likely (or hopefully) never planned by a single person. Thus, on a further attempt, the discussion will address the question how the interaction of various individuals with different attitudes and believes affect the decision of groups. Hence, in quest of the sparsely treated phenomenon of the group planning fallacy, examining whether overoptimistic biases of subjects are reinforced or evened out through discussion.

2.4.1. The Group Polarization Effect

For almost half a century now, the tendency of groups to make more extreme decisions than the initial inclinations of its members, has been a topic of large interest. Already in 1969, Moscovici and Zavalloni defined the term group polarization as, "a group-produced enhancement of a prevailing individual tendency" (found in Lamm, H. and Myers, D.,1978, p. 146). The concept is based on an experimental finding of the MIT student James Stoner from 1961, called the "risky shift". Subsequently often found under the term of the 'choice shift', thus not only portraying a group's decisions towards greater risk, but also towards greater caution, in case of individuals' initially cautious tendencies (Aronson, 2010).

Despite the number of similarities between group polarization and choice shifts, there is one distinct difference, according to Myers and Lamm (1976). Group polarization, on the one hand, refers to attitude intensification on the individual level due to group influence and discussion (attitude polarization). Choice shifts, on the other hand, are the outcome of that attitude change, namely, from a group member's pre-discussion attitude to the outcome of the group decision.

However, not all group settings are affected by polarization in the same way. As Myers and Lamm (1975) suggested, well-established groups or topics that are well-known to them are in general less prone, whereas with newly formed groups that are confronted with novel tasks, group polarization will have a more profound influence on the decision making.

Empirical evidence for that argument exists plenty. Bray and Noble (1978) confirmed the effect on jury trials. Their findings include that high authoritarian subjects reached guilty verdicts more frequently and imposed more severe punishment than low authoritarians. Latter ones even shifted towards increased lenience through deliberation. Kalven and Zeisel (1966), summarized "the deliberation process might well be likened to what the developer does for an exposed film: it brings out the picture, but the outcome is predetermined" (p. 489).

On a more recent attempt, Yardi and Boyd (2010) found examples of groups that are not even physically together. Online media outlets like Facebook and Twitter demonstrate, to our all knowledge, huge platforms to share opinions and involve in discussions. Following the above-mentioned argument by Myers and Lamm (1975) concerning the familiarity of group members, the internet, given the anonymity of users, is an archetypal breeding ground for group polarization. Not least also because of the nature of the topics and issues, being trending or current headlines. Nevertheless, attitudes strengthen only among like-minded individuals, Yardi and Boyd (p. 2) conclude, whereas "replies between different-minded individuals reinforce in-group and out-group affiliation".

2.4.2. The Group Planning Fallacy

Combining the two just discussed concepts, this section examines whether the polarization effect also affects optimistic time predictions of groups. The relevance of that approach is highlighted by Buehler, Messervey and Griffin (2005) who argue that most projects are planned and carried out by groups such as committees, teams or work groups. They further outlined that there is only a limited supply of research in collaborative predictions concerning the social interactive dimension, whereas the majority focused on the benefits and drawbacks of aggregating independent forecasts (e.g. Clemens & Winkler, 1993). The relevance of integrating various opinions is not only dominant in a professional environment, but also when it comes to personal future concepts. Especially in these cases, the superiority of social interactions over aggregating induvial advices and opinions is evident.

To adduce the earlier discussed concept from Kahneman and Tversky (1979) about the inside and outside view of team members, there could probably arise a doubt whether optimistic biases that reside in individuals also manifest in group tasks. Since they argue that observers, contrary to active agents, are less prone to overconfidence given their objective viewpoint, people in an

inter-dependent group task, while being in the state of waiting for contributions of others, might as well adapt to a more observant stance and thus mitigate a potential planning-based optimistic bias. However, as Buehler, Messervey and Griffin (2005) argue, membership in a collective group, irrespective of the strength of bonds, generates significantly higher levels of involvement compared to neutral observers. Furthermore, next to being partially in an observer role for other colleagues' contributions, the active participation in the own task makes the agent become a stakeholder of the big picture. In sum, the tendency of an actor to think as an observer gets overridden by the active contribution role and therefore constitutes the fundament of optimistically biased predictions for group tasks.

Moreover, since successful task completion is commonly a shared goal among group members, their exaggerated optimism (or supressed pessimism) might be understood as commitment and loyalty to the project (Kahneman & Lovallo, 1993). Next to these concerns about group approval, the lack of individual accountability for wrong group estimates, may further promote unrealistic views.

Another factor that is considered to add to the group bias is selective focus. Due to discussion, the members' attention is drawn solely on information that accentuates their own initial inclination. This, in most of the cases, are thoughts about successful task completion. A disproportionatel regard on task relevant information, like task or team member qualities, leads to the assumption of prompter completion and consequently to neglect information suggesting a task may take longer than desired (Buehler, Messervey and Griffin, 2005). This idea of selective attention is also thematised in the concept of the team scaling fallacy (TSF) by Staats et al. (2012). The effect results from a propensity to excessively focus on process gains (e.g. associated with increasing team size), relative to the process losses. Per definition, the TSF is "a tendency to increasingly under-estimate task completion time as team size grows" (p. 132).

2.5. Research Gap

Even though there exist recognised research attempts describing the group planning fallacy that managed to establish itself, empirical contributions similar to Buehler and colleagues (2005) ones, are to my best knowledge of scarce volume. Their findings were consistent with the underlying assumption that people underestimate completion time for group projects. The effect was significantly evident for individual as well as for collaborative predictions, even in the presence of distributional information (previous group time). Further they confirmed the effect of group polarization/accentuation, indicated by more optimistic (and more biased) collaborative estimates than by the individual group members. On top of that, they could prove the mediating effect of selective focus on the overall group discussion bias. Admittedly, these findings represent a milestone in the research field of group biased time estimations, however the plain matter of the under-proportional amount of empirical research, compared to similar topics like the group polarization (Lamm and Myers ,1978; Myers and Lamm, 1975, 1976; Bray and Noble, 1978) or accentuation (Kerr & Tindale, 2004; Hinsz et al., 1997; Tindale, 1993; Kerr et al., 1996) effect, calls for additional examination.

Second the role of team member interdependency in group tasks has been fairly neglected in previous studies yet. Indeed Buehler, Messervey and Griffin (2005) highlighted the discrepancy of contributor and observer roles, but not only did they left out the variable of task interdependency in their experimental design, but also, they just mentioned the sequential part of interdependence. Thus, the question whether the nature of the group task, being more individual (parallel setting) or more interdependent, has a different effect on time predictions, is still to be answered. Furthermore, approaching task interdependency not from a sequential point of view, where "people often wait for contributions of others", but rather from a simultaneous aspect, where the team members actually work on the same object at the same time, displays an unobserved model that necessitates empirical analysis.

Finally, the present-day research has not taken group decision schemes on board, which may impact the determination of the project time-frame. Models that analyse the within group process to reach a consensus exist plenty in social science (e.g., Social Judgment Scheme, Davis, 1996; Social Decision Scheme for Quantities, Hinsz, 1999; Revision and Weighting model, Sniezek & Henry, 1990). However, in the context of biased time predictions, an unanswered question is still, whether the group estimate would follow e.g. the most optimistic group member (minority influence / polarization) or rather fall back on averaging opinions. Furthermore, how these different schemes relate to the accuracy or bias of the group prediction.

3. Research design

Building on the scarce empirical examination of overconfident time predictions in group tasks, in the face of the above several mentioned significance, the present research in this paper attempts to connect the literate stream of the planning fallacy with strands concerning attitude polarization in groups. A large focus is thereby drawn on the discrimination of greater or lesser task interdependency as well as on consensus approximation patterns.

Within the established literature on the overall effect of the planning fallacy, there is little doubt about the existence of this bias. Likewise, the vast number or empirical evidence does not necessarily require additional dissection of the 'judgemental error'. However, in the spirit of good scientific conduct, a research design that addresses derived implications about the group planning fallacy and still attempts to be holistic, demands an empirical testing of the grassroots.

Working hypothesis 1: Due to the distorted self-perception of overconfidence, individuals tend to systematically underestimate the time and resources required to complete projects.

Based on that assumption and paired with insights from group interaction literature, Buehler, Messervey and Griffin (2005), discussed the degree of bias in predications that are generated collectively. The solely inexistence of follow up data confirming their significant results, inevitably commands to replicate parts of their study. Therefore, the subsequent working hypothesis is proposed:

Working hypothesis 2: The interaction and discussion between group members accentuates the effect of the planning fallacy, resulting in even more overoptimistic time predictions.

As Kozlowski and Bell (2001) suggested, research that fails to consider interdependence has limited value for building knowledge about team effectiveness. Per definition, task interdependence refers to the degree of task-driven interaction among members (Shea & Guzzo, 1987) and may range in its manifestation from pooled to sequential to reciprocal (Thompson, 1967). Contextual variables such as task characteristics, type of goal, or consequential rewards/punishment, are acknowledged among scholars as key drivers for interconnections. Considering a task that provokes independency just by its very nature at levels even higher than reciprocal, e.g. conjunctive simultaneous ones that require all members to perform effectively at the same time (Gully et al. 2002), most likely will diminish the accuracy of completion time predictions due to the underlying neglect of additional coordination needs. This leads to the formulation of the following working hypothesis:

Working hypothesis 3: Task interdependence will moderate the relationship between overconfidence and accuracy of time predictions such that higher levels of interdependence will be related more strongly to overoptimistic time predictions.

