MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

"Managing Exploration in Organizations:
The Effect of Managerial Interruptions on Subordinate Search Behaviour"

verfasst von / submitted by

Elisabeth Letea, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirementsfor the degree of

Master of Science (MSc)

Wien, 2023 / Vienna, 2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 915
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Betriebswirtschaft
Betreut von / Supervisor:	Univ.-Prof. Dipl.-Chem. Dr. Markus Georg Reitzig, MBR

Acknowledgments

I would like to express my deepest gratitude to Professor Markus Reitzig and Valentina Richter for their guidance, support, and patience. Sincere thank you to everyone who participated in my experiment, this thesis would have been impossible without you. Special thanks to my partner and my closest family members, for always being by my side.

Abstract

The purpose of this thesis is to examine the effect of managerial appraisal on subordinates' search behaviour compared to non-managerial interruptions during a search task on a complex landscape. It is proposed that interruptions initiated by a superior who provides managerial appraisal would increase the search distance, when compared to non-managerial interruptions, due to the desire of the subordinates to signal a high level of effort to their manager. A laboratory experiment based on the NK landscape was conducted, however, the results are unable to support the hypothesis. Nevertheless, this thesis contributes to the research field by emphasising that managerial practices can constitute an interruption of subordinate's work task. This new perspective can lead to additional insights and explanations when studying a manager's effect on search behaviour.

Table of Content

1 Introduction

Search is a fundamental process in organisations, since in many situations decision-making alternatives are not given at the beginning, but need to be elaborated by search endeavours first (Simon, 1957). Hence in many cases search is necessary for activities like developing new business strategies (Porter, 1991; Rivkin, 2000), creating innovative technologies or new products (Eisenhardt & Tabrizi, 1995; Fleming & Sorenson, 2001; Schillebeeckx, Yimin & Gerard, 2019), solving structural design issues (Ethiraj & Levinthal, 2004; Siggelkow & Rivkin, 2005) and building organisational competences (Bruderer & Singh, 1996). When the search task is complex and choice attributes are interdependent, finding the right balance between exploration and exploitation becomes challenging, which creates the risk of overexploration (Billinger et al., 2014) or underexploration (Piao & Zajac, 2016).

In these situations, however, with increasing complexity the search behaviour of subordinates is progressively challenging to monitor and assess by a superior. Since the search problems are complex and the environment is unknown, a manager is lacking objective benchmarks for comparison purposes. Nevertheless, providing managerial appraisal can still influence search behaviour. As subordinates are aware of the challenges superiors face when trying to assess their subordinates, they tend to pursue activities which display a high level of effort when engaging in a complex task like searching on rugged landscape (Richter et al., 2020). Therefore, managerial appraisal could be a valuable tool to steer subordinates search distance in a desired direction, even though this niche of the search literature is scarcely researched yet.

Considering that management practices such as providing performance appraisal can be an interruption of a work task in certain settings, the exploration enhancing effect could be caused by the managerial appraisal or by the mere interruption of the search process. Hence, the question arises whether a subordinate who is interrupted by a superior has a different search behaviour than those who experience non-managerial interruptions. Therefore, the underlying research question of this thesis is: Can superiors influence the search distance of their subordinates by providing regular managerial appraisal

and how does the effect of this managerial interruption differ from non-managerial interruptions during a complex search task?

In the attempt of finding answers to these questions a laboratory experiment was conducted based on the Alien Game of Billinger et al. (2014) including the main adaptations introduced by Richter et al. (2020). By comparing managerial with non-managerial interruptions, the effect of managerial appraisal can be isolated. It is proposed that the effect of the managerial interruption should outweigh the mere interruption and regular managerial appraisal should result in an increase of search distance.

To provide a proper background for the experiment the theoretical framework is outlined in the first half of this thesis, which consists of the chapters 2 and 3. Since a profound understanding of search in organisations is key, chapter 2 gives an overview of the current state of research in this field. First, search is defined in the organisational context, then the necessity for search is outlined and later explained what makes search complex. After that, the differences between exploration and exploitation are depicted. Next, the concept of capturing search behaviour as search distance is presented, including an explanation of the NK landscape. The last part of this chapter discusses previous findings about the actual search behaviour of individuals on complex landscapes.

The following chapter 3 deals with the manager's effect on search and starts by illustrating the challenges managers face when engaging in assessment of search behaviour. After providing a definition of managerial appraisal, previous studies are reviewed. Starting with findings on how the superior-subordinate relationship impacts subordinate behaviour, followed by a detailed discussion of how managerial appraisal can affect search behaviour. Since management practises constitute an interruption of a work task in certain cases, a closer look at the effect of interruptions on individual's behaviour is provided. Then, non-managerial interruptions (in terms of a mere temporal interruption) are compared to managerial interruptions (in terms of managerial appraisal). Discussing their different potential impact on search distance leads to the hypothesis presented at the end of that chapter.

After that, chapter 4 describes the experimental method. First, the setup and the procedure of the laboratory experiment are explained in detail. Then the differences between the treatment groups and

the control group are depicted. The next part delivers some information about the participants. Lastly, the dependent and independent variables are listed including the explanation of their measurements.

The results of the experiment are presented in chapter 5, which is divided into two parts. First, the descriptive statistics give insights into the observed search behaviour by summarising the payoff outcome and the search distance, then by having a closer look at the search distance per group type over time. The second part consists of testing the hypothesis. The nonparametric Wilcoxon rank-sum test is one of the focal points of the estimation strategy to compare the results per group type. In order to gain a more detailed insights into the effects of the individual variables various Poisson regression models using pooled treatment groups as well as each group individually are presented. This is followed by the discussion of the robustness checks and the chapter is closed by the presentation of the demographic distributions.

In the end, chapter 6 summarises the results of the laboratory experiment and discusses their meaning for the research question. Then the potential limitations of the experimental setting are addressed, including an outlook on future research. Finally, the conclusion highlights the contribution of this thesis to the research field.

2 Search in Organisations

Since the term search is broadly used in a wide variety of topics, it can initially be misleading in the context of organisations. In everyday life the first association is to physically search for an object. However, in organisations it means creating decision alternatives which contribute to problem solving, and later strengthen competitive advantage (Porter, 1991).

The problems organisations have to face are usually nontrivial, since they involve multiple decision variables and are complex due to factors like globalisation, different technologies, interdependencies within business processes as well as with external stakeholders (Mihm, Loch, Wilkinson & Huberman, 2010). Coherently, the complexity increases if the organisation has to operate in an environment, which is turbulent and constantly evolving (Levinthal, 1997). Due to this complexity and environmental setting the decision maker is usually unable to solve the problem by optimizing, but has to choose a satisficing option from a set of alternatives (Simon, 1979). Since these decision-making alternatives are not given at the beginning of the process, they need to be elaborated by search endeavours, which therefore constitute an essential activity in organisations (Simon, 1957; Newell, 1990).

In order to generate said alternatives the decision parameters can either be changed incrementally by altering only few factors (local search) or radically by modifying all parameters (distant search) (Levinthal, 1997). However, possible interdependencies between the decision-making variables need to be taken into consideration. Changing one parameter can influence one or multiple other parameters and thereby the overall outcome leaving the effects within the search task inscrutable. Since an individual's capability to deal with such task complexity is limited, it is necessary to distribute the search efforts within the organisation in order to fully process the required information, capture all decision-making variables and assess the resulting alternatives (Rivkin & Siggelkow, 2003). In order to benefit from economic specialisation, the search endeavours are divided into smaller tasks and allocated to individuals, which on the other hand are in need of managerial coordination (Mihm et al., 2010). Therefore, "such corporate exploration can best be conceived of as the process of aggregating individual search behavior through organizational structure" (Richter, Janjic, Klapper, Keck & Reitzig, 2020, p.2).

Search tasks can vary highly in scope and context depending on the organisational problem, like developing new business strategies (Porter, 1991; Rivkin, 2000), creating innovative technologies or new products (Eisenhardt & Tabrizi, 1995; Fleming & Sorenson, 2001; Schillebeeckx, Yimin & Gerard, 2019), solving structural design issues (Ethiraj & Levinthal, 2004; Siggelkow & Rivkin, 2005) and building organisational competences (Bruderer & Singh, 1996).

The context of the search task determines the decision parameters which need to be (re)combined in order to create new alternatives. For instance, in the case of Porter's (1991) development of business strategies the decision variables include among others product type, pricing policy, degree of vertical or horizontal integration, cost efficiency and geographic location. He explains that by choosing the scope a competitive position within the industry can be gained and, in some cases, might even impact industry structure. "These principles make it clear that the essence of strategy is choice" (Porter, 1991, p.101). Since the alternatives which the decision-maker can choose from need to be elaborated first (Simon, 1957), this quote once more emphasises the necessity of search in organisations for competitive advantage.

2.1 Exploration vs. Exploitation

A fundamental theory is that organisations seek for solutions only if they are unable to meet expectations. Instead of aiming for a global optimum individuals tend to satisfice, such that they stop their search activities as soon as their expectations are met (Simon, 1979). As a result, expectations have a high impact on search behaviour. Since agents have to perform in complex and often novel environments, they cannot always rely on previous knowledge or social comparisons. Therefore, the feedback they receive directly from the search process plays an essential role in the formation of expectations (March, 1988).

One way of conceptualising the process of search is defining it as a two-stage process, where the first step is to assess whether the agent is satisfied with the current situation. Only in case of dissatisfaction and choosing to search, the agent continues the second stage and chooses how many decision-making variables to change (Billinger et al., 2021). Using this two-stage approach, exploration can be defined

in different ways. During the first stage (whether to search) exploration means to strive for new knowledge (March, 1991). However, in the second stage (where to search) exploration is viewed as distant search (Levinthal, 1997).

When examining exploration in terms of the quest for new knowledge many studies have utilized approaches like the multi-armed bandit model for conducting laboratory experiments (for example Posen & Levinthal, 2012). It originates from the one-armed bandit and is therefore an analogy for a slot machine. The participants of such a game are asked to repeatedly choose from a set of alternatives with unknown payoffs and are tasked to maximise their total gains. Since the payoffs are uncertain and stochastic the participants need to get to know the alternatives and their expected payoffs by repeated trial and outcome observation of each lever, thereby pursuing so-called reinforcement learning (Sutton & Barto, 2018).

Each trial's outcome contributes to the decision-maker's expectations and forms subjective beliefs that have a direct impact on the following choices. Each trial represents a trade-off between playing the arm with the currently highest expected payoff (exploitation) or choosing a different alternative which is yet uncertain but might yield higher outcomes (exploration). When put into the context of an organisation, where decision-makers have to select for instance a policy alternative, they can either try a new policy or they decide to remain with the previously implemented one (Posen & Levinthal, 2012).

The organisation's goal is to maximise the payoffs over time, however each choice equals drawing from a probability distribution with unknown and alternative-specific means. Due to a lack of better information each choice depends on subjective beliefs about the expected payoffs of each decision alternative. The organisation would stop searching and exploit the known alternative when the payoffs meet or exceed the aspiration levels. Otherwise, if the alternative currently believed best falls below the target, the organisation would engage in exploration (Cyert & March, 1963; March, 1991).

A general assumption is that the organisation's resources are limited, since the number of feasible search trials is substantially less than the number of alternatives (Rivkin, 2000). Therefore, the organisation has to allocate its resources between exploiting the alternative considered best at the present (maintaining

the status quo), and exploring other options which appear less attractive at first in aspiration to reveal outperforming alternatives (shifting the status quo). At the same time, finding the right balance between exploitation and exploration is crucial. Both processes are essential for company success, since exploitation is connected to refinement, implementation, execution and efficiency, while exploration is associated with risk taking, innovation, variation and flexibility. A structure that overemphasises exploitation and omits exploration is in danger of being trapped in a suboptimal status quo. On the opposite, a system that puts too much effort into exploration while neglecting exploitation is likely to collect too many new ideas without developing them into new products, business strategies or specialised capabilities. In this case, the organisation would "suffer the costs of experimentation without gaining many of its benefits" (March, 1991, p.71).