4. Experimental design

4.1. Overview

Twenty-three two-person teams were asked to participate in a decision making / team efficiency study. Subjects were confronted with two different crafting tasks, which they had to perform in teams of two, one after another. Prior to the actual crafting, each person had to individually estimate the time their team will take to finish the task, as well as the level of performance. Groups were then randomly assigned by immediate vicinity matching. Each group was subsequently asked to provide a joint judgement of expected completion time and performance. The purpose of the study was not explicitly revealed, rather participants were told that the experiment was conducted to observe collective decision-making processes.

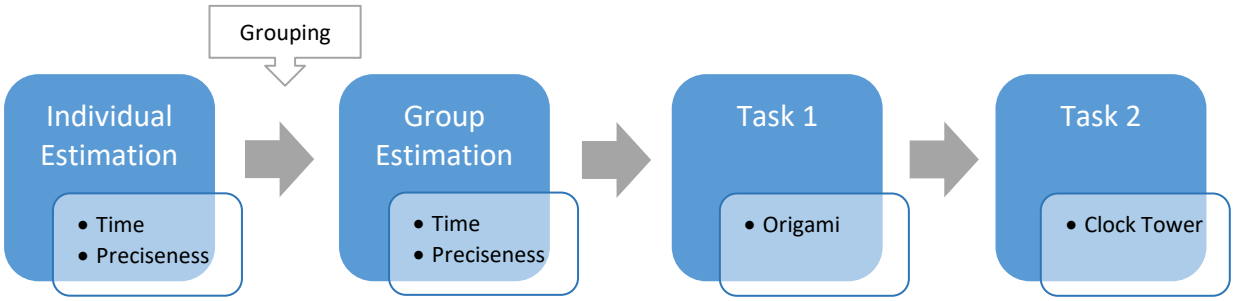


Figure 1: Experimental Flow Chart

Upon the arrival all participants were given a written introduction sheet (Appendix: figure 12), laying out the procedure and task essentials. These also contained assembly instructions for both, the origami creatures (Appendix: figure 13) as well as the clock tower (Appendix: figure 14). After having familiarized themselves with the challenges of each task, subjects were then asked to fill out the individual estimation sheet (Appendix: figure15), predicting completion time (mins) and preciseness (%) for both of the tasks. Thereafter, participants were pooled with their neighbours to form teams and discuss their assumptions and strategies to reach the target.

Consequentially they had to put their joint prediction, on the same variables as before, down on the group estimation sheet (Appendix: figure 16).

Contrary to the probably assumed smaller significance of the two different estimations (individual & group) each candidate would provide, the information that was derived from these predictions did almost represent the centrepiece for further discussion. Especially the variable of time was crucial for the determination of the planning fallacy. In order to divert the attention of the participants away from this importance, next to the experiment's stated purpose of team work efficiency, the variable of task accuracy / performance pretty much served as an interference factor. Additionally, it was also taken as a measure of confidence, when counterbalanced the indicated performance estimation, with the prior experience and general talent in crafting stated on the questionnaire (Appendix: figure 17) that was to be filled in after the completion of the second task.

The intent of the experimental setting was to highlight divergencies in completion time estimations and the actual time required to finish. The trait of (over)confidence, which naturally resides within our human nature, was in this case the trigger or talking in modelling terms, the dependent variable and thus out of any possible interference for the experimenter. On the other hand, letting each candidate make their own private prediction first and then discuss with their partner to come up with a joint estimation, intended to examine the existence of a group accentuation effect hence a group planning fallacy. Further manipulations as the layout of the different tasks served to test team member interdependencies Apart from subjective talents and experience, the tasks were designed to match in their level of difficulty as well as in their time to complete them. However, the clock tower construction was meant to be objectively more interdependent and require more efforts of coordination, in order to observe whether this kind of challenge would cause even stronger misjudgements of time.

4.2. Participants

Forty-six bachelor students (25 female, 21 male) at the University of Vienna took part in a team exercise during a class in strategy. The exercise was completed for a class in which students were not accustomed to participating in social experiments. Moreover, no cover story was given to the students about the exercise, and no material relevant to the group planning fallacy was covered in class prior to the exercise. Participants were randomly assigned to two-person teams (23 teams). All of the teams were then asked to complete the same group crafting and construction tasks. The subjects received no financial incentives or any other kind of credential reward for participation. It took about 50 minutes to finish all stages of the experiment.

4.3. Treatment

Participants of the experiment were asked to perform in a two (respectively four) layered exercise. Having the preparational part of planning, discussing and estimating, in the beginning, the more active part or the actual crafting and construction was to be followed up on a second step. Each of them further subdivide into two further dimensions, the individual / group judgements and the origami / clock tower construction.

4.3.1. Judgements

i. Individual estimations…

… of preciseness and especially completion time, aimed to set a benchmark for the participants, which they presumably derived solely by the integration of singular information. Following the concept of Kahneman and Tversky (1979), assumptions about the scope of the task, considering efforts and potential division of labour, most likely lead to the formulation of an independent strategy that was catered towards the specific task features. Hence failing to incorporate distributional information of past performances and

experience. This is supposed to hold particularly true for the origami exercise, where most of the subjects already have a reference value. Additionally, the step by step instruction for each paper fold was further assumed to potentially direct subjects to neglect outside information and instead focus on incremental procedures. Finally, also the influence of (over)confidence was supposed to be higher in the origami task. On the one hand due to the selection of the objects (a sailboat and a koala bear) to fold, which generally appear to be less complex than the prominent crane for instance. On the other hand, not having any prior experience with the construction of a tower out of marshmallows and spaghettis, could trigger intimidation and consequently decrease overconfidence within subjects. However, the earlier mentioned contradicting irrationality within target estimations gets notably significant in the origami task, where even in the face of prior experience, assumptions would still not be based on.

ii. Group estimation …

… and discussion intended to prove the existence of a general group planning fallacy, by examining whether choices towards the task completion were subject to risky shifts. In other words, if reinforced overconfidence through team member interaction would lead to an accentuated misjudgement of time and performance. Active discussion about singular issues (strategies, plans, division of labour) was expected to result in that choice shift towards even shorter time periods. Averaging of the individual benchmark values, on the other side, was expected to follow the wisdom of the crowds effect and balance out over/under estimations. With regards to the higher level of task interdependency for the clock tower construction compared to the origami exercise, inaccuracies in the estimated values were suspected to be higher in the second task.

4.3.2. Tasks

In the process of designing the two different tasks, efforts were concentrated towards creating common intersection, in order to ensure comparability and causality. It was attempted to have exercises with approximately similar durations (10 mins) and levels of complexity. The team tasks should further require some extent (not necessarily the same) of coordination and provide opportunities for division of labour. More distinguishing variables and precise task details are outlined on the following.

i. Origami exercise

The task of folding origami objects for experimental research, has been a team project that can be found in other publications already, yet so far in a slightly different manner. Dan Ariely (2010) for instance, in relation to the earlier discussed Ikea effect, pointed out why we overvalue things that we create, by making people bid for their own origami creatures. Contrary to this ex post evaluation, the setting in this experiment contained an ex ante estimation of team efficiency, in terms of time and preciseness.

Given away in the introduction sheet, students were told that the origamis "have to be built in the right way", meaning that "they should resemble the original [models] as much as possible." Moreover, indicated on the instruction sheet, the exercises required each subject to build "four origamis together with your partner" or in other words "your team has to fold two sailboats and two koalas". The chosen origami models are in terms of complexity two of the easiest objects that could be found in the broad expanses of the internet (see figure 2). This was largely in order to make sure participants could easily comprehend all steps of the instruction and therefore achieve to completely finish the task within a certain time frame (up to 20 mins). The worst case to avoid was that people give up due to excessive difficulty or having to cut the task at a certain time, for progressing motives of the whole experiment.

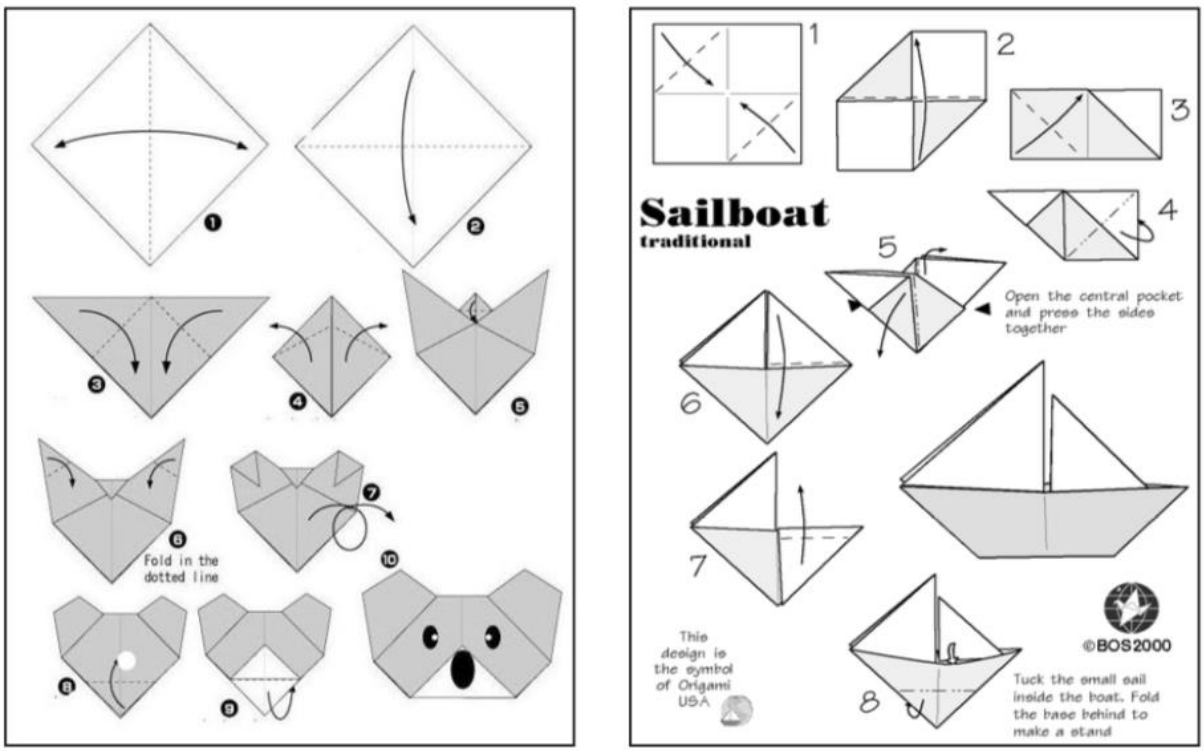


Figure 2: Origami Instructions

The main distinction variable to the second task, however, was the 'parallel' setting of the origami exercise. Asking a team of two people to fold four origami objects (two of each kind), allowed them to divide the task and allocate the functions according to their preferences and competences. If done efficiently, this could made teams gain from additional learning effects.