In the same way, these problems apply to organisational learning as well. Naturally a business has to improve and refine their existing competences, however building too much on the current structures makes experimenting with other procedures less appealing. Vice versa, striving for new skills and ideas, leaves less resources for perfecting the current capabilities (Levinthal & March, 1993). As a result, it is crucial to find the right balance in the trade-off between exploration and exploitation.

2.2 Search Distance

When describing the search strategy by either implementing incremental changes or rather radical ones, search behaviour can be conceptualised using the search distance. In this case, exploration means that the quest for new business strategies exceeds current competences or competitive positions (distant search). Correspondingly, an organisation practices exploitation if the current situation is refined only by marginal adaptations (local search). Such alternations in the close neighbourhood represent more routine activities and imply that the future possibilities are determined by historical effects. When search is exclusively undertaken by marginal improvements all outcomes depend on the initial starting position. Whereas, when the search efforts leave the immediate environment by undertaking more radical change (a so called long-jump), a variety of novel opportunities unfolds. In some cases, such long jumps might be caused by previous failures which demand for organisational reorientation, but then provide a revived

opportunity for local improvements. Local search might be profitable in the short-run, however in the long-run exploration in form of distant search is essential in order to create the possibility for innovation (Levinthal, 1997).

This trade-off between local and distant search becomes increasingly complex when decision-making variables interact. Changing only one parameter can release unforeseen effects and makes predicting future outcomes difficult. Using a rugged performance landscape allows for visualising such interdependencies while at the same time providing a conceptual framework for the search distance. Such a landscape can be constructed using the NK model, which originates from evolutionary biology, but was successfully established in the management literature by Levinthal (1997). By that he stimulated a comprehensive stream of research which has repeatedly proven that the NK model is a suitable instrument for examining search processes as it provides a highly formalised structure that can be used in a variety of decision-making contexts (for a review see Baumann, Schmidt, & Stieglitz, 2019).

The NK algorithm uses two key parameters for creating complex performance landscapes. The parameter N is defined by the number of choice attributes a search alternative consists of. Since each attribute has the option to be chosen or not, these attributes are binary and result in a total of 2^N possible combinations. The payoff of each search alternative is combination specific and determined by drawing from a uniform distribution. When put visually, each point of the search landscape represents one potential combination of choice attributes. The height of that point illustrates the profitability, the higher the peak the higher the payoff (Levinthal, 1997).

The second parameter K is the complexity parameter, as it determines the interdependency of the choice attributes and thereby sets the number of peaks in the search landscape. At its minimum $K = 0$ the variables are completely independent, such that the search involves no complexity. The search landscape then contains only one peak and since the variables are independent the payoff can be improved by changing only one attribute at a time until all attributes are set at the value that results in the highest outcome. The complexity increases with a higher value in K, since more variables become interdependent, which leads to a higher number of peaks in the landscape. The parameter reaches its maximum at $K = N - 1$ when each of the choice attributes depend on all others. In that case one choice

attribute contributes to the payoff depending on its own value as well as depending on the value of all other choice attributes as well. As a change in one attribute affects the payoff contribution of the other attributes, neighbouring search alternatives do not necessarily have a similar performance. Consequently, adapting a single attribute might decrease the payoff, but changing the same attribute together with other ones might have a positive impact on the performance (Levinthal, 1997).

Overall, the property of K regulates the ruggedness of the search landscape, since the interdependencies of the attributes determine the number and shape of the peaks and valleys. This complexity is what makes combinatorial search demanding. The higher the K parameter the more local peaks exist in the search landscape, which become increasingly difficult to differentiate from the global maximum. Since search agents are subject to bounded rationality, a highly rugged landscape incorporates the risk of the search agents being attracted to local peaks instead of keep searching for the alternative with the highest payoff (Baumann et al., 2019).

The challenge for the organisation is then to provide managerial support which incentivises more distant search in order to avoid being trapped on a local search and encourage finding the best performing alternative. Contributing to this managerial challenge Siggelkow and Rivkin (2005) elaborated on formal designs which should support explorative search. Moreover, the appropriate design choice depends on the turbulence of the environment as well. An environment which is complex but stable demands distant search, but an environment which is complex and turbulent additionally requires rapid improvement (Siggelkow & Rivkin, 2005).

2.3 Search Behaviour

With this concept of capturing exploration as a search distance on a search landscape, it is possible to measure the search activities in a continuous variable, in contrast to the bandit model, where it is only binary. This allows for a more comprehensive view on search behaviour. While some papers theorize about the ideal search process (e.g. Siggelkow & Rivkin, 2005), Billinger and colleagues (2014) were interested in studying the actual search behaviour of individuals. They therefore developed an experiment design called the Alien Game, which is a laboratory experiment built on the NK model that

allows for a highly controlled environment. The search task in this experiment was to combine various attributes in order to design a product which then will be sold to an alien customer. The purpose of introducing aliens as customers was to avoid the potential effect of prior knowledge. Since the participants have no knowledge about the preferences and predispositions of the aliens an initial bias of the searcher can be eliminated. Consequently, since the preferences of the alien customers cannot be anticipated, the payoff of the specific combination is unknown as well and has to be uncovered by trial-and-error search.

Since Billinger and colleagues (2014) were mainly interested in the effect of task complexity on search behaviour they varied the complexity of the search landscape by setting the K parameter to 0 for a smooth landscape with no complexity, to 5 for a rugged landscape with intermediate complexity and to 9 for a highly rugged landscape with maximal complexity. For all three complexity types the N parameter was 10 as the participants were given 10 choice variables for the product design. When starting the search process the choice variables were set to display the combination with the lowest payoff such that the starting point was fixed for all participants. Then they had 24 rounds to search for better performing combinations. Since the number of choice attributes is 10, there are 1,024 (2^N) potential combinations. Therefore, the number of points in the landscape by far outweighs the number of trials which constitutes a resource constraint. The participants were given the goal to find the best performing combination and were incentivised to do so by financial remuneration. All participants received a voucher, but the players with the top 3 final performance earned cash prizes.

As expected, the participants performed much better in the smooth than in the rugged landscapes and many participants were able to find the global maximum in the smooth landscape as opposed to the rugged ones. What all landscapes had in common, is that human agents combined local and distant search, in contrast to computational agents which undertook only incremental changes. Both tend to start with a long jump, but the search behaviour of human agents displayed a temporal pattern as the search distance increased over time especially in the rugged landscapes. In general, with increasing number of search trials, the payoffs of the individual combinations increased but the marginal improvements decreased (Billinger et al., 2014).

Considering that a smooth landscape does not contain any local optima, searching locally results in immediate gradual improvements. Hence, local search is beneficial in a low complexity environment. On the contrary, on a rugged landscape local search contains the risk of getting trapped on a local peak. In these settings distant search can be more useful to increase the payoff through long jumps (Levinthal, 1997). The results of the Alien Game show that since human search agents engage in both local and distant search, they are inclined to abort a local search strategy too early in environments with no or intermediate complexity. However, in a highly rugged landscape human participants were able to outperform a computational agent by engaging in more distant search (Billinger et al., 2014).

As outlined before, expectations and subjective beliefs have a high impact on the search process. This is especially relevant in the Alien Game, as there are no objective benchmarks, previous knowledge or comparisons to other searchers (Lant, 1992). Since decision-makers tend to form reference points by using simple heuristics (Baucells, Weber & Welfens, 2011), search agents define reference points based on the outcome of previous trials. Therefore, in this search task reference points are endogenous, subjective and change over time. Billinger and colleagues (2014) defined a feedback variable using the combination with the highest payoff in prior trials to represent the reference value and found strong support for adaptive search (Levinthal & March, 1981).

Adaptive search means that the "search behavior in a combinatorial task gradually adapts to performance feedback, […] a pattern of failure-induced exploration and success-induced exploitation" (Billinger et al., 2014, p.102). A successful combination of choice attributes frequently leads to staying in the neighbourhood of that successful point and undertaking only local search efforts, since positive feedback might be an indication for a promising cluster of alternatives which could be investigated further. Repeated failure on the other hand, increases search distance and thereby leads to more exploratory search behaviour. Whether a current outcome is viewed as success or failure is derived from the subjective reference value, which in this case is the currently highest payoff combination. There are no random shifts between local and distant search as the search behaviour is adapted only gradually. However, the search process tends to display cycles of local and distant search. When a search agent discovers a successful combination during a long jump of "failure-induced exploration", the search

behaviour is likely to return to a strategy with reduced search distance. Additionally, the type of feedback (being positive or negative) affects how quickly adjustments of search behaviour are made. Success results in a rapid downward adjustment, as opposed to a slower upward adjustment in response to failure (Billinger et al., 2014).

Billinger and colleagues (2014) were especially interested in the influence of task complexity and therefore used three differently rugged landscapes (K = 0, 5 and 9). However, they found that the complexity of the search task has no direct effect on search distance, but systematically impacts the feedback which leads to an indirect effect based on the process of adaptive search. On a smooth landscape performance can easily be improved by incremental changes, resulting in more successful combinations and frequent positive feedback, which leads to a local search strategy around the status quo. In a rugged environment on the other hand, it increasingly becomes difficult to identify high performing peaks, which leads to repeated failures that increase the search distance over time (Billinger et al., 2014). These findings are in line with previous work on risk preferences and managerial decision making, which showed that individuals are willing to undertake higher risks when previous efforts were unable to meet a certain aspiration level (March, 1988).

In order to have a more differentiated look on the effect various feedback types have on search behaviour Billinger and colleagues (2021) repeated their Alien Game with a slight modification. They changed the remuneration system to introduce opportunity costs of active search by setting the monetary incentive on the accumulated payoff instead of the payoff of the final trial. They left unchanged that everyone received a voucher to compensate for their participation, but this time the top 3 participants with the highest total payoff earned additional cash prizes. Thereby, the search agents have to consider if the benefits of search outweigh the potential downside risk. This trade-off is especially relevant in later trials when there are only few decisions left, since benefits are uncertain, but the search costs are known.

Billinger and colleagues (2021) differentiated between initial, average and immediate feedback and found that the effect of performance feedback is dependent on time. Yet, their results show that initial and average feedback have no impact on search distance neither in combination with the number of trials nor on their own. However, immediate feedback (comparing the current trial with the previous

one) has a significant effect on search distance, but only in the later trials of the search task. When the search task is rather progressed positive recent feedback leads to a reduced search distance. Billinger and colleagues (2021) suggest that this effect is triggered by the opportunity costs of search. When the number of remaining search trials reduces, the search agents prefer to improve their payoff locally since the risk of long jumps appears higher.

3 Manager's effect on search

Since the described search behaviour is situated in an organisational context and in most cases organisations inherently involve a hierarchy, the search agents are part of a superior-subordinate relationship. As mentioned, such search endeavours on complex landscapes require managerial coordination when divided into subtasks within an organisation (Mihm et al., 2010), but at the same time with increasing complexity they are progressively challenging to monitor and assess by a superior. Furthermore, since search behaviour is not always optimal in complex settings – risk of overexploration (Billinger et al., 2014) or underexploration (Piao & Zajac, 2016) – the question arises how search behaviour of subordinates can be influenced. Does a superior have an impact on search behaviour? If yes, how and in which direction can a manager steer the subordinate's search distance?

3.1 Management Appraisal

There are various ways superiors can engage in performance management in order to improve the performance of their subordinates, the main levers are feedback, goal setting, training and reward systems. All of these methods depend on performance appraisal as a starting point to build on (DeNisi & Murphy, 2017). However, these methods of performance management are hardly applicable to a search task in a complex and unexplored environment. The superiors face the problem of lacking objective benchmarks, which would be necessary for comparisons or goal settings. Additionally, managers have no broader knowledge about the challenges of the search task or landscape than their subordinates, such that they are unable to provide feedback containing any novel information (Richter et al., 2020).

Despite the challenging situation managers can still show their presence and interest by engaging in performance appraisal, which provides an opportunity for assessing the subordinates search endeavours and leads to information exchange (Pichler, 2012; Townley, 1993). "Performance appraisal refers to a formal process, […] by which employees are evaluated by some judge (typically a supervisor) who assesses the employee's performance along a given set of dimensions, assigns a score to that assessment, and then usually informs the employee of his or her formal rating" (DeNisi & Murphy, 2017, p.421).

Furthermore, the benefits of performance appraisal include the possibility of information collection for the managers. By regular assessments the superiors have the opportunity to observe their subordinate's approach to the search task and to gain new insights about the characteristics of the search landscape (Richter et al., 2020).

3.2 Managerial interruptions and the effect on search distance

Since organisational structure and managerial feedback are determinants of subordinates' future behaviour, these factors could provide guidance for search behaviour as well. However, as described above, a complex search task on an unexplored landscape limits a manager's room for manoeuvre. Despite the challenges, engaging in performance appraisal is still viable, which makes this kind of managerial process particularly relevant in the search context. Nevertheless, to the current state this niche is scarcely researched yet (Richter et al., 2020).

However, there are some findings about the superior-subordinate relationship that could be transferrable to the current field of interest in order to understand how a search agent could react to performance appraisal. Various studies have shown that interactions with a manager lead subordinates to modifying their behaviour (Detert & Treviño, 2010; Milliken, Morrison, & Hewlin, 2003; Morrison, See, & Pan, 2015). Reitzig and Maciejovsky (2015) found that agents pass up fewer ideas when operating in a steep hierarchy, especially when concerned about receiving negative feedback from their direct superior. This fear of being assessed and given negative feedback is termed evaluation apprehension and according to these findings it restrains valuable knowledge transfer. However, regardless of their fear of negative response, when confronted with administrative costs and thereby perceived lack of control the subordinates were less interested in submitting risky ideas, since these would demand for profound explanations.

Furthermore, Fang and colleagues (Fang, Kim, & Milliken, 2014) showed that evaluation apprehension and the fear of negative consequences results in subordinates omitting unfavourable feedback, such that during reporting to the superior they might either withhold their true performance level or distort it in a way that is hardly interpretable. The propensity to misrepresent the truth is influenced by social factors

and overall company culture. At the same time, their findings showed that under certain conditions performance can be improved by a moderate rate of information distortion. Unfortunately, that rate is much higher in the real world than would be optimal for a performance enhancing effect.

In addition, Keum and See (2017) took a refined look at the innovation process by applying a multimethod approach with an experimental as well as a field study. They discovered that hierarchy of authority is harmful during the phase of idea generation, due to evaluation apprehension and perceived lack of control, which result in self-censoring, internal filtering or reduced motivation and overall decreased idea generation or sharing. On the other hand, hierarchy of authority can bring certain benefits during the phases of screening and selection, such as decreasing the propensity of agents to select their own ideas as opposed to those of their peers.

Building on these findings in the field of behavioural strategy Richter and colleagues (2020) adapted the Alien Game in order to examine how managerial monitoring would impact subordinate behaviour during a complex task like search on a rugged landscape, in contrast to previous studies which included rather simple tasks. For this purpose, they modified the Alien Game from Billinger and colleagues (2014 & 2021) in following ways: First of all, a virtual boss was introduced in order to create regular managerial monitoring. The managerial interruption happened after every fifth round and consisted of performance appraisal and a request, where the subordinate would have to justify their previous actions. Due to the complexity of the search landscape and the lack of objective benchmarks the managers cannot include novel information to their feedback, such that the appraisal contains solely summarised statistics about the previous search progress. Nevertheless, by engaging in managerial monitoring the superiors signal that they observe the actions of the search agents and thereby might guide their behaviour in a desired direction.

Additionally, the number of search rounds was increased from 25 (Billinger et al., 2014 & 2021) to 40 for two reasons. First, with a larger number of search trials, the agents have "more time to develop a search strategy" (Richter et al., 2020, p.21). Second, the treatment groups received enough managerial interventions, which was eight times in total. The control groups on the other hand searched the performance landscape for 40 rounds without any interruptions.

Another modification was made in the remuneration system of the participants. The task goal was set to maximising the cumulative payoff over all rounds, which the participants actually received in form of a cash payment. Billinger et al. (2021) used a similar approach, however in their experiment only the participants with the top 3 highest payoffs earned the cash prices. Therefore, in the experiment of Richter et al. (2020) the monetary incentive might have been stronger since everyone received a cash payment based solely on their own performance, which could have increased motivation and the mean remuneration payoff as well. Billinger et al. (2014) on the other hand set the goal to finding the best performing combination. In that case, exploration has no opportunity costs. However, in the study of Richter et al. (2020) it was suitable to establish opportunity costs by rewarding the overall payoff in order to recreate a more realistic managerial monitoring situation.

A further adaption is that in the experiment of Richter and colleagues (2020) the participants can decide for themselves where to start searching, such that the decision of the first round represents the individual starting point. In contrast, Billinger et al. (2014) determined a fixed starting position for all agents, which was the configuration with the lowest payoff.

Overall, Richter et al. (2020) conducted four experimental studies and their findings constitute valuable contributions to the development of the neo-Carnegie School (Gavetti, Levinthal, & Ocasio, 2007). In their main study, a virtual manager provides performance appraisal by using a frequency score, that emphasises "how often recent search improved the prior status quo […], thus adequately reflect the way in which managers illustrate past performance of innovation pipelines" (Richter et al., 2020, p.20). This appraisal was either framed positively or negatively by emphasising the number of trials that increased or decreased the combination-specific payoff, depending on the treatment groups the participant was allocated to. Even though the managerial appraisal contained no new information, the framing could indicate the superior's subjective satisfaction or dissatisfaction to the subordinate. The results showed that for both framings managerial monitoring had a significant effect on search behaviour by increasing the search distance.

The underlying mechanisms of this effect are different depending on the framing. In the case of positively framed appraisal the search agent feels motivated and supported, since there is no reason to

fear negative consequences even if exploration efforts might by unsuccessful. On the other hand, negatively framed appraisal can trigger two counteracting mechanisms. First, while engaging in exploratory search a temporary decline of performance is unavoidable but leads to progressively negative response during managerial appraisal. In this case, evaluation apprehension might come into effect, which then can result in a decrease of search distance due to risk adverting behaviour. Second, subordinates who receive punishments as a direct response to an unsuccessful activity tend to promptly change their behaviour that caused the unsuccessful activity in order to prevent further distress and negative consequences. Transferring this to the search context, the subordinate will leave the part of the landscape which caused the disappointing performance results by undertaking long jumps and aiming for higher peaks. This latter effect seems to outweigh the evaluation apprehension, since the results of Richter and colleagues (2020) showed that both positively and negatively framed appraisal lead to an increase in search distance. Therefore, providing managerial appraisal encourages the subordinates to undertake more exploration irrespective of the framing, as opposed to search agents who receive no managerial interaction at all.

However, the managerial monitoring consisted of a combination of performance appraisal as well as a request for justification, such that the described effect could result from one of these elements or from both of them appearing together. In order to isolate each of the managerial interventions Richter and colleagues (2020) conducted a supplementary experiment, where one of the treatment groups received only the managerial appraisal and the other was regularly asked to justify their actions. The pure-justification group had a similar search distance as the control group, which indicates that the process of reflecting and explaining own behaviour has no impact on future search distance. This demonstrates the necessity to assign the request for justification to a clearly defined issue, otherwise "asking the subordinate for an explanation without specifying what exactly this explanation should refer to was not perceived as meaningful" (Richter et al., 2020, p.29). In contrast, providing managerial appraisal without the request for explanation remains increasing the search distance, which underlines a robust independent effect of managerial appraisal.

For the purpose of testing, whether it is the manager who has the crucial impact or if it is appraisal in general regardless of hierarchy, Richter et al. (2020) conducted an additional study in which the performance appraisal was provided by either a colleague on the same hierarchy level or a lower-ranked intern. They found no significant effect during this complementary study. Hence the results highlight that it is necessary for performance appraisal to be provided by a manager to have an effect on subordinate search behaviour. Thus, they concluded that for increasing the agents' search distance by providing positively or negatively framed appraisal the asymmetry in hierarchy, which is intrinsic to the superior-subordinate relationship, is particularly essential (Richter et al., 2020).

As discussed previously, the higher the complexity of the environment the increasingly challenging it is for the managers to compare the performance of their subordinates due to the lack of objective benchmarks. Yet, a subordinate has to expect that the superior will assess their performance (Detert & Treviño, 2010; Milliken, Morrison, & Hewlin, 2003) and Richter and colleagues (2020) argue that the subordinates are aware of these challenges their superiors have to face. Thus, managerial monitoring has a different effect in complex tasks than in other cases. In less complex tasks subordinates, who perceive evaluation apprehension or lack of control due to managerial monitoring, usually engage in risk adverting behaviour in order to prevent negative outcomes which would be visible to their superiors. However, when performing tasks with a higher complexity this effect is outweighed by the desire to impress their manager, such that the subordinates pursue activities which display a high level of effort and can be justified easier such as making long jumps. To demonstrate the impact of task complexity Richter et al. (2020) conducted another experiment with a noncomplex search landscape (K=0) and found no significant impact of managerial monitoring on search behaviour. Therefore, this additional study again highlights that an increase in complexity leads to higher levels of exploration and that complexity is a crucial boundary condition, since only if pursuing a complex search task managerial appraisal will result in a higher search distance.

Taken together, Richter et al. (2020) provided valuable and differentiated insights into a manager's effect on search behaviour by showing that presenting managerial appraisal to subordinates increases their level of exploration in complex search tasks. At the same time, since the manager interrupts the

search task regularly the managerial appraisal constitutes a work intrusion, defined as "an unexpected encounter initiated by another person that interrupts the flow and continuity of an individual's work and brings that work to a temporary halt" (Jett & George, 2003, p. 495). Thus, viewing managerial appraisal as an interruption can bring additional insights and explanations for a profound understanding of how managerial appraisal impacts search behaviour. For this reason, the next chapter will outline the influence of interruptions on subordinates' behaviour with a focus on non-managerial interruptions, in order to later compare the mere interruption with the managerial interruption (in the form of managerial appraisal) and discuss their different effects on search behaviour.

3.3 Non-managerial interruptions and the effect on search distance

Interruptions have a substantial impact on organisational life as they appear very frequently and are omnipresent in every kind of workplace. Rarely employees[1] could finish a work task without being interrupted, as on average three external interruptions occur per hour and up to 22 per day (Wajcman & Rose, 2011). Additionally, interruptions can be time-consuming as well, since survey results from Leroy & Glomb (2018) show that every forth interruption takes more than 20 minutes.

Since interruptions are that common in the daily work, this topic has been studied in connection with various aspects of work life. Studies have found that interruptions have a significant impact on decision making (Speier, Valacich & Vessey, 1999), task performance (Leroy & Glomb, 2018), stress and job satisfaction (Puranik, Koopman & Vough, 2021), time pressure and task accomplishment (Sonnentag, Reinecke, Mata & Vorderer, 2018). However, these studies showed that in most cases interruptions can have beneficial as well as detrimental effects at the same time. For that reason, it is ambiguous how exactly interruptions would impact a complex search task on a rugged landscape.