ii. Clock Tower

Another prominent example in team performance research was introduced by Tom Wujec (2015), called the marshmallow challenge. The setting in his task were the following: "in eighteen minutes, teams must build the tallest free-standing structure out of 20 sticks of spaghetti, one yard of tape, one yard of string, and one marshmallow. The marshmallow needs to be on top." By nature, in the context of a PF experiment, the dimensions of a pre-set deadline has to be crossed out. Taking further away the parts of height maximization and resource

constraints, there was a need to formulate another measurable goal. Having subjects made to assemble the materials in a pre-defined structure, as good and as fast as they can, met the criteria that was demanded. Even though the instructions contained clear and explicit descriptions about how to assemble the spaghettis and marshmallows (figure 4), as well as of which geometric forms the whole tower consists (figure 3), the instructions gave away that "there are several ways to construct the clock tower". Subjects were further told that the final product should match the construction model in the description closely and has to be able to stand on its own for 3 seconds.

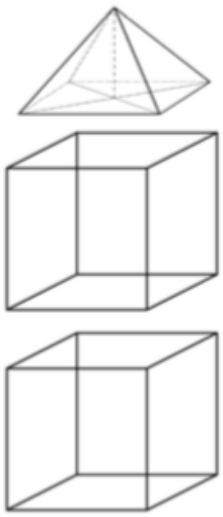


Figure 3: CT Basic Layout

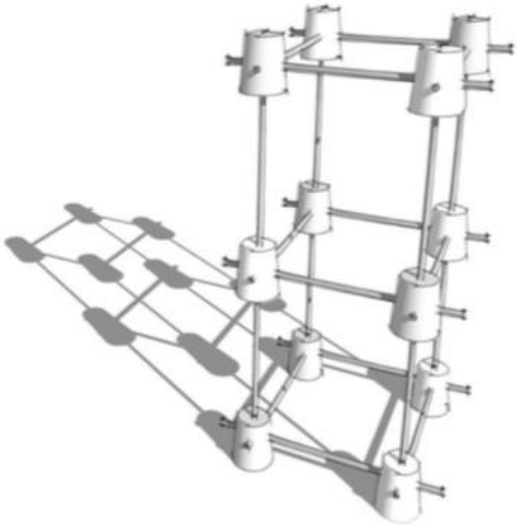


Figure 4: Material Assembly

Again, an ex post evaluation of time was requested, but for this task this was assumed to be even harder compared to the origami task. Unlike the prominent paper folding exercise that the majority of people at least tried and failed once in their lives, the overall experience with marshmallow and spaghetti tower constriction was assumed to be scarce. The additional higher difficulty in conceptualizing the single steps, was further expected to hamper participants deriving a strategy based on singular information, thus consequently to settle on a benchmark.

As mentioned earlier, the most relevant difference to the origami exercise, was the level of task interdependency and coordination of individual efforts. The underlying assumption of building a tower in a team of two, was to induce the students to simultaneously operate on the same object, instead of dividing the task and working side by side. The higher level of interdependence was expected to cause efficiency losses, which would result in even higher project time misjudgements.

5. Results

With regard to the further discussion of the hypothesized concepts from the research design, the main focus within this section has been the participants' estimations about time. The results include individual and group judgements, as well as the actual times needed to complete each of the tasks. Summary statistics describing the actual and estimated total minutes of labour needed to complete the exercises are provided in Table 1. Confirming the comparability of the two exercises, the actual times needed are hovering around the intended length of ten minutes for each task. However, the origami task, in fact required more time to complete (M = 13,23) than the clock tower construction (M = 8,65), with a mean difference of 5,58 minutes.

Table 1: Estimated and actual completion times (in mins) for the two different tasks

Task	Individual estimation	Group estimation	Actual time
Origami			
M	8,96	7,96	13,23
(SD)	(5,64)	(3,62)	(2,33)
Clock Tower			
M	12,02	10,65	8,65
(SD)	(8,89)	(7,07)	(1,98)

Means, standard deviations (in parentheses), sample size n = 46

5.1. Results of working hypothesis 1

Directing the focus on hypothesis 1, dealing with the 'generic' concept of the planning fallacy, the comparison of estimated versus actual times reveals significant differences. Summarized in table 2, estimation errors were calculated by subtracting individual / group estimations (columns one / two, respectively), from the collected actual times (column three) in table 1. Verified with a one sample t-test, mean errors in the origami task for both, individual and group level, were highly significantly different from zero (t[45]=5,83; p<.001; t[45]=12,19; p<.001; respectively). Similar results were found for the marshmallow construction, yet mainly on the individual stage (t[45]= -2,61; p<.05), whereas the group error was slightly not significant (t[45]=-1,88; p=.066).

Table 2: Individual & group estimation errors (in mins)

Task	Individual error	Group error
Origami	4,27	5,27
Clock Tower	-3,37	-2,00

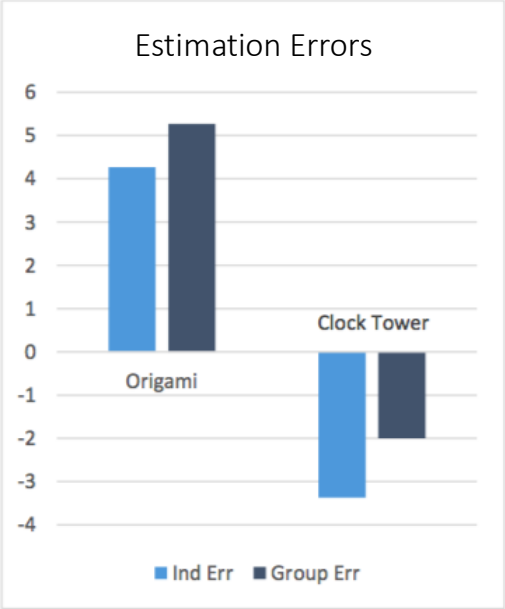


Figure 5: Estimation errors for task 1 and task 2

As illustrated in figure 5, the misjudgements of time are mirror-inverted functions in terms of over / underestimation for the two different tasks. The clearly positive values for the origami folding exercise (4,27/5,27 mins) align with the concept of the planning fallacy, namely that people underestimate completion times for (group) projects. Contrary assumptions have to be drawn on the findings of the tower construction task, where participants actually overestimated their completion times (-3,37 /-2,00), hence do not fit in the classical theory of the planning fallacy.

Subsequent paired samples tests, performed for more profound substance, reveal similar results as shown in figure 6. Effects of underestimation were highly significant for the origami task (t[45]=5,83; p<.001), whereas the mean difference of +3,37 in the clock tower construction suggested a significant (p=.012) overestimation of project time.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	OR Ind Est - OR Act T	-4,274	4,972	,733	-5,750	-2,797	-5,830	45	,000
Pair 2	CT Ind Est - CT Act T	3,373	8,777	1,294	,767	5,979	2,607	45	,012

Figure 6: Paired Samples Test, comparing individual estimations (Ind Est) with actual completion time (Act t), for task 1 (OR = Origami) and task 2 (CT = Clock Tower)

5.2. Results of working hypothesis 2

A group accentuated effect of that planning fallacy, as proposed in hypothesis two, or a general underestimation of to the time frame by groups, was statistically evident. The latter finding is based on similar calculations as previously on the individual level. A paired sample test (figure 7) reveals highly significant differences (t[45] = -12,19; p<.001) between the group estimations and the actual times needed in the origami task. Again, for the clock tower construction, the overall positive test value of +1.88 rather speaks in favour of an overestimation of needed time, however in this case only to an insignificant manner (p=.066).

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	OR Gr Est - OR Act T	-5,274	2,933	,432	-6,145	-4,403	-12,194	45	,000
Pair 2	CT Gr Est - CT Act T	2,003	7,214	1,064	-,139	4,146	1,884	45	,066

Figure 7: Paired Samples Test, comparing group estimations (Gr Est) with actual completion time (Act t), for task 1 (OR = Origami) and task 2 (CT = Clock Tower)

In order to test whether initial tendencies of underestimation were accentuated by interaction and discussion, aggregate values (averages of the individual (group) member estimations), were compared to the group consensus values, expressed as the group estimation. As summarized in table 3, discussion-based group estimations were more optimistic (A/N: indicated by smaller numerical values) than the aggregated initial estimations of subjects, holding true for both tasks.