In general, interruptions are occurrences which prevent members of an organisation from continuing with their work task and bring the previous workflow to a temporary break off. Jett and George (2003) defined four categories: intrusions, breaks, distractions and discrepancies.

[1] In this chapter the term employees is purposely used instead of the term subordinates since the following findings are not specific to the superior-subordinate relationship and are meant to discuss non-managerial interventions.

Intrusions are initiated externally by another person, which can be a manager or a colleague, and happen without being scheduled, such as a visit in person and receiving messages by email or phone call (Jett & George, 2003). This interruption type usually has a negative connotation, since time is a scarce resource, intrusions can quickly instill a perception of time pressure and stress (Puranik et al., 2021; Sonnentag et al., 2018). On the other hand, interruptions can incorporate some benefits as well. Sonnentag et al. (2018) showed that when the intrusion comes in a form of a smaller task, such as answering a message immediately, the level of responsiveness increases which leads to a positive impact on perceived task accomplishment. At the same time, since the interruption constitutes a social interaction with the interrupter, intrusions can have positive effects on social factors as well. Puranik and colleagues (2021) found that intrusions contribute to a feeling of belongingness, which positively impacts self-esteem and job satisfaction. Therefore, the effect of the intrusion can further depend on the primary reason. Being interrupted for societal reasons, such as having a short chat, can differ from being interrupted for work related reasons, such as assigning new tasks or asking for help or information (Puranik et al., 2021). Individual's attention regulation is especially challenged when the intrusion consists of a task, that needs to be carried out immediately. In that case, the interrupted task and the interrupting task both require resources and when individuals struggle with engaging with the new task instead of the previous one, this will likely result in attention residue. In that case, the individual continues to think about the interrupted task, instead of focusing on the interrupting, currently more urgent task. Attention residue negatively impacts performance, as it reduces cognitive availability. This negative effect will be reinforced when individuals expect to continue the previously interrupted work under time pressure. However, one possibility to mitigate this harmful effect is to reflect on the interrupted task and plan the next steps before shifting the focus to the interrupting task. This ready-to-resume plan minimises attention residue and increases task performance (Leory & Glomb, 2018).

Breaks on the other hand are either set by a fixed schedule or can be self-initiated in a planned or spontaneous way. They serve the purpose of satisfying the employee's needs and allow for a rest period (Jett & George, 2003). Regular switching between work and rest and the change of work pace brings various benefits to the employee. The main purposes are recreation and rejuvenation from demanding

but also from monotonous tasks when employees experience fatigue. Even though, breaks positively contribute to recovery as well as performance, the size of their effect depends on break length. For instance, even microbreaks of just 40 seconds can have substantial recovery benefits, which then lead to an increase in performance (Conlin, Hu & Barber, 2021). However, longer breaks have a significantly higher contribution to immediate recovery than short breaks, but in turn, employees tend to perform better after a short break (Lim & Kwok, 2016). One explanation is that breaks can demand for a start-up period in case the employee feels less involved in the task and needs time to remember the details. Furthermore, as breaks interrupt the workflow, they inherently lead to a loss of available time or might even result in procrastination (Jett & George, 2003). On the other hand, when elaborating complex problems which require creativity, breaks are necessary in order to process information at a subconscious level. During this process of incubation, the subconsciousness attempts a multitude of compositions and if a stable combination of mental elements is successfully created it appears in the consciousness as an insight (Csikszentmihalyi & Sawyer, 1995).

Distractions however are associated with rather negative than positive consequences, yet some findings show that the effect depends on task type and task complexity. Distractions are defined as "psychological reactions triggered by external stimuli or secondary activities that interrupt focused concentration on a primary task" (Jett & George, 2003, p.500). Such that a distraction does not necessarily interrupt but disturbs a work task. Distractions can be especially detrimental when a person depends more on the working memory instead of the long-term memory. Hence, when the employee operates in an unknown field, information stored in the short-term memory can easily be lost due to distractions (Jett & George, 2003). Additionally, the harmfulness of distractions depends on task complexity. For simple tasks decision-making performance can be enhanced, yet for complex tasks disruptions will reduce decision accuracy, but expand decision-making time (Speier et al., 1999).

Discrepancies are another type of interruptions which happen unexpectedly and emerge when observations about the external environment contradict internal expectations. They trigger an emotional reaction and the person's attention is diverted towards the cause of the inconsistency, thereby disrupting the information processing of the previous task. In cases when the interrupted task is time-dependent

and complex, individuals might not have the means to appropriately handle the emotional reaction and examine the situation objectively (Jett & George, 2003). However, discrepancies can be initiated on purpose as well. A mentor can guide the focus of a mentee in a desired direction by actively questioning the aspirations and actions of the mentee. This kind of interruption can result in increased flexibility, which is crucial for task settings with high uncertainty and enables successful changes to more promising strategies (Okhuysen, 2001). Another benefit of discrepancies is breaking with potentially outdated schemas. Building on previous experiences and knowledge individuals tend to develop simplified schemas, which support the processing of novel information. However, when individuals face discrepancies, they are more likely to actively analyse and interpret the latest information, thereby initiating mindfulness and active thinking (Levinthal & Rerup, 2006).

When taking these findings together and applying them to the search context on a rugged landscape, interruptions can have various counteracting effects on search behaviour. On the one hand, the potential benefits include that breaks stimulate recovery, which reduces fatigue and thereby might enhance concentration (Lim & Kwok, 2016). Furthermore, an interruption can cause attention residue, such that the person will keep thinking about the unfinished task even while doing something else (Leory & Glomb, 2018). Especially for demanding tasks like searching on a rugged landscape, interrupting a task can create an incubation effect, during which the unconscious mind can help solving complex problems (Csikszentmihalyi & Sawyer, 1995). In addition, the interruption of a workflow can cause an escape from a routine and thus encourage mindfulness and active thinking (Levinthal & Rerup, 2006). Therefore, interruptions can help the search agent to step back from the process, reflect and take time to creatively elaborate new solutions.

On the other hand, interrupting a focused and productive workflow can bring some detrimental effects to the search process. Searching a landscape which is rugged and unfamiliar without any previous experiences is a demanding and novel task. In such a setting, interrupting the search agent's cognitive processes can result in diversion of attention and engagement, later potentially in the loss of information from the short-term memory (Jett & George, 2003). When returning to a complex search task, previous interruptions can further negatively impact decision-making by reducing the decision accuracy and

increasing decision-making time (Speier et al., 1999). Additionally, resuming with the task after an interruption might require a start-up period in order to accommodate to the landscape again and remember the last trials and details, causing costly lags (Jett & George, 2003).

Overall, it is ambiguous whether the beneficial effects will outweigh the detrimental effects of interruptions in a complex search task, or vice versa. Nevertheless, it is relevant to incorporate the interruptions aspect when studying search behaviour in organisations due to the substantial impact on subordinate behaviour and the high frequency they appear within everyday work. An interruption can be initiated by a manger as well, and as shown by Richter et al. (2020) search behaviour of a subordinate can be influenced by a superior through providing regular managerial appraisal. Considering that the manager thereby interrupts the search task regularly, the exploration enhancing effect could be caused by the managerial appraisal or by the mere interruption of the search process. Hence, the question arises whether a subordinate who is interrupted by a superior has a different search behaviour than those who experience non-managerial interruptions. Therefore, the underlying research question of this thesis is: Can superiors influence the search distance of their subordinates by providing regular managerial appraisal and how does the effect of this managerial interruption differ from non-managerial interruptions during a complex search task?

As discussed, interruptions have various counteracting effects on subordinate behaviour which makes predictions for the search context difficult. However, building on the findings of Richter et al. (2020), it can be argued that it is indeed the managerial appraisal that causes the changes in search behaviour, since they found that a pure justification approach or receiving appraisal from a peer or intern had no effect, even though these settings represent an interruption as well. Therefore, it can be proposed that the effect of the managerial appraisal will exceed potential effects of the mere interruption and will result in an increase of search distance when comparing to interruptions that do not involve any managerial interaction. Such that the hypothesis states:

***Hypothesis**: Subordinates will have a higher search distance when being interrupted by a superior who is providing managerial appraisal as opposed to experiencing non-managerial interruptions.*

4 Experimental Method

In order to test the hypothesis and contribute to answering the research question, the chosen method is conducting a laboratory experiment based on the Alien Game of Billinger et al. (2014) including the adaptations introduced by Richter et al. (2020). The treatment remained the same, yet for the purpose of isolating the effect of managerial appraisal from the potential effect of interruptions, non-managerial interruptions have been included in the search process of the control group. This main modification and further slight differences are outlined in a detailed description below.

Nevertheless, the group structure was not altered. There are three groups, two treatment groups and one control group. The search process of the treatment groups includes regular interruptions initiated by a superior, who provides either positively or negatively framed managerial appraisal. The search process of the control group includes regular interruptions as well, however without any form of managerial presence or appraisal, which are therefore termed non-managerial interruptions.

The experiment of this thesis was coded[2] in the programming software o-Tree (Chen, Schonger & Wickens, 2016). The search task lasts 40 rounds, and the search landscape is based on the NK algorithm. Since the searchers are provided with ten symbols to create artistic illustrations, this determines the number of choice attributes to N = 10. This parameter creates a large landscape with a total of 1.024 different combinations, which requires a relatively high number of attempts from the participants in order to discover successful combinations. Richter et al. (2020) set the parameter K to 2 and 9, such that their participants had to complete the search task twice, both for a low as well as a high complexity landscape. In contrast, in the Alien Game of this thesis the participants complete the search task once with the parameter K = 2, which results in a landscape with lower complexity, where it is easier to identify performance peaks, yet it is demanding enough to fulfil the boundary condition of task complexity.

[2] The technical appendix consisting of the entire code can be accessed online via the following link: https://drive.google.com/drive/folders/1RQyQKHqkoE6Wj9b2LwcWnWZW7XSORiaS?usp=sharing

4.1 Detailed procedure

Regardless of the group type or the treatment all participants had to complete two parts of the experiment. The first part consisted of the introduction and trial rounds, the second one represented the main part including the treatment.

First of all, the welcome page asked the participants to imagine that they work for an alien boss in an alien company. On the second page they were introduced to their task, which was described as creating artistic illustrations for alien customers in order to sell them and earn as much payoff as possible. The instructions explained that there are ten different symbols which can be combined freely and without any restrictions to create said illustrations. This means that the players could select either none, some or all of the ten symbols. In the next page the search mask was presented for the very first time and the participants were asked to familiarise themselves with the checkboxes and buttons. After they created their first illustration their design was displayed visually on a fictional canvas, so that the participants could imagine how such an illustration would look like, since later on only the search mask including the ten symbols had been shown. Furthermore, this page included an explanation about which information could be derived from the search mask, that will be outlined in more detail in the next paragraphs.

After that the instructions continued by indicating that the main part consists of 40 rounds, however before that they would have to complete five trial rounds in order to get accustomed to the task. Again, it was emphasised two more times that the goal is to maximise the total payoff. Following the five trial rounds the alien boss appeared for the very first time, but only to those in the treatment groups. It was disclosed that the alien boss is highly interested in the search process and in the players development in the alien company, for that purpose the alien boss was going to observe the average payoff after every fifth round. In order to reinforce the virtual presence of a boss a picture of a green alien in a black suit was displayed every time the alien boss appeared throughout the experiment. Then a summary table of the search mask and the average payoff of the five trial rounds was shown next to the picture of the alien boss. These two pages have been included in the game only for the treatment groups since these pages were the preparation for the treatments in the main part.

After that the last page of the introductory part appeared to all of the players regardless of their group type, stating the end of the trial. Further, it explained that the payoffs from the trial rounds do not contribute to the final payoff in the main part and that the players cannot make any use of their previous experiences, since the alien customers in the main part are different from the first part and have other preferences. In the background this means that the NK landscape used to calculate the payoffs in the trial rounds was not the same that was used in the main part, such that submitting the exact same combination of symbols would lead to different payoffs comparing the trial with the main part. Richter and colleagues (2020) included some further explanations about the preferences of the alien customers, since in their experiment the participants had to play the alien game twice once on a performance landscape with K = 2 and once with K = 9. Even though this was not the case for the experiment described in this thesis, as the main part was only played once, their description of the customer preferences for the landscape with K = 2 was included as well, in order to ensure a high comparability of the experiments and change as little as possible. Due to the landscape having a relatively low complexity it was outlined, that the aliens tend to have favourite symbols to which they assign higher value. Therefore, including such specific symbols would be a better search strategy than to concentrate on the overall design.

In the main part participants were asked to answer demographic questions about their age and gender. After that there was just a short warning, that their main search task would start on the next page such that the players can get ready. As already mentioned, the main search task consisted of 40 rounds and for each round they were given 40 seconds to choose an individual combination of symbols indicated by a countdown at the top of the page. In case the participant did not select any symbols or did not click on the “Next”-button before the time runs out, an empty combination with no symbols selected would be submitted. After each round the combination-specific payoff was displayed without any countdown such that the players could take their time in between rounds to process the new information and develop a search strategy for the next steps. The provided information as well as the search interface itself is presented in form of a large table as can be seen in Figure 1 on the next page.

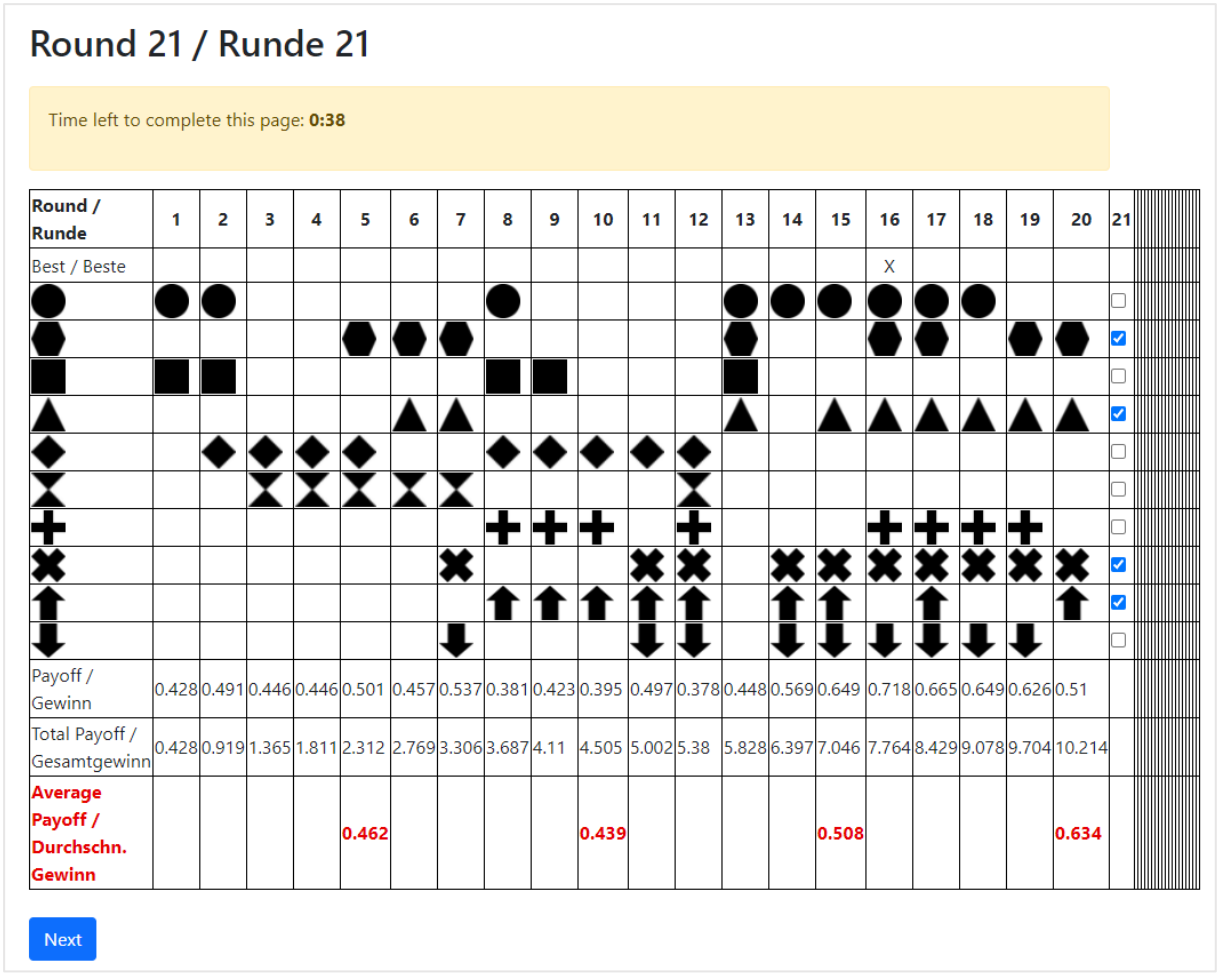


Figure 1: Search interface

This search mask is used throughout the whole alien game for both treatment groups as well as the control group. The title of the page indicates the current round. The yellow box contains the countdown for each round. The columns of the table represent the individual rounds, thereby the table grows as the game progresses. The first row numbers the rounds. The main part contains one symbol per row such that the players can interact with the interface to select the symbols one by one for the artistic illustrations. At the beginning the column of the first round is empty, since there is no starting point provided. Thus, the player has to tick the boxes in order to design their first artwork. However, in all other rounds those symbols which have been used in the previous round are already preselected. The exemplary screenshot in Table 1 shows a game where the first twenty rounds are completed, such that in the beginning of the 21st round the symbols which were used in round 20 are ticked by default, yet can be changed if desired. Since for every round a new column is added to the table the participants can see all of their previously submitted combinations.

The bottom part of the rows contains various indicators about the payoffs. First of all, after each round a combination-specific payoff is derived from the underlying NK landscape. In addition, the total payoff is being displayed which is the cumulative payoff over all rounds. The last row represents the average payoff that is calculated after every fifth round for the last five rounds and functions as a key performance indicator for the treatments. In addition, the best combination with the hitherto maximum individual payoff is marked with an X in the second row (see Round 16 in Figure 1). Due to this structure the table delivers all relevant information about the previous combinations and their respective payoffs. Since there is no need to memorise the previous performance, the player can concentrate on their search strategy without having to rely on their memory skills.

4.2 Treatment Groups

For both treatment groups the fictional alien boss appears after every five rounds in order to reflect on their performance. These managerial interruptions of the search process consist of two pages. First, there is just the announcement that the alien boss enters with the intention to appraise the average payoff of the last five rounds. This page underlines the presence of the imaginary supervisor. The next page involves much more input and therefore constitutes the main part of the treatment (see Appendix Figures A1 and A2). The alien boss summarises the search output of the last five rounds and compares the current average payoff with the previous average payoff.

For that purpose, the left side of the page contains the search table as usual, however it shows just the last five rounds. On the right side the current payoffs are compared to the previous successes by the fictional boss. For players part of the "positively framed appraisal" group the supervisor counts the number of attempts, that had a higher payoff than the average payoff of the previous five rounds. In this way, the emphasis lies on the successful submissions. The following sentence was presented to the players: "Compared to your previous average, your payoffs in this section were above average X out of five times" (see Appendix Figure A1). In contrast, the participants in the "negatively framed appraisal" group were shown the number of combinations, which resulted in a lower payoff than the average payoff of the previous five rounds in order to focus on their failed attempts. The following statement was

displayed: “Compared to your previous average, your payoffs in this section were below average X out of five times” (see Appendix Figure A2). Additionally, the current as well as the previous average payoffs are presented individually. At the end of the page a chart underlines the managerial appraisal graphically. By drawing a red line which represents the previous average and blue dots that show the payoffs of the current five rounds, it is instantly comprehendible what the current performance appears like compared to the previous average.

However, no comparisons are possible during the first managerial interruption after the fifth round since there is no previous average. Therefore, the left side contains the output table as usual, but on the right side of the page only the average payoff of the first five rounds is presented without the above-mentioned treatment statement. Thus, the actual performance appraisal comes into effect after round ten for the first time and seven times in total over the course of the experiment.

The described treatment including the table, chart and statement has been used by Richter et al. (2020) as well. Nevertheless, there are some differences since they applied additional levers to encourage the participants to reflect on their performance. Before the alien boss appeared, the players were asked to rate their own performance compared to the previous section on a five-point scale from “much better” to “much worse”. The second difference is that during the managerial appraisal Richter and colleagues (2020) provided a textbox and instructed their participants to explain and justify their search process. However, in a complementary study they have isolated the appraisal treatment and the justification treatment and found that those who had to explain their search process without receiving any appraisal did not behave any differently from the control group. On the other hand, the search behaviour of those who received managerial appraisal without having to justify their own strategy was similar to those of the main study. Since these findings showed that justification has no impact on search, the alien game for this thesis does not include any demand for explanations.

One last point which should be emphasised about the treatment is that the managerial appraisal does not include any information which would be new to the participant. All figures presented during the treatment can be derived from the search mask and are simply presented in a way that either the successful endeavours or the failed attempts are highlighted within the managerial appraisal. Since the

participants of the treatment groups do not receive any more information than those in the control group, it is possible to examine how the plain framing of reflecting on previous performance can affect the search behaviour.

4.3 Control Group

Even though the control group does not receive any managerial intervention, the participants are still interrupted in their search process after every fifth round. During these interruptions they do not receive any managerial input, the instructions simply state that: "Now you have 10 seconds to reflect. Then the Next-button appears and you can continue the search" (see Appendix Figure A3). The "Next"-button becomes visible only after ten seconds, such that it is not possible to skip the interruption in order to proceed their search task immediately. This way the players are forced to stay at this page for that minimum amount of time, yet if desired they can extend the interruption even longer since the task continues only after clicking the button. During this non-managerial interruption the participants have the chance to reconsider their search process and previous performance or to plan their next steps and develop ideas, however they are free to choose how they spend this time.

These interruptions during the search task of the control group constitute the major modification of the experiment by Richter et al. (2020), in which the participants of the control group had no interruptions at all and completed all 40 rounds in a row. Thus, they introduced a combined effect of the managerial appraisal together with an interruption when comparing the treatment groups with the control group. In order to isolate the effect of the managerial appraisal and eliminate any unintended influences of the interruption, the control group was modified by implementing the non-managerial interruptions in the Alien Game of this thesis.

4.4 Participants

Overall, 60 people participated in the Alien Game and were allocated equally to the three groups. Each group consists of 20 participants with a relatively even age and gender distribution, on which more details will be provided later. In total, the age ranges from 18 to 76 years while 46% of the participants

is male and 53% is female[3]. The participants were selected from a subject pool consisting of family, friends and colleagues of the author of this thesis. This subject pool includes various occupations ranging from students, to being employed or retired.

In contrast to previous studies using the Alien Game (Billinger et al., 2014, Billinger et al., 2021, Richter et al., 2020) there were no monetary incentives provided to the participants. This lack of financial remuneration and its contingent effect shall be discussed later together with the results.