Table 3: Aggregated and group estimations (in mins)

Task	Aggregated value	Group estimation
Origami	8,96	7,96
Clock Tower	12,02	10,65

Given the previously confirmed group-based underestimation of times in the origami task, the mean difference of one minute between the aggregated individual estimates and the group estimates, proves the existence of an accentuation effect. However, test values from a paired sample test (see figure 8) imply that this tendency was not significantly strong in this setting ($t[22] = 1,38$; $p = .179$). The opposite was found for the marshmallow tower construction, where the mean difference of 1,37 minutes was indeed statistically significant. Based on the fact that teams finished this task earlier than they predicted, the results do not approve an accentuated underestimation of the actual completion time.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	OR Agg - OR Gr Est	1,000	3,458	,721	-,495	2,495	1,387	22	,179
Pair 2	CT Agg - CT Gr Est	1,370	3,181	,663	-,006	2,745	2,065	22	,051

Figure 8: Paired Samples Test, comparing aggregated estimations (Agg) with group estimations (Gr Est), for task 1 (OR = Origami) and task 2 (CT = Clock Tower)

5.3. Results of working hypothesis 3

Incorporating the dimension of interdependency, forthcoming tests were performed to observe the moderating effect of task interdependency on overconfident time predictions, as proposed in hypothesis three. First and foremost, following the experimental design's intent of having two different tasks that significantly differ in their level of task interdependency, was a premise for any further testing of the theoretical framework. In the post experimental feedback questionnaire (see appendix: figure 17), subjects reported their perceived level of coordination, communication and team work needed in order to successfully complete the task. The results of these three variables were then aggregated to represent the perceived level of interdependence for each task. In accordance with the experimental design, participants indicated higher task independency for the marshmallow tower than for the origami folding (table 4).

Table 4: Perceived level of task interdependency

Task	Mean	Standard deviation
Origami	3,27	0,97
Clock Tower	4,36	0,75

N=46

Moreover, a paired sample test was performed to highlight statistical differences in terms of interdependency between the two tasks. As figure nine reveals, the results were highly significant (p<.001).

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	OR Intdp - CT Intdp	-1,087	,879	,130	-1,348	-,826	-8,388	45	,000

Figure 9: Paired Samples Test, comparing task interdependency (Intdp) of task 1 (OR = Origami) and task 2 (CT = Clock Tower)

Based on these insights, a subsequent analysis of variances, characterized by its enhanced statistical power, was executed to investigate both, the interactions between as well as the effects of the individual factors, task interdependency and time misjudgement. Given the underlying existence of causality in the research model, the remaining just proven significant distinction point between task one and two, is the factor of task interdependency. Therefore, a significant main effect of the project component (in figures 10 and 11 entitled as "task") would suggest statistical differences of completion time in the face of differences in the interdependence level. These main effects were calculated for both, individual estimations (figure 10) and group estimations (figure 11), to examine whether they hold true on both conditions. In other words, whether the effect of task independency on overconfident time misjudgements is present in both settings. Further comparisons of the effect intensities were supposed to provide additional insights whether groups are more prone to that fallacy.

The results from a 2 (*task*: origami (OR) vs clock tower (CT)) x 2 (*structure*: individual vs actual) within-subjects ANOVA (figure 10) report that the main effect of *task* was slightly insignificant (F [1, 45] = 2,16, p = .148). Thus, indicating that, even in the face of significant

differences in the interdependency level between the two different tasks, individual time estimations were not additionally biased.

Descriptive Statistics

	Mean	Std. Deviation	N
OR Ind Est	8,96	5,641	46
OR Act T	13,23	2,326	46
CT Ind Est	12,02	8,893	46
CT Act T	8,65	1,983	46

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
task	Sphericity Assumed	26,448	1	26,448	2,163	,148
	Greenhouse-Geisser	26,448	1,000	26,448	2,163	,148
	Huynh-Feldt	26,448	1,000	26,448	2,163	,148
	Lower-bound	26,448	1,000	26,448	2,163	,148
Error(task)	Sphericity Assumed	550,246	45	12,228		
	Greenhouse-Geisser	550,246	45,000	12,228		
	Huynh-Feldt	550,246	45,000	12,228		
	Lower-bound	550,246	45,000	12,228		
structure	Sphericity Assumed	9,333	1	9,333	,245	,623
	Greenhouse-Geisser	9,333	1,000	9,333	,245	,623
	Huynh-Feldt	9,333	1,000	9,333	,245	,623
	Lower-bound	9,333	1,000	9,333	,245	,623
Error(structure)	Sphericity Assumed	1713,719	45	38,083		
	Greenhouse-Geisser	1713,719	45,000	38,083		
	Huynh-Feldt	1713,719	45,000	38,083		
	Lower-bound	1713,719	45,000	38,083		
task * structure	Sphericity Assumed	672,473	1	672,473	52,562	,000
	Greenhouse-Geisser	672,473	1,000	672,473	52,562	,000
	Huynh-Feldt	672,473	1,000	672,473	52,562	,000
	Lower-bound	672,473	1,000	672,473	52,562	,000
Error(task*structure)	Sphericity Assumed	575,721	45	12,794		
	Greenhouse-Geisser	575,721	45,000	12,794		
	Huynh-Feldt	575,721	45,000	12,794		
	Lower-bound	575,721	45,000	12,794		

Figure 10: 2x2 ANOVA - Individual Estimations vs Actual Completion Times

Similar procedures were applied for the group comparison, depicted in figure 11. Taking into consideration the insights from prior discussion on hypothesis two, the results of this two times wo ANOVA unsurprisingly indicate that there are significant differences for the main effect of *task* in the group setting (F [1,45]= 5,36, p = .025). Hence, advocating that different levels of task interdependence lead to different results in the estimation discrepancy dilemma. However, notabene, the numbers only confirm a deviation of estimations from actual times, but certainly not the direction. Therefore, drawn on earlier arguments, the effect of interdependence, which

was apparently significant in group estimations, does not approve an underestimation at all, but only a significant misjudgement of time.

Descriptive Statistics

	Mean	Std. Deviation	N
OR Gr Est	7,96	3,621	46
OR Act T	13,23	2,326	46
CT Gr Est	10,65	7,072	46
CT Act T	8,65	1,983	46

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
task	Sphericity Assumed	40,909	1	40,909	5,386	,025
	Greenhouse-Geisser	40,909	1,000	40,909	5,386	,025
	Huynh-Feldt	40,909	1,000	40,909	5,386	,025
	Lower-bound	40,909	1,000	40,909	5,386	,025
Error(task)	Sphericity Assumed	341,785	45	7,595		
	Greenhouse-Geisser	341,785	45,000	7,595		
	Huynh-Feldt	341,785	45,000	7,595		
	Lower-bound	341,785	45,000	7,595		
time	Sphericity Assumed	123,001	1	123,001	5,455	,024
	Greenhouse-Geisser	123,001	1,000	123,001	5,455	,024
	Huynh-Feldt	123,001	1,000	123,001	5,455	,024
	Lower-bound	123,001	1,000	123,001	5,455	,024
Error(time)	Sphericity Assumed	1014,751	45	22,550		
	Greenhouse-Geisser	1014,751	45,000	22,550		
	Huynh-Feldt	1014,751	45,000	22,550		
	Lower-bound	1014,751	45,000	22,550		
task * time	Sphericity Assumed	609,045	1	609,045	78,384	,000
	Greenhouse-Geisser	609,045	1,000	609,045	78,384	,000
	Huynh-Feldt	609,045	1,000	609,045	78,384	,000
	Lower-bound	609,045	1,000	609,045	78,384	,000
Error(task*time)	Sphericity Assumed	349,649	45	7,770		
	Greenhouse-Geisser	349,649	45,000	7,770		
	Huynh-Feldt	349,649	45,000	7,770		
	Lower-bound	349,649	45,000	7,770		

Figure 11: 2x2 ANOVA - Group Estimations vs Actual Completion Times

An additional comparison of the effect strength gives away that groups (F = 5,36) tend to display higher likelihood to fail deadlines for tasks with higher interdependencies between group members than individuals (F= 2,16) do.

6. Discussion

Given the nature of the results, the discussion of the findings asks for an early limitation to the experimental design. Contrary to the intended similar task lengths that provide comparability, the actual completion times differed in 5,58 minutes, representing a total deviation of 42%. However crucial for the further discussion is that finishing times for the two tasks point in two different directions from the intended length, resulting in over- and underestimations. To some extent this might has been due to learning effects, which are going to be addressed in detail shortly, but in a repeated study with almost similar settings, the design would most likely have been adapted, to counteract these findings.

6.1. Discussion of working hypothesis 1

Tackling the overall topic from the grassroots, hypothesis one was intending to replicate findings from well-established past research, dealing with the generic model of the planning fallacy. The results from the present study partly align with the classical models from Kahneman and Tversky (1979), Buehler, Griffin and Ross (1994) or Roy et al. (2005). To be more accurate, the general tendency to underestimate the task complexity and consequently the time and efforts needed to finish it, was significantly present in the origami folding exercise. On average, individuals confidentially underestimated the completion time by 4,27 minutes out of 13,23, representing a judgemental error by something more than 32%.