The volunteers participated in the online experiment via individual links using their private electronical devices. The completion of the Alien Game took approximately 30 to 45 minutes, and all of the 60 participants completed the game within two weeks. All instructions and captions were provided in English as well as in German such that the participants could choose a preferred language, thereby reducing possible language barriers.

4.5 Variables

Dependent variable:

Search Distance reflects the search behaviour and is therefore the central dependent variable for this thesis. In alignment with previous studies using the Alien Game (Billinger et al., 2014, Billinger et al., 2021, Richter et al., 2020) it measures the number of search attributes that have been changed in the current trial compared to the highest performing combination so far. This way the variable *Search Distance* captures whether and how far the searchers deviate from the hitherto most successful combination. Therefore, it represents a Hamming Distance, which is widely used for programming literature (Hamming, 1950) as well as for researching organisational search behaviour (e.g. Levinthal, 1997; Rivkin, 2000).

Since the artistic illustrations consist of up to 10 symbols, the *Search Distance* can take any discrete value between 0 and 10, where 0 means that the combination of symbols remained unchanged and 10

[3] The demographic questions also included a non-binary option for gender termed "Other". However, none of the participants selected this answer.

means that all symbols which previously have been selected are now removed from the combination or vice versa. This measure allows for depicting whether the participant engages in rather distant or more local search, and thereby shows the search agent's willingness to deviate from successful combinations.

Independent Variables:

- *Appraisal* is the main independent variable, which takes the value 1 if the search process includes a treatment in form of a positive or negative managerial appraisal, and 0 if the participant was part of the control group. Thereby, this treatment dummy compares the pooled treatment groups with the control group, without distinguishing between positive or negative framing.
- *Positively Framed Appraisal* and *Negatively Framed Appraisal* are additional treatment variables for analysing each treatment group individually. These are dummy variables as well, since a value of 1 specifies the respective treatment group.
- *Payoff Feedback* represents the payoff from the previous round, which was initially derived from the NK algorithm for that specific combination of symbols. The values are continuous, but non-negative.
- *Round Number* indicates the player's progress within the search process. Thus, it can adopt discrete values between 1 and 40.

Demographic Variables:

- *Age* is measured in years and ranges from 18 to 76. For some analyses it will be clustered in five categories (18-29, 30-39, 40-49, 50-59, over 60).
- *Gender* is a binary variable, the value 1 represents female participants and the value 0 represents male participants[4].

[4] As pointed out in footnote 3, none of the participants chose the non-binary answer to the question about gender, therefore the *Gender* variable is binary in this study.

5 Results

First of all, a variety of descriptive statistics have been prepared mainly using the average *Search Distance* over all rounds per person. After that, the nonparametric Wilcoxon rank-sum test is one of the focal points of the estimation strategy to compare the results per group by again using the dependent variable as an average per person. According to the Shapiro-Wilk test ($p = 0.000$) the variable *Search Distance* follows a non-normal distribution, such that the Wilcoxon rank-sum test is an appropriate analysis in this case.

In order to gain a more detailed insight into the effects of the individual variables various Poisson regression models have been run using pooled treatment groups as well es each group individually. This time using participant-round panel data, which allows analyses on the round level instead of an aggregated level per player. In addition, the Poisson regression models have been rerun excluding the first ten rounds, since the first real treatment happens during the second managerial intervention. For the purpose of robustness checks all of these models are tested using a linear regression analysis as well. The results of the ordinary least squares models are included in the Appendix (see Tables A2 and A3).

5.1 Descriptive Statistics

The total payoff off all participants ranges from 17.933 to 25.957 and the mean is 22.006 (all payoff values are in Euro). The final payoff by group type is shown in Table 1, however the mean as well as the minima and maxima are very similar. Interestingly, both the overall minimum as well as the overall maximum are in the Positively Framed Appraisal group.

Final Payoff by Group Type

Final Payoff	Obs	Mean	Std. Dev.	Min	Max
Positively Framed Appraisal	20	22.033	1.967	17.933	25.957
Negatively Framed Appraisal	20	22.059	1.685	19.129	25.182
Control Group	20	21.925	1.515	19.069	25.122

Table 1: Final payoff by Group Type

The average *Search Distance* per person ranges from 0.051 to 3.769 and the mean is 1.403 which indicates that on average seven symbols are changed every five rounds. As can be seen in Table 2, the Negatively Framed Appraisal group has the highest average *Search Distance* while the Positively Framed Appraisal group has the lowest one.

Avg Search Distance by Group Type

Avg Search Distance	Obs	Mean	Std. Dev.	Min	Max
Positively Framed Appraisal	20	1.297	0.783	0.051	2.667
Negatively Framed Appraisal	20	1.546	0.926	0.231	3.538
Control Group	20	1.365	1.024	0.128	3.769

Table 2: Average Search Distance by Group Type

In order to gain a first insight into the patterns of the search behaviour during the 40 rounds, the *Search Distance* per round has been plotted in Figure 2. For enabling comparisons between the group types the *Search Distance* has been averaged per group. At the beginning of the search process the Negatively Framed Appraisal group had a higher *Search Distance* than those with positive framing, yet those from the control group had a relatively high *Search Distance* as well. After the first actual managerial intervention in round 10 the *Search Distance* of the Negatively Framed Appraisal group dropped suddenly, while those who received positively framed appraisal started to search more distantly.

The control group had a peak around round 15, then decreased together with the Positively Framed Appraisal group. Both went back up after round 20, however the *Search Distance* had a decreasing trend from this point, while the Positively Framed Appraisal group had two more peaks after the fourth and fifth managerial interventions in rounds 25 and 30 before their *Search Distance* decreased in the last ten rounds. In contrast, it is salient that in the second half of the Alien Game the participants in the Negatively Framed Appraisal group had a much higher *Search Distance* compared to the other two groups. For this group the *Search Distance* seemed to peak before and after round 25 and after round 30, and then decreased like the other groups. Therefore, especially the fourth and fifth managerial interventions could have had an influence on the search behaviour for both treatment groups, however this is just the impression gained from the chart and not a statistical test yet.

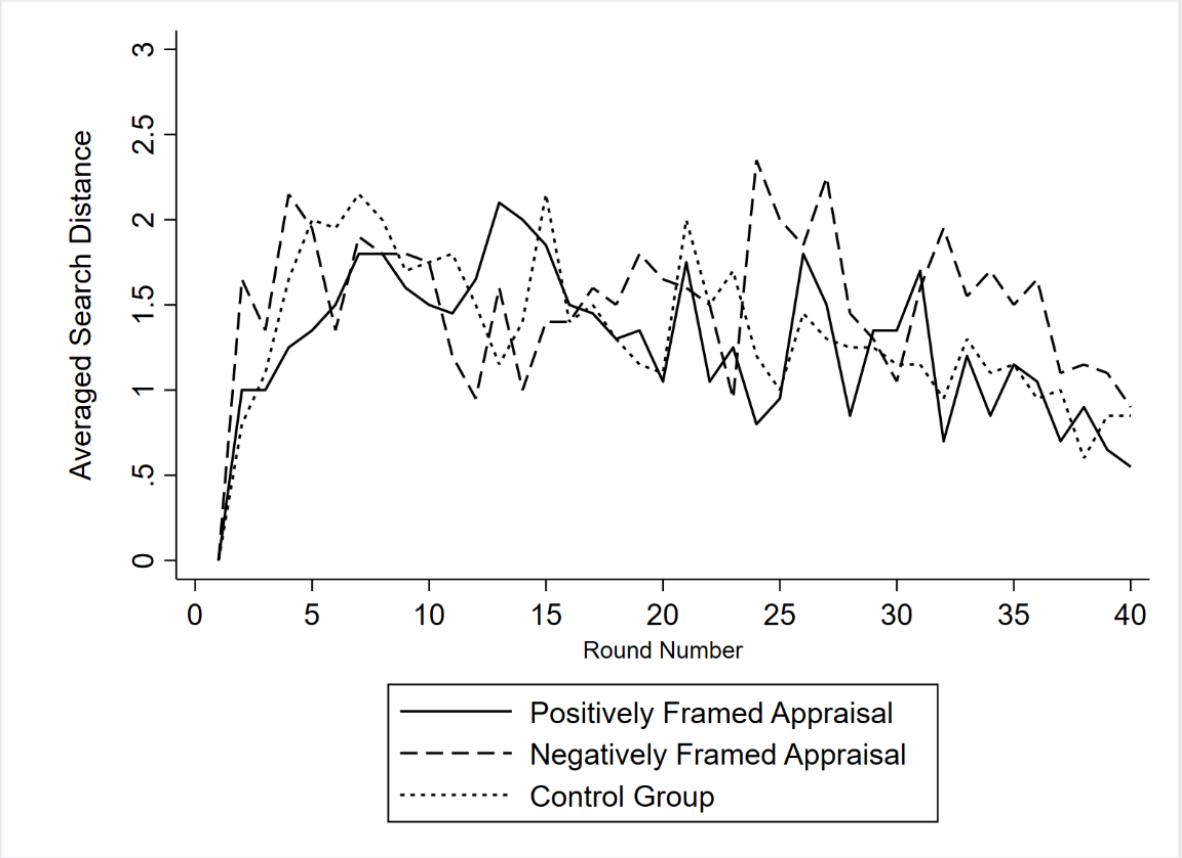


Figure 2: Average Search Distance per group type over time

As anticipated, regardless of the group type there is a slightly decreasing trend towards the end of the game. Yet, this effect is much weaker compared to previous studies (e.g. Richter et al. 2020). However, some other trends are rather unexpected, such as the control group and those with positive framing having quite similar behaviours in some sections as described above, as well as the Negatively Framed Appraisal group having a much higher *Search Distance* than the Positively Framed Appraisal group towards the end of the game. Overall, contradicting to the hypothesis there seem to be no clear patterns that would imply a substantial difference between the treatment groups and the control group, since the *Search Distance* of the control group exceeds the treatment groups several times. This first impression is in alignment with the correlation matrix (see Appendix Table A1), as it shows weak correlations for all of the variables.

5.2 Inferential Statistics

In order to make statistical comparisons between the group types, the average *Search Distance* is used to conduct nonparametric Wilcoxon rank-sum tests, which are an appropriate estimation method since this variable is non-normally distributed according to the Shapiro-Wilk test (p = 0.000). The results of several Wilcoxon rank-sum tests can be found in Table 3.

WRT			
Comparisons	**N**	**z**	**p**
Appraisal vs. Control Group	60	-0.439	0.6606
Positively Framed Appraisal vs. Control Group	40	0.041	0.9676
Negatively Framed Appraisal vs. Control Group	40	0.717	0.4734
Positively vs. Negatively Framed Appraisal	40	-0.703	0.4818

Table 3: Wilcoxon rank-sum tests

Despite the mean of the *Search Distance* being higher for the pooled treatment groups (mean *Search Distance* = 1.422) than for the control group (mean *Search Distance* = 1.365), this difference is not statistically significant according to the Wilcoxon rank-sum test (WRT: p = 0.6606). The results remain insignificant when using only the Positively Framed Appraisal group (mean *Search Distance* = 1.297, WRT: p = 0.9676) or the Negatively Framed Appraisal group (mean *Search Distance* = 1.546, WRT: p = 0.4734) to compare with the control group. Further, when comparing the treatment groups with each other the difference is again not significant (WRT: p = 0.4818). Therefore, it is not possible to reject the null hypothesis at this point. According to these Wilcoxon rank-sum tests, there is no effect of providing managerial appraisal on the *Search Distance* regardless of the positive or negative framing.