With the second task however, the numbers contradict the expected results and therefore partly the first hypothesis of this paper. Participants predicted the completion time to be considerably higher than it actually was, with an average error of 3,37 minutes out of 8,65, thus by almost 40%. Explanations for these findings can be related to several factors. First, bringing up the nature of the task again, the likelihood that participants have already prior experience with folding an origami object is assumed to be higher by certain factor, than crafting a tower out of spaghetti and marshmallows. Hence there are two conclusions drawn on that. On the one

hand subjects most likely will not have had a reference point of the resources needed when they were asked to provide a precise estimation. The lack of such a benchmark could have led, as discussed by Kahneman and Tversky (1979), that participants fall back on singular information predictions, derived by strategic step by step thinking. As examined at an earlier occasion already, this would have rather led to an underestimation of the time needed, which is why an alternative and perhaps even simpler explanation should be favoured. The way the estimation sheet was outlined, subjects were asked for time and success predictions for the marshmallow task and origami task in this order. Therefore, in the face of a missing reference point for the second task, a comparison to the first one seems not far to seek. This, leading to the second conclusion, that people did overestimate the level of complexity of the second task, and by doing so, when having estimated the first task with 8.96 minutes on average, just added up an assumed extra time needed and settled on a prediction that was 34% higher (12.02 minutes). However, following the concepts of Boltz, Kupperman and Dunne (1998), they argue that for tasks that are not time-consuming or task with a low level of complexity, people are more likely to overestimate completion times and thereby underestimate their future performance.

Another explanation for the surprising findings of this study could be linked to missing incentives to finish the task as fast as possible, e.g. in terms of monetary rewards. As Bower, Ashby, Gerald & Smyk (2002) argued, the process of incentivization should motivate an agent to achieve extra 'value-added' services over those specified originally. So, in a professional environment results may differ from this experimental study, due to the presence of this exact factor. Further they mentioned that for the dimension of time delays, penalties are even more efficient than incentives. A slightly different explanation was given by Burt and Kemp (1994), namely that the tendency of making overestimations is a strategy that gives a feeling of control over time and helps avoid stress caused by an inability to complete tasks in the allocated time.

6.2. Discussion of working hypothesis 2

Taking a closer examination of the results of the second hypothesis, the two different tasks share, at least for the group accentuation effect, significant similarities. The proposed trend of a general underestimation of time frames by groups, was again only statistically evident in one of the two exercises.

In resembling manner as the pioneer contribution by Buehler et al. (2005) about the group planning fallacy or it's slightly modified replication by Staats, Milkman and Fox (2012), the experimental design of the origami task produced overlapping findings. To put it in the words of the latter researchers, estimators are "relatively insensitive to the impact that team size can have on the total amount of effort that a project requires" (p.135). The numerical confirmation of this argument is given by the large discrepancy of the average group estimation value (7,96 mins) compared to the average actual time needed (13,23) to finish the origami objects as given in the preview. The corresponding paired sample test further was highly significant ($p<.001$).

On the other hand, the clock tower construction, delivered alternative results to the ones just mentioned. Although not being statistically significant, the overall tendency did not align with the concept of overconfident time predictions. Subjects rather underestimated their performance, hence overestimated completion time (10,65 mins vs 8,65 mins, estimated vs actual, respectively). In the group setting, potential explanations for these findings can be linked to various factors. Among others, learning effects, due to repeated group interaction in randomly assembled groups, are of relevant magnitude. More precisely, in accordance with Hall and Williams (1966), the performance level of groups increases with repeated interactions of its members. Mainly due to superior communication and abilities to deal with conflicts and opposing opinions. Broken down to the symptoms of the present study, superior performance in the second task (13,23 / 8,65 mins, origami / clock tower, respectively), might have been a consequence of group members gaining familiarity of the strengths and weaknesses of their teammates, hence were able to adopt their strategies on dividing the task more efficiently.

This explanation, admittedly, bears a limitation to the generalizability of the findings, namely not having altered the chronological order of tasks in the experiment for half of the participants. This could have contributed to counterbalance distorting effects of learning and group interaction.

Nonetheless, current findings of this experiment significantly, and not just partly, stand with established concepts, in the dimension of estimation shifts. As Buehler et al. (2005) found, the bias of optimisms gets accentuated by group discussion. In the current experimental analysis, individual group member estimations were aggregated and subsequently compared to the group consensus values, indicated as the group estimation. Results in both of the tasks reflect a noticeable tendency of optimistic estimation shifts. While in the origami task, the mean difference amounted to exactly one minute (8,96; 7,96), in the marshmallow and spaghetti crafting exercise, group discussion induced estimations to decrease from 12,02 to 10,65 minutes (Δ 1,37). A potential explanation for that phenomena could be that through the discussion, participants' attention was drawn on information implying the task could be completed quickly, and in particular on their plans and strategies for achieving this goal. Thus, being affected by the already mentioned selective focus dilemma (Buehler et al., 2005; Staats, Milkman and Fox, 2012). Given the relatively small sample size of 23 teams, results from a paired sample test, did not (or almost) reflect sufficient statistical significance [p(OR) =.179; (p(CT) =.051)]. However, backing up the findings of this study, in both tasks, there was a noticeable increase of optimism owing to group discussion and in both tasks, this shift amounted to something more than eleven percent (OR= 11,16% and CT= 11,39%).

As an approach of novelty, this analysis of the group planning fallacy aims to further address the process of reaching a collaborative consensus. Even though not being separately included in the research design or statistically in depth examined by various tests, an observation of subjects during the experiment and a post-evaluation of their provided answers, give insights

of the applied decision schemes. They also could contribute to explain slightly insignificant accentuation effects, treated in the previous paragraph. By doing so, numbers confirm that the anticipated effect of reinforcing initial overconfident tendencies, accentuation, was indeed noticeable, but mainly in homogenous group constellations (gender, age, mother tongue, nationality). The larger proportion of the samples, however, were heterogenous groups, which fell back on averaging in order to reach a final group prediction, balancing out individual over- and underestimations within a team.

Other than those explicitly expressed demographic factors, variables like social relation, common history and familiarity among team members were crucial in defining which approach to pursue to reach consensus. Similar to heterogenous groups, ad-hoc groups with unfamiliar group members, did not involve in profound discussions by contradicting or reinforcing each other's opinions. They rather converged towards a final joint estimation, midway from their initial points of view. By contrast, within teams where members shared some sort of personal relation beforehand, an exchange of private predications was mostly followed by ambitious attempts of persuasion, by elaborating on task details and possible strategies to tackle them.
The overall larger number or frequency of teams rather choosing averaging as the appropriate way to settle on a value, was to a certain extent due to the higher number of diverse group constellations, which was an intentional attempt to maximize variety among teams to guarantee the representativeness of the sample. Thus, being an indicator for group constellation frequency, and not for the (expected) quality of the decision strategy.

Speaking of performance, Hall and Williams (1966) found that established groups are significantly superior to ad hoc groups in decision processes. A major reason for that is the way to deal with conflict and opposing opinions. Whereas ad hoc groups more likely resolve conflict through compromise procedures, mostly leading to faster but not necessarily optimal solutions, established groups responded with increased creativity and sensitivity.

Next to those subjectively observed interrelations of participants, tangible statements about their personal characteristics also had a considerable impact on performance metrics. Similar to the findings of this thesis, Keck and Tang (2017) found that in gender homogenous (all male) groups, the minor level of interpersonal sensitivity has a direct and negative impact on the quality of estimations, driven by lower information sharing. Thus, compared to simply averaging of individual judgements, a group discussion in a gender homogenous setting, most likely results in an inferior outcome. Taking a closer look on the judgmental errors from both tasks of the experiment, heterogenous groups tendentially performed better than homogenous groups, since they predominantly relied on averaging estimations. Intense discussions in heterogenous team structures, which according to Keck and Tang (2017) would also outperform homogenous discussion estimations, were not sufficiently represented in the sample.

6.3. Discussion of working hypothesis 3

Thematised in the final hypothesis, the relationship between overconfidence and the preciseness of time predictions was supposed to be intensified by tasks that require the executive agents of a team to commonly perform. Task interdependence, or as Langfred (2000) puts it, the "structural constraint on the ability of group members to successfully complete tasks" (p.56), was perceived significantly (t[45]=8,39; p<.001) different in the two exercises of the experiment. As shown in table 5, not only on the aggregated averages of the two tasks this was evident, but also for each of the components. Therefore representing the fundamentals that make direct assumptions about moderating forces possible.

Table 5: Interdependence level summary

	Origami	Clock Tower
Coordination	3,19	4,23
Communication	3,42	4,38
Team work	3,19	4,42
Interdependence	3,27	4,36

N=46

Regardless of requirements having been met, hypothesis three has to be rejected without any need of complex calculations. The fact that the second task simply not even took shorter than the first one, but also shorter than on average expected by individuals and groups. Hence contradicting the concept of moderating effects of task interdependence. Appropriate reasoning for these results vary from very trivial explanations to more sophisticated academic concepts.

Not far to seek is definitely the idea of synergies. Opposite to the assumed process losses due to increases coordination and team efforts needed, the division of labour bears several upsides. For example, the development and/or selection of experts, the use of multiple perspectives and information, team learning, and social identification processes all contribute to performance gains in teams (Hertel, 2011).

An alternative concept was introduced by Langfred (2000), considering the dimensions of task interdependence ('structural constraint') and group autonomy ('imposed structuring of group work'). He argued that under high task interdependence, there would be a positive relationship between group autonomy and group effectiveness. Linked to the findings of the present study, in the clock tower construction, which in fact was subject of high task interdependence, the increased group anatomy ("note that there are several ways to construct the clock tower", Appendix: figure 14) could have led to the higher effectiveness.