As a next step, the treatment variables are combined with the other independent variables *Payoff Feedback* and *Round Number* to create several regression models. These regression analyses lead to a more detailed insight into the different effects, and it can be checked whether the results of the Wilcoxon rank-sum test are reflected. Again, the variable *Appraisal* provides an analysis of the pooled treatment groups, in addition some of the regression models include the *Positively Framed Appraisal* and the *Negatively Framed Appraisal* individually. In contrast to the previous tests, participant-round panel data is used for the regression analyses, which allows testing on the round level instead of an aggregated

level per player. Since the dependent variable *Search Distance* is based on count data, Poisson regressions with robust standard errors have been used, the results are listed in Table 4.

Dependent variable: Search Distance	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Appraisal	0.036			
	(0.057)			
	[0.535]			
Positively Framed Appraisal		-0.065		-0.180
		(0.067)		(0.063)
		[0.332]		[0.005]
Negatively Framed Appraisal			0.061	
			(0.033)	
			[0.061]	
Payoff Feedback	-4.068	-4.804	-4.342	-3.237
	(0.294)	(0.363)	(0.377)	(0.336)
	[0.000]	[0.000]	[0.000]	[0.000]
Round Number	0.000	-0.002	0.003	0.000
	(0.002)	(0.003)	(0.003)	(0.003)
	[0.969]	[0.480]	[0.396]	[0.983]
Constant	2.481	2.904	2.573	2.172
	(0.156)	(0.186)	(0.193)	(0.170)
	[0.000]	[0.000]	[0.000]	[0.000]
Log likelihood	-4,177	-2,683	-2,853	-2,785
Pseudo R-squared	0.0501	0.0697	0.0529	0.0374
Number of observations	2,340	1,560	1,560	1,560
Number of total searchers	60	40	40	40
Excluded category		Negatively Framed Appraisal	Positively Framed Appraisal	Control Group

Note: Robust standard errors in parentheses; p-values in brackets.

Table 4: Poisson regression with robust standard errors

As can be seen in Table 4, four different models have been created during the regression analysis. Model 1 includes the variable *Appraisal* such that the effect of providing managerial appraisal in contrast to no appraisal at all is examined by using the data from all 60 participants with a total number of 2,340

observations. The Models 2 to 4 each exclude one group type to isolate the impact of the appraisal framing, hence these models are derived from 1,560 observations from 40 participants.

According to Model 1, regular managerial interventions would have an increasing effect on the *Search Distance*, however the *Appraisal* parameter is not significant ($p = 0.535$). Model 2 on the other hand suggests that providing *Positively Framed Appraisal* would lead to a decrease of the *Search Distance*, but again the treatment variable is not significant ($p = 0.332$).

In alignment with the hypothesis Model 3 depicts a treatment parameter which indicates a positive impact on the dependent variable and is significant on the 10% level ($p = 0.061$) when focusing on the *Negatively Framed Appraisal*. Although, when excluding the control group and comparing the positive with the negative framing in Model 4, the treatment variable is significant on the 1% level ($p = 0.005$), however the parameter is negative in this case, proposing a decreasing effect on the *Search Distance*.

All four models have in common, that the parameter of the *Payoff Feedback* is negative and highly significant ($p = 0.000$). This clearly underlines that the higher the payoff in the previous round, the lower the *Search Distance*, such that the searcher is less likely to deviate from a successful combination.

The *Round Number* on the other hand is insignificant in all models and its parameter is zero or almost zero, indicating that there is no influence of the progress of the game on the search behaviour. This result is rather unexpected, as it is contradicting to previous studies (Billinger et al., 2014, Billinger et al., 2021, Richter et al., 2020).

Overall, the results of this regression analysis show no support for the hypothesis. Even though, some models display a significant treatment effect, the differences seem rather arbitrary. The effect of the treatment variable is insignificant in Models 1 and 2, but (highly) significant in Models 3 and 4. Yet, Models 1 and 3 indicate an increasing impact of the managerial appraisal, but Models 2 and 4 suggest the opposite. Since these contradicting results are not backed by theory, it has to be assumed that the effects are random in this case.

Further regression analyses have been conducted attempting to clarify the indistinctness and contradictions. Given that during the first managerial interruption after round five there is no previous

average to compare the performance with, the actual appraisal treatment comes into effect after round ten for the first time. Hence, the Poisson regression models have been rerun excluding the first ten rounds. As can be seen in Table 5 below, the regression models have been constructed in the same manner as before, except that less observations are used in each model due to focusing on the period after the first treatment.

Dependent variable: Search Distance	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Appraisal	0.073			
	(0.067)			
	[0.279]			
Positively Framed Appraisal		-0.026		-0.178
		(0.078)		(0.074)
		[0.738]		[0.016]
Negatively Framed Appraisal			0.079	
			(0.038)	
			[0.037]	
Payoff Feedback	-5.182	-5.957	-5.477	-4.368
	(0.327)	(0.414)	(0.412)	(0.365)
	[0.000]	[0.000]	[0.000]	[0.000]
Round Number	-0.006	-0.014	0.001	-0.005
	(0.003)	(0.004)	(0.004)	(0.004)
	[0.085]	[0.002]	[0.743]	[0.193]
Constant	3.215	3.814	3.190	2.929
	(0.193)	(0.233)	(0.236)	(0.211)
	[0.000]	[0.000]	[0.000]	[0.000]
Log likelihood	-3,089	-1,967	-2,106	-2,075
Pseudo R-squared	0.0813	0.108	0.0825	0.0658
Number of observations	1,800	1,200	1,200	1,200
Number of total searchers	60	40	40	40
Excluded category		Negatively Framed Appraisal	Positively Framed Appraisal	Control Group

Note: Robust standard errors in parentheses; p-values in brackets.

Table 5: Poisson regression with robust standard errors for rounds > 10

Overall, there is no substantial change in the results of the regression analysis when excluding the first ten rounds. It is still the case, that the *Appraisal* and *Positively Framed Appraisal* are not significant in Model 1 ($p = 0.279$) and Model 2 ($p = 0.738$). Besides that, the treatment variables stayed significant in Model 3 ($p = 0.037$) and Model 4 ($p = 0.016$) both on the 5% significance level, instead of previously on the significance levels 10% and 1%.

The directions of the treatment variables remained unchanged as well. Thus, Models 1 and 3 keep suggesting that *Appraisal*, and especially *Negatively Framed Appraisal*, would have a positive impact on the *Search Distance*. Whereas Models 2 and 4 continue to indicate a decreasing relationship between the treatment and the search behaviour.

Similarly, the variable *Payoff Feedback* was not affected by focusing on the period after the first treatment, as it continues to be highly significant ($p = 0.000$). The *Round Number* on the other hand, is now weakly significant in Model 1 ($p = 0.085$) and strongly significant in Model 2 ($p = 0.002$), whereas previously it was insignificant in all of the models.

Altogether, excluding the first ten rounds in this second regression analysis had almost no impact on the results. Therefore, the differences about direction and significance of the treatment variables remain puzzling, such that the lack of support for the hypothesis remains.

To examine whether the results are robust all of the regression models have been recomputed based on the ordinary least squares method. The robustness checks are displayed in the Appendix, see Tables A2 and A3. The directions of the treatment parameters did not change when using a linear regression method, such that the regression models again show increasing as well as decreasing effects of managerial appraisal on the *Search Distance*. However, the treatment variables are not significant in any of the models of the robustness checks. There were no changes for the *Payoff Feedback*, as it stayed significant in all cases. The *Round Number* is insignificant, except for Model 2 when excluding the first ten rounds (see Appendix Table A3).

Therefore, the only variable that stayed robust without any changes in the significance through all four regression analyses was *Payoff Feedback*. Thus, in alignment with theoretical explanations and to no

surprise it can be stated with quite certainty that the higher the achieved payoff in the previous round the lower the chances for distant search. This once again emphasises that, searchers are unlikely to deviate from successful combinations.

In general, the robustness checks give no reason to change the conclusion about the hypothesis. Since it is not possible to reject the null hypothesis, it has to be assumed that providing managerial appraisal had no impact on the *Search Distance* when comparing with non-managerial interruptions. Yet it remains unclear, why there are some dissimilarities between the regression models, as the Wilcoxon rank-sum tests revealed no differences between the group types regardless of the positive or negative framing. Since the deviations between the *Positively Framed Appraisal* group and the *Negatively Framed Appraisal* group could not be explained by statistical testing, a closer look on the demographic distributions is needed. Overall, 53% of the participants is female and 46% is male (see Figure 3). This gender ratio is reflected in both treatment groups as can be seen in Figures 4 and 5.

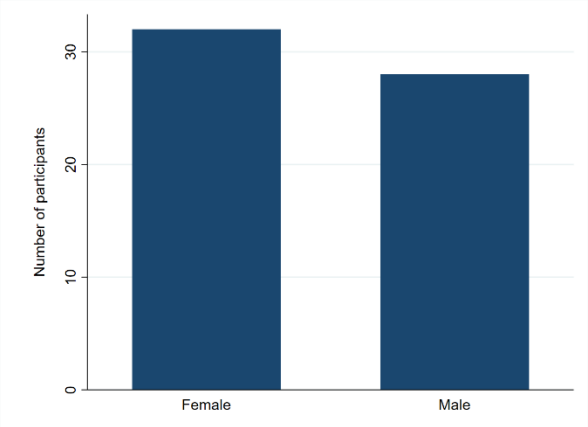


Figure 3: Gender distribution overall

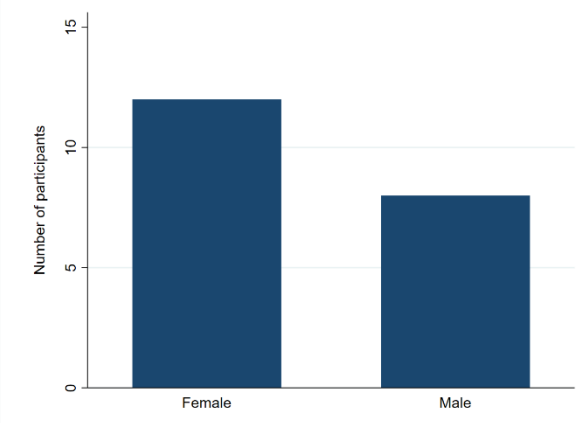


Figure 4: Gender distribution – positive framing

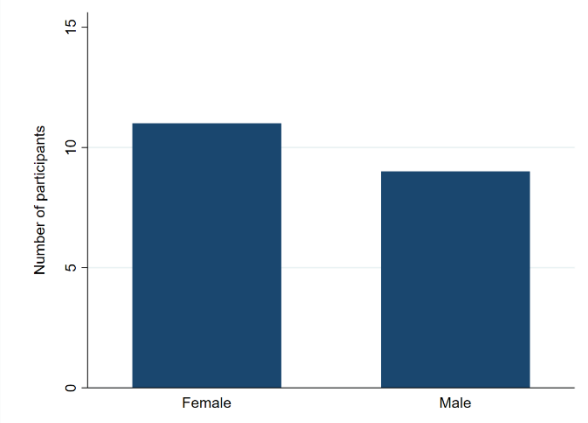


Figure 5: Gender distribution – negative framing

Likewise, it is a similar case for the age distribution. In total, the age ranges from 18 to 76 years, although the majority of the participants (35 out of 60 people) was younger than 30 years old (see Figure 6). Both, the *Positively Framed Appraisal* group (Figure 7) and the *Negatively Framed Appraisal* group (Figure 8) have an age structure which are similar to each other and to the overall distribution.