Turning back the attention of the discussion on the results from the previous chapter, the ANOVA on the individual judgements reported an insignificant relation of task interdependence and time estimation errors. The same testing on the group dimension, on the other hand, indicated a significant relation of these variables. Therefore arguing that due to the before mentioned estimation shifts in group discussion, more optimistic judgements lead to a stronger effect, than on the individual level. Again, this does support the argument of hypothesis three, namely that the (group) planning fallacy would be intensified by higher task interdependence, but rather aligning with the concepts of unexpected synergies and the positive externalities of the interdependence and autonomy interactions.

7. Conclusion

Collaborative failure in time and efficiency estimations, or rephrased in a softer manner, the central human challenge of linking past experiences to our understanding of the present and our expectations for the future, was the underlying subject of this thesis. This dilemma is even amplified by the social interaction with our surroundings. Yet all of us make predictions every day in our lives and often they are overoptimistic or just plain irrational. Speaking of paradoxes, examples of well renowned public figures are likewise not an exception, as a former British prime minister approved - "I don't make predictions, I never have, and I never will".

The general cognitive bias of rating one's own pessimistic scenarios much less plausible than the optimistic ones, leading to the best guess scenario being closely aligned to the latter one, finds its utterly form of damage in us spreading our 'knowledge'. Hereby affecting other people's thought processes and accentuating initially wrong tendencies. Especially professional teams that have been established for a while, are often predisposed to neglect additional outside information. A homogenisation of team members, due to normative and social values, leads to a tendency towards more unrealistic predictions about deadlines and resources.

Finally, next to the constellation of the team and their norms, the nature of the task was supposed to have a significant stake in the accuracy of predictions. Group tasks of a more independent kind (executed parallelly by different agents), are due to the integration of observer perspectives for the other group members' achievements, a smaller tendency of being misjudged.

The quantitative research of this thesis tried to investigate those ideas and models by an experimental study with twenty-three two-person teams. Participants were asked to perform in two different crafting exercises, one after another, in teams of two. Prior, they had to individually and subsequently with their group partner, estimate their needed time to complete each of the tasks as well as their performance level. The private initial estimation that was followed by a group discussion-based consensus judgement was supposed to highlight potential

estimation shifts towards their confidence level. Differences in the two task characteristics, were tried to be reduced to the dimension of interdependence, by leaving the level of complexity and time needed on similar terms, in order to guarantee causality.

The research design consisted of two different layers with three hypotheses. The first part aimed to replicate 'well-established' concepts of human interaction in project management realms. The underlying assumption was that individuals and, even more so, groups, are prone to make predictions about future scenarios that were overoptimistically biased. The second proposition was incorporating a missing piece in the puzzle of existing contributions. Structural constraints to the nature of the task (interdependencies) were supposed to intensify the negative relationship of overconfidence and time predictions.

The results gathered from the experiment, partly support and partly contradict prior research findings. Additionally, they provide novel insights about aggregated decision processes. Namely the integration of the subjects' demographic factors and social relations, in order to explain the origin of discussion-based estimations shifts towards more optimistic judgements. Doing so, the results indicate that homogenous groups (gender, age, mother tongue, nationality) as well as established groups (contrary to ad-hoc) in fact are prone to accentuated (biased) predictions. In accordance with existing literature, a general individual and group tendency to underestimate time and efforts needed to complete a task, was significantly evident in the origami folding exercise. The marshmallow and spaghetti tower, on the other hand, did not fit in the framework of the planning fallacy, since completion times were undercutting individual as well as group estimations. Thereby highlighting the often in research neglected phenomena of (unexpected) synergies. These process gains were basically the main driver to reject hypothesis three, which was proposing and intensified group planning fallacy because of higher task interdependence.

7.1. Practical implications

The results of the origami exercise are especially interesting in practice, since they reflect a common pitfall that is related to the features of a task itself. In a professional environment, within the frontiers of everyday work, agents will be confronted with tasks that they are 'anyway' familiar with and capable of executing. Consequently, their confidence is high enough to underestimate the complexity. Contrary, as already mentioned in chapter 6, the marshmallow and spaghetti tower construction appeared to be totally obscure, making it more difficult to estimate it and leaves subjects behind with larger respect of the increased novelty. Talking about an average hierarchical system in the professional environment, subjects would most likely rather be confronted with task that to some extent fall under their competences and do not require them to reinvent the wheel every time.

Besides of matching the competence and prior experience of managers with their assigned projects, when the level of within-group task interdependence is predetermined by the organization's technology, it is crucial for the success to carefully match the degree of autonomy. For instance, in a sequential-dependent task nature, where agents often wait for contributions of others before performing their own part, increasing autonomy, enables them to adapt their viewpoints to more observant stances that might give rise to an outside perspective and mitigate the planning-based optimistic bias. However, in highly decentralised team structures, increased autonomy might lead to higher planned risk seeking and consequently to higher levels of actual risk seeking behaviour, which in turn negatively related to multiteam system performance.

7.2. Limitations & future directions

As already shortly addressed in the discussion above, one feature that may limit the generalizability of the present study was, is the inexistence of a between subject comparison. Potential synergies that might have led in the second task to earlier completion times, could have been counterbalanced by switching the task order for half of the participants.

A further limitation of this, and any other study relying on perceptual reported data, is the bias inherent in such observations. Yet using such data to make judgements about applied decision schemes, does not fulfil the standards of quantitative research. Not having undertaken in-depth statistical evaluation of these findings in fact further limit the generalizability of the arguments but was following the intent of not building numerical equations on intangible data.

An obvious direction for future research is the establishment of quantifiable models dealing with decision process structures in regard to the overall topic of the planning fallacy. Thereby also enabling a more sophisticated analysis of group compositions and member demographics. Likewise, worth to examine would be whether cultural dimensions and normative guidelines (e.g. not to disagree with a stranger), would produce different results in different environments for the overconfident group estimation dilemma.

References

Alicke, M. D., & Govorun, O. (2005). *"The better-than-average effect*". In M. D. Alicke, D. Dunning & J. Krueger (Eds.), "*The self in social judgment"* (pp. 85-106). New York: Psychology Press.

Ariely, Dan (2010). „*The upside of irrationality: The unexpected benefits of defying logic at work and at home"*. New York: Harper

Ariely, Dan and Wertenbroch, Klaus (2002). *"Procrastination, Deadlines, and Performance"*. American Psychological Society Vol. 13, No. 3, May 200

Aronson, Elliot (2010). "*Social Psychology".* Upper Saddle River, NJ: Prentice Hall. p. 273

Boltz, Marilyn G.; Kupperman, Cara and Dunne, Jessica (1998). *"The role of learning in remembered duration"*; Memory & Cognition; September 1998, Volume 26, Issue 5, pp 903–921

Bower, D., Ashby, G., Gerald, K., & Smyk, W. (2002). "*Incentive Mechanisms for Project Success".* Journal Of Management In Engineering, 18(1), 37-43.

Bray, R. M.; A. M. Noble (1978). "*Authoritarianism and decisions of mock juries: Evidence of jury bias and group polarization*". Journal of Personality and Social Psychology. 36 (12): 1424–1430

Buehler, Roger; Griffin, Dale, & Ross, Michael (2002). *"Inside the planning fallacy: The causes and consequences of optimistic time predictions"*. In Thomas Gilovich, Dale Griffin, & Daniel Kahneman (Eds.), Heuristics and biases: The psychology of intuitive judgment, pp. 250–270. Cambridge, UK: Cambridge University Press.

Buehler, Roger; Dale Griffin; Michael Ross (1995). *"It's about time: Optimistic predictions in work and love"*. European Review of Social Psychology. American Psychological Association. 6: 1–32)

Buehler, Roger; Dale Griffin; Michael Ross (1994). *"Exploring the "planning fallacy": Why people underestimate their task completion times"*. Journal of Personality and Social Psychology. American Psychological Association. 67 (3): 366–381

Buehler, Messervey, Griffin (2005) *Collaborative planning and prediction: Does group discussion affect optimistic biases in time estimation*?; Organizational Behavior and Human Decision Processes 97 (2005) 47–63

Burt, C. D. and Kemp, S. (1994), *Construction of activity duration and time management potential*. Appl. Cognit. Psychol., 8: 155-168.

Cialdini, Robert (2009). *"Influence: Science and Practice"*. Boston, MA: Pearson Education.

Clemens, R. T., & Winkler, R. L. (1993). *Aggregating point estimates: A Xexible modeling approach*. Management Science, 39, 501-515

Davis, J. H. (1996). *Group decision making and quantitative judgments: A consensus model*. In E. Witte & J. H. Davis (Eds.), Under- standing group behaviour: Consensual action by small groups (Vol. 1, pp. 35–59). Hillsdale, NJ: Erlbaum.

Dichter, E. (1964). „*Handbook of consumer motivations: The psychology of the world of objects"*. New York: McGraw-Hill.

Ernst and Young (2014). *"Spotlight on oil and gas megaproject*"; Oil and gas capital projects series; 2014 EYGM Limited.

Gallup (2012). "*The truth about how we think*". Gallup Business Journal, March 20, 2012

Gully, S. M., Incalcaterra, K. A., Joshi, A., & Beaubien, J. M. (2002). *A meta-analysis of team-efficacy, potency, and performance: Interdependence and level of analysis as moderators of observed relationships*. Journal of Applied Psychology, 87, 819-832.

Hall, J., & Williams, M. S. (1966). *A comparison of decision-making performances in established and ad hoc groups*. Journal of Personality and Social Psychology, 3(2), 214-222.

Hertel, Guido (2011) "Synergetic effects in working teams", Journal of Managerial Psychology, Vol. 26 Issue: 3, pp.176-184

Hinsz, V. B. (1999). *Group decision making with responses of a quantitative nature: The theory of social decision schemes for quantities.* Organizational Behavior and Human Decision Processes, 80, 28–49.

Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. (1997). *The emerging conceptualization of groups as information processors.* Psychology Bulletin, *121*, 43–64.

Hirschman, A. O. (1967). "*Development Projects Observed"*. Washington, D.C.: The Brookings Institution

Kahneman, Daniel (2011). "*Thinking, fast and slow".* New York: Farrar, Straus and Giroux

Kahneman, D., & Lovallo, D. (1993). *Timid choices and bold forecasts: A cognitive perspective on risk taking.* Management Science, 39, 17–31.

Kahneman, D., & Tversky, A. (1977). *"Intuitive prediction: Biases and corrective procedures".* Management Science. 12, 313–327.

Kalven, H. G., Jr., & Zeisel, H. (1966). "*The American jury".* Boston: Little, Brown

Keck, S. and Tang, W. (2017)."*Gender composition and group confidence judgment: The perils of all male groups*", Management Science, Articles in Advance, pp. 1–22

Kerr, N. L., MacCoun, R. J., & Kramer, G. P. (1996). *Bias in judgment: Comparing individuals and groups*. Psychological Review, 103, 687–719.

Kerr, N. L., & Tindale, R. S. (2004). *Group performance and decision making*. Annual Review of Psychology, 55, 623–655.

Kozlowski, S. W. J., & Bell, B. S. (2001). *Work groups and teams in organizations*. In W. C. Borman, D. R. Ilgen, R. J. Klimoski, & I. B. Weiner (Eds.), Handbook of psychology: Industrial and organizational psychology. New York: Wiley.

Kruger, Justin (1999*). "Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments".* Journal of Personality and Social Psychology. 77 (2): 221–232.

Kruger, J., Dunning, D. (1999). *"Unskilled and unaware of it. How difficulties in recognizing one's own incompetence lead to inflated self-assessments"*. Journal of Personality and Social Psychology. 77, Nr. 6, 1999, pp. 1121–1134

Lamm, H. and Myers, D. (1978). "Group-Induced Polarization of Attitudes and Behavior". Advances in Experimental Social Psychology, pp.145-195.)

Lanaj, K., Hollenbeck, J. R., Ilgen, D. R., Barnes, C. M., & Harmon, S. J. (2013). "*The double-edged sword of decentralized planning in multiteam systems*." Academy of Management Journal, 56(3), 735-757

Langfred, C. W. (2000). 'Work-group design and autonomy: a Field Study of the Interaction Between Task Interdependence and Group Autonomy'. Small Group Research, vol. 31, no. 1, p. 54 - 70.

Lovallo, D., & Kahneman, D. (2003). *"Delusions of Success: How Optimism Undermines Executives' Decisions"*. Harvard Business Review: 56–63.

Mannes, Albert E. and Moore, Don A. (2013). *"A Behavioral Demonstration of Overconfidence in Judgment"*, Psychological Science Vol 24, Issue 7, pp. 1190 – 1197

Moore, Don A.; Healy, Paul J. (2008). *"The trouble with overconfidence"*. Psychological Review. 115 (2): 502–517.

Moscovici, S., & Zavalloni, M. (1969). "The group as a polarizer of attitudes". Journal of Personality and Social Psychology, 12(2), 125-135.

Myers, D.G.; H. Lamm (1976). "The group polarization phenomenon". Psychological Bulletin. 83: 602–627

Myers, D.G.; Lamm, Helmut (1975). "*The polarizing effect of group discussion*". American Scientist. 63 (3): 297–303.

Norton, Michael I., Daniel Mochon, and Dan Ariely. (2011). *"The IKEA Effect: When Labor Leads to Love."* Journal of Consumer Psychology 22, no. 3 (July 2012): 453–460.

Plous, S. (1993). "*The psychology of judgment and decision making"*. New York: McGraw-Hill.

Read, Daniel; Loewenstein, George; Kalyanaraman, Shobana (1999). *"Mixing virtue and vice: combining the immediacy effect and the diversification"*. Journal of Behavioral Decision Making; Dec 1999; 12, pp.257-273

Roy, Michael M.; Christenfeld, Nicholas J. S.; McKenzie, Craig R. M. (2005). "*Underestimating the Duration of Future Events: Memory Incorrectly Used or Memory Bias?"*. Psychological Bulletin. 131 (5): 738–756.

Shakespeare, William (1986). *"The Globe illustrated Shakespeare. The complete works, annotated, Deluxe Edition, Hamlet, Act II, scene 2, page 1879"*. Greenwich House, Inc., New York, USA.

Shea, G.P. & Guzzo, R.A. (1987). *Group Effectiveness: What really matters*? Sloan Management Review, 28(3), 25-31.

Sniezek, J. A., & Henry, R. A. (1990). *Revision, weighting, and commit- ment in consensus group judgment.* Organizational Behavior and Human Decision Processes, 45, 66–84.

Staats, Milkman, Fox (2012) *The team scaling fallacy: Underestimating the declining efficiency of larger teams;* Organizational Behavior and Human Decision Processes 118 (2012) 132–142

Stoner, J.A. (1961). "*A comparison of individual and group decision involving risk"*. Unpublished master's thesis, Massachusetts Institute of Technology.

Svenson, Ola (1981). *"Are we all less risky and more skillful than our fellow drivers?"*. Acta Psychologica. 47 (2): 143–148

Thaler, R.H., & Shefrin, H.M. (1981). *"An economic theory of self-control"*. Journal of Political Economy, 89, 392–406.

Thompson, J. D. (1967). *Organizations in action.* New York: McGraw- Hill.

Tindale, R. S. (1993). Decision errors made by individuals and groups. In N. J. Castellan (Ed.), *Individual and group decision making: Cur- rent issues* (pp. 109–124). Hillsdale, NJ: Erlbaum.

Twain, Mark (1923). "*Europe and Elsewhere*", Harper & Brothers Publishers, New York 1923

Zenger, T. R. (1992). *"Why do employers only reward extreme performance? Examining the relationships among performance, pay, and turnover*". Administrative Science Quarterly, 37, 198-219.

Yardi, Sarita; Danah Boyd (2010). "Dynamic Debates: An analysis of group polarization over time on Twitter". Bulletin of Science, Technology and Society. 30 (5): 316–27.

Online

Derstandard (2017). „*Mehrkosten: Wien gesteht "Fehlentscheidungen" bei Krankenhaus Nord ein*; By Oona Kroisleitner; November 21, 2017, 17:27

Original text: „Ja, es gab Fehlentscheidungen in der Vergangenheit"; „Wir wussten schon, dass die Kosten über der Planung liegen und dass der Bau sich verzögern wird"; „…ein Projekt dieser Komplexität so nicht durchsetzbar ist."

Accessed 18.12.2017 via http://derstandard.at/2000068209187/Rechnungshofbericht-Offenbar-mehr-als-8-000-Maengel-bei-Spital-Nord

NYT (2016). *Santiago Calatrava's Transit Hub Is a Soaring Symbol of a Boondoggle*; By Michael Kimmelman; March 2, 2016

Accessed 18.12.2017 via https://www.nytimes.com/2016/03/03/arts/design/santiago-calatravas-transit-hub-is-a-soaring-symbol-of-a-boondoggle.html

Wujec, Tom (2015) Marshmallow Challenge; Accessed 15.06.2018 via https://www.tomwujec.com/design-projects/marshmallow-challenge/

Appendix

List of Figures

List of Tables

Introduction

Thank you for participating in this experiment!

Task

On the following you will perform two different tasks, one after another. Both of them you will tackle in teams of two. The first task is an Origami exercise as shown on handout A. The second task is a construction exercise as shown on handout B. You will solve both tasks with the same partner. Once you have familiarized yourself with the instructions of both exercises, we ask you to **individually estimate** the time it will take your team to complete the tasks as well as to which extent your sculpture(s) will resemble the models. You have 3 minutes to do so, **please don't share your ideas with others**.

While we are collecting your individual estimations, please get together with your neighbour, to form teams of two. You will be given a **group estimation** sheet that we will ask you to fill out thereupon, answering the same questions again. You are allowed to discuss your ideas.

Goal

For the origami, it is necessary that you build it in the right way, which means they should resemble the original (handout A) as much as possible. After you finish building the sculptures, we will assess whether they are correct or not.

For the construction task, you have to build a clock tower out of marshmallows and spaghettis. The final object should again match the construction plan (handout B), however, there is no 'right way' to proceed. Once you are ready we will again assess the correctness.

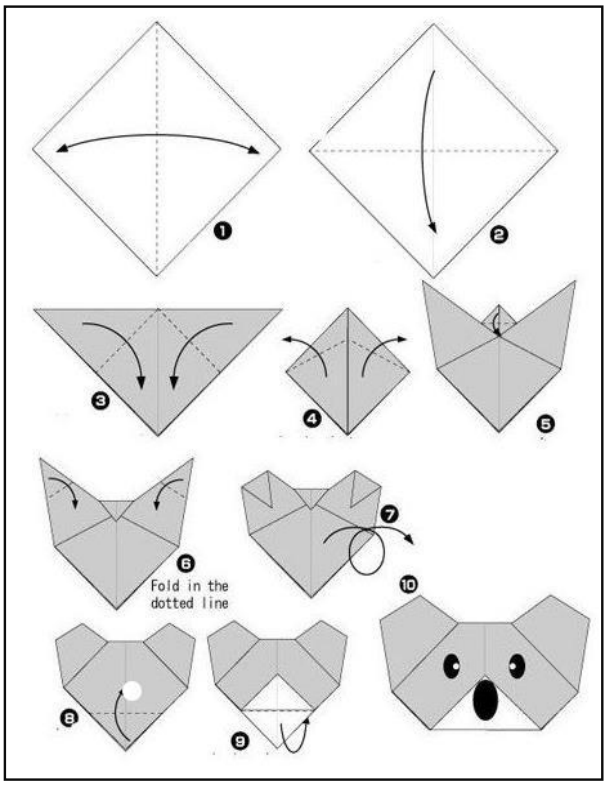
Data Collection

The personal information collected in this experiment will only be stored for research purposes, will only be made available to members of the research group, and will not be passed on. Results of your experiments will only be published in anonymous fashion. Please contact us if you have questions on this.

Good look and have fun!

Figure 12: Introduction

Handout A



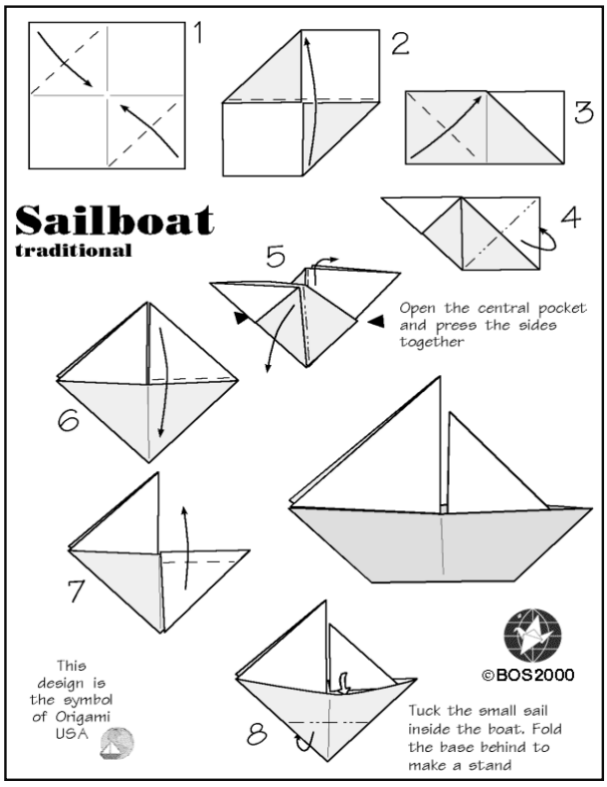


Figure 13: Origami Instructions

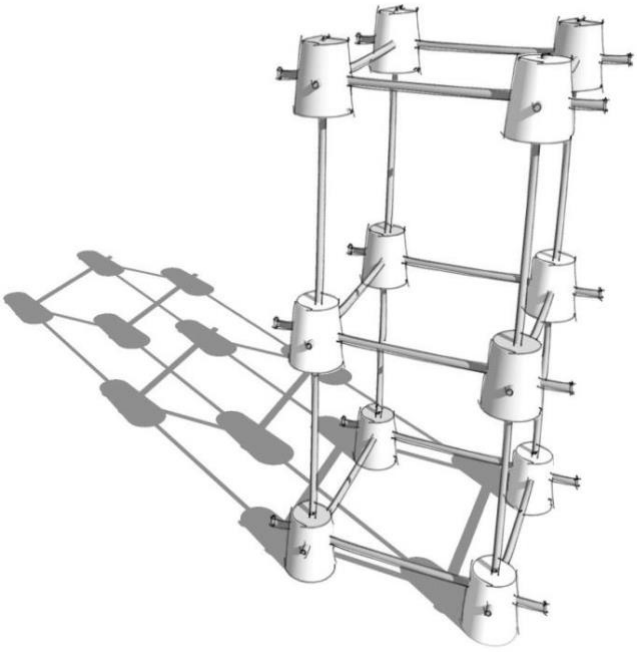
Handout B

In this task, you have to build a clock tower just like the one on the picture 1, out of spaghettis and marshmallows. An idea of how to connect the materials is given in picture 2. The building consists of several different geometric forms that resemble the model on picture 3. Note that there are several ways to construct the clock tower. However, we suggest you to break down the spaghetti to the given length below. Good luck!

Suggested length of spaghetti

Picture 1: Clock tower

Picture 2: Material assembly



Picture 3: Basic layout

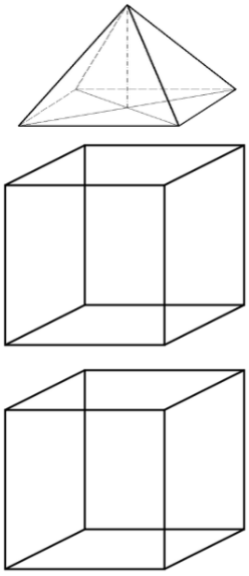


Figure 14: Clock Tower Instructions

Individual judgement

Please write down your age and ID number in numerical form:

Age: _____ ID: _____

Please indicate your gender by circling the answer. I am: Male / Female

Please indicate your native language: _____

1. Task – Origami

a. Please indicate to which extent you expect your **Origamis** to match the model.

Preciseness: _____ (%)

b. Please indicate your estimated time to finish the **Origamis** as given in the preview.

Time: _____ (mins)

2. Task – Clock Tower Construction

a. Please indicate to which extent you expect your **Clock Tower** to match the model.

Preciseness: _____ (%)

b. Please indicate your estimated time to finish the **Clock Tower** as given in the preview.

Time: _____ (mins)

Figure 15: Individual Estimation

Group judgement

Please now answer the same questions as before in groups of two. Before you start please write down the age, gender and ID of each group member in numerical form.

Member1: Age: _____ Male / Female ID: _____

Member2: Age: _____ Male / Female ID: _____

1. Task – Origami

a. Please indicate to which extent you expect your **Origamis** to match the model.

Preciseness: _____ (%)

b. Please indicate your estimated time to finish the **Origamis** as given in the preview.

Time: _____ (mins)

2. Task – Clock Tower Construction

a. Please indicate to which extent you expect your **Clock Tower** to match the model.

Preciseness: _____ (%)

b. Please indicate your estimated time to finish the **Clock Tower** as given in the preview.

Time: _____ (mins)

Figure 16: Group Estimation

Questionnaire

Please write down your age and ID number in numerical form:

Age: _____ ID: _____

Please indicate your gender by circling the answer. I am: Male / Female

Please indicate your native language: _____

Do you think you are talented in spatial visualization tasks (manipulate 2D & 3D objects)?

(not at all) 1 2 3 4 5 (extremely)

Do you have experience with crafting (working with materials)?

(not at all) 1 2 3 4 5 (extremely)

How well did the members in your group work together?

(not at all) 1 2 3 4 5 (extremely)

How much coordination inside the team was required to complete the Origamis?

(not at all) 1 2 3 4 5 (extremely)

How much coordination inside the team was required to complete the Clock Tower?

(not at all) 1 2 3 4 5 (extremely)

How much communication inside the team was required to complete the Origamis?

(not at all) 1 2 3 4 5 (extremely)

How much communication inside the team was required to complete the Clock Tower?

(not at all) 1 2 3 4 5 (extremely)

Please turn the page for the remaining questions**!**

Figure 17: Questionnaire p. 1

How much team work was required to complete the Origamis?

(not at all) 1 2 3 4 5 (extremely)

How much team work was required to complete the Clock Tower?

(not at all) 1 2 3 4 5 (extremely)

How difficult was the Origami task?

(not at all) 1 2 3 4 5 (extremely)

How difficult was the Clock Tower task?

(not at all) 1 2 3 4 5 (extremely)

Figure 18: Questionnaire p. 2

Zusammenfassung

Im Einklang mit den Leitsätzen der Verhaltensökonomie, versucht diese Arbeit Irrationalitäten im menschlichen Verhalten aufzudecken, welche zu nicht optimalen Entscheidungen führen. Das vielgepriesene Abbild der Menschheit, wie etwa dargestellt in dem Auszug aus Hamlet, wird dabei schrittweise aufgebrochen, durch die Analyse von kognitiven Einschränkungen in unseren täglichen Gedankengängen. Plausible und realistische Vorhersagen zu machen, über zukünftige Ereignisse, stellt nicht nur eine unserer größten Schwächen dar, sondern ist auch der Grundstein dieser Studie. Es wird angenommen, dass durch die Diskussion mit anderen, die Formulierung von überoptimistischen Szenarien zusätzlich verstärkt wird. Das Herzstück dieser Abhandlung ist eine experimentelle Studie zum Thema kollektiver Ressourcenfehlplanung unter verschiedenen Aufgabenbedingungen. In zwei aufeinanderfolgenden Gruppenaufgaben wurden Teilnehmer dazu aufgefordert eine individuelle und eine gemeinsame Prognose abzugeben, über die zu erwartende Fertigstellungzeit und Leistung. Die Ergebnisse des Experiments bekräftigen nur teilweise bereits etablierte Erkenntnisse anderer Forschung im Bereich des (Gruppen-) Fehlschlusses. Auch die Existenz von Schätzungsverlagerungen hinzu zu mehr Optimismus, konnte nur zu unzureichender Signifikanz belegt werden. Die dritte Annahme, nämlich dass wechselseitige Abhängigkeit in Gruppenarbeiten zu einer erhöhten Fehleinschätzung aufgrund von Optimismus führt, konnte mit diesem Experiment nicht nachgewiesen werden. Abschließende Diskussionen beschränken sich nicht nur auf theoretische Implikationen, sondern beabsichtigen auch Anregungen für die praktische Welt zu liefern.