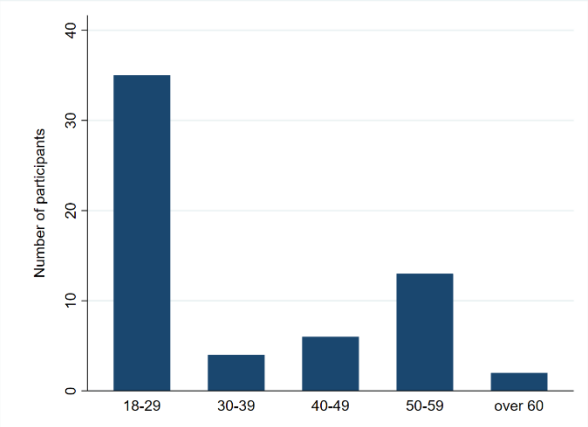


Figure 6: Age distribution overall

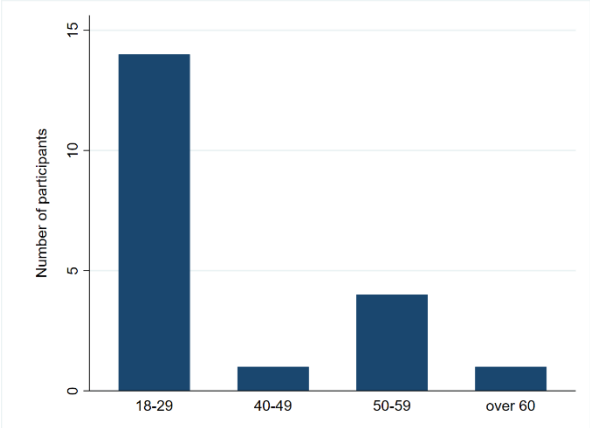


Figure 7: Age distribution – positive framing

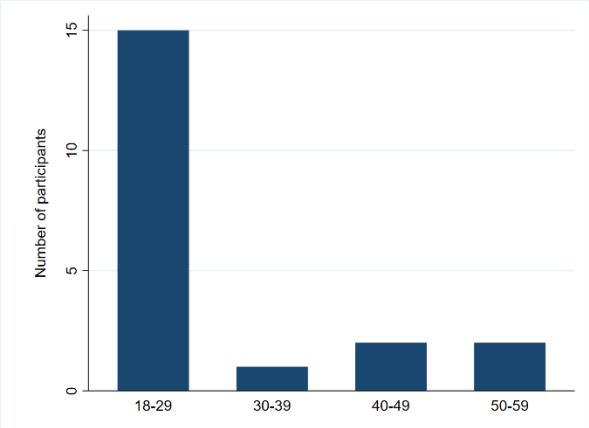


Figure 8: Age distribution – negative framing

Since the demographic distributions of the experiment groups are comparable, both in terms of gender and age, it can be concluded, that the demographic factors cannot explain the differences between positive and negative framing of appraisal found during the Poisson regression analysis.

6 Discussion

The purpose of this thesis was to examine the effect of managerial appraisal on subordinates' search behaviour compared to non-managerial interruptions during a search task on a complex landscape. The chosen method was to conduct an experiment based on the Alien Game of Billinger et al. (2014) including the main adaptations introduced by Richter et al. (2020). In the treatment group participants were regularly interrupted by positively or negatively framed managerial appraisal. The participants of the control group received non-managerial interruptions, without any form of appraisal or managerial presence. Unfortunately, the results of the laboratory experiment at hand showed no support for the hypothesis, which proposed that interruptions that are initiated by a superior who provides managerial appraisal would increase the search distance, when compared to non-managerial interruptions.

Various statistical tests have been applied, yet the results were mainly insignificant and partly contradicting previous findings. According to the nonparametric Wilcoxon rank-sum tests there were no differences in search distance regardless of interruption type or framing of managerial appraisal. Similarly, the Poisson regression model showed no significant effect of managerial appraisal on search distance. Except for negatively framed appraisal, which had a weak exploration increasing effect. Since this effect is contradicting the nonparametric Wilcoxon rank-sum tests, that showed no differences at all, the effect of negative framing in the regression model seems arbitrary.

When comparing the positive to the negative framing the regression analysis suggested a decreasing effect on search distance. However, there is no theoretical background which would suggest a relevant difference between positive and negative framing and previous studies found no effect either (Richter et al., 2020). Furthermore, these differences are not explicable by demographic characteristics, since both treatment groups have similar age and gender distributions. Hence, it has to be assumed the effect displayed in the Poisson regressions is random in this case.

Even though the results in this thesis were unable to show a significant effect of managerial appraisal on search distance, according to the theoretical background a managerial interruption should be more meaningful than a mere temporal interruption. Both interruption types can stimulate reflective and active

thinking, however the non-managerial interruption can be used as a simple break instead of staying focused on the task. The managerial interruption however leads to information exchange, although this information is not new, it still draws the attention to previous successes or failures.

The non-managerial interruptions in general can have various counteracting effects on subordinate behaviour. The potential benefits include reduction of fatigue (Lim & Kwok, 2016), encouragement of mindfulness and active thinking (Levinthal & Rerup, 2006) or idea generation in the unconscious mind for solving complex problems (Csikszentmihalyi & Sawyer, 1995). However, interruptions can negatively impact decision-making (Speier et al., 1999), lead to a loss of attention and engagement and require a costly start-up period (Jett & George, 2003).

However, managerial appraisal should have a particularly strong impact on search. Numerous studies have emphasised that interactions with a manager lead subordinates to modifying their behaviour (e.g.: Fang et al., 2014; Keum & See, 2017). Although for less complex tasks perceived lack of control and the fear of negative consequences leads subordinates to choose less risky actions (Reitzig & Maciejovsky, 2015). In the case of complex situations like searching on a rugged landscape subordinates tend to pursue activities which display a high level of effort, since they are aware of the lack of objective benchmarks and the challenges superiors have to face when trying to assess their subordinates. Thereby, the asymmetry inherent in the superior-subordinate relationship is substantially relevant, since appraisal that is provided by someone on the same hierarchy level or below, such as a colleague or an intern, has no effect on search behaviour. Likewise, a request for justification without appraisal has no impact on search distance. Both of these cases constitute an interruption, yet only the managerial appraisal had a significant effect in the study of Richter et al. (2020). Therefore, it can be argued that the effect of the managerial appraisal should exceed potential effects of the mere interruption. Hence, there must be other causes for the non-significant results in this thesis.

6.1 Limitations and future research

An important reason for the lack of significant results might be the number of participants. The size of the search distance increasing effect of managerial appraisal was originally relatively small in the experiment of Richter et al. (2020). With lower number of participants in this thesis, the effect might have disappeared, such that the sample size may not be adequate to answer the hypothesis. Additionally, the participants do not represent an objective subject pool, since they are volunteers consisting of family, friends and colleagues of the author of this thesis.

A further factor, which could have caused the insignificant effect is the lack of financial remuneration for the participants. Previous Alien Game experiments (Billinger et al., 2014; Billinger et al., 2021; Richter et al. 2020) incorporated different financial incentives to motivate the participants to perform better. Since in this thesis the participants did not receive a compensation for their efforts, they might have not perceived any consequences of the managerial appraisal. Hence, in future research it would be valuable to conduct a replication of this experiment including a monetary incentive.

Like it is the case for most laboratory experiments, the external validity is probably limited due to the simplification and abstraction of the search process. Nevertheless, using a similar approach as previous studies (Billinger et al., 2014; Billinger et al., 2021; Richter et al. 2020) yields important benefits, such as the comparability of the findings across studies.

This thesis focused on analysing search behaviour captured by search distance, however in future research it might be a constructive addition to study the performance of the search outcomes, which is scarcely researched yet. Especially since numerous findings point out that interruptions have a substantial impact on task performance (Leroy & Glomb, 2008; Lim & Kwok, 2018; Conlin, Hu & Barber, 2021), it would be interesting to examine how managerial interruptions can benefit or harm subordinates' performance during a search task with high complexity.

6.2 Conclusion

The connection between managerial appraisal and search behaviour is a phenomenon worth studying since finding the right balance between exploration and exploitation is challenging when searching a rugged landscape. Due to the risk of overexploration (Billinger et al., 2014) or underexploration (Piao & Zajac, 2016), search behaviour of individuals is not always optimal in complex settings. Therefore, it is crucial to further understand how superiors can impact search behaviour. Managerial appraisal could be a valuable tool to steer subordinates search distance in a desired direction.

Even tough, the laboratory experiment was unable to provide significant results to the conceptualisation of the effect of managerial appraisal on subordinates' search behaviour during a complex search task, this thesis constitutes a contribution to this research field by offering a new perspective on managerial appraisal when seen as an interruption. Considering that management practices such as providing performance appraisal sometimes are an interruption of a work task opens up new ways of viewing the effect managers have on subordinate behaviour. Hence, it is a promising future direction to further analyse how the managerial interruptions (especially in the form of performance appraisal) differ from non-managerial interruptions in terms of their impact on subordinate search behaviour.

References

Baucells M., Weber M., Welfens F. (2011). Reference-point formation and updating. *Management Science*. 57(3), 506–519.

Baumann, O., Schmidt, J., & Stieglitz, N. (2019). Effective search in rugged performance landscapes: A review and outlook. *Journal of Management*, 45(1), 285–318.

Billinger, S., Stieglitz, N., & Schumacher, T. R. (2014). Search on rugged landscapes: An experimental study. *Organization Science*, 25(1), 93-108.

Billinger, S., Srikanth, K., Stieglitz, N., & Schumacher, T. R. (2021). Exploration and exploitation in complex search tasks: How feedback influences whether and where human agents search. *Strategic Management Journal*, 24(2), 361-385.

Bruderer, E., & Singh, J. V. (1996). Organizational evolution learning and selection: A genetic-algorithm-based model. *Academy of Management Journal*, 39(5), 1322–1349.

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree - An open-source platform for laboratory, online, and field experiments. *Journal of behavioral and experimental finance*, 9, 88-97.

Conlin, A., Hu, X., & Barber, L. K. (2021). Comparing Relaxation Versus Mastery Microbreak Activity: A Within-Task Recovery Perspective. *Psychological Reports*, 124(1), 248–265.

Csikszentmihalyi, M., & Sawyer, K. (1995). Creative insight: The social dimension of a solitary moment. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight*: 329-363. Cambridge, MA: MIT Press.

Cyert, R. M., & March, J. G. (1963). A behavioral theory of the firm. Englewood Cliffs, NJ: Prentice-Hall.

DeNisi, A. S., & Murphy, K. R. (2017). Performance Appraisal and Performance Management: 100 Years of Progress? *Journal of Applied Psychology*, 102(3), 421– 433.

Detert, J. R., & Treviño, L. K. (2010). Speaking up to higher-ups: How supervisors and skip-level leaders influence employee voice. *Organization Science*, 21(1), 249–270.

Eisenhardt, K. M., & Tabrizi, B. N. (1995). Accelerating Adaptive Processes: Product Innovation in the Global Computer Industry. *Administrative Science Quarterly*, 40(1), 84–110.

Ethiraj, S. K., & Levinthal, D. A. (2004). Modularity and innovation in complex systems. *Management Science*, 50(2), 159–173.

Fang, C., Kim, J. H., & Milliken, F. J. (2014). When bad news is sugarcoated: Information distortion, organizational search and the behavioral theory of the firm. *Strategic Management Journal*, 35(8), 1186–1201.

Fleming, L. (2001). Recombinant uncertainty in technological search. *Management Science*, 47(1), 117-32.

Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: Evidence from patent data. *Research Policy*, 30(7), 1019–1039.

Gavetti, G., Levinthal, D. A., & Ocasio, W. (2007). Perspective-Neo-Carnegie: The Carnegie School's past, present, and reconstructing for the future. *Organization Science*, 18(3), 523–545.

Hamming, R. W. (1950). Error detecting and error correcting codes. *The Bell System Technical Journal*, 29(2), 147-160.

Jett, Q. R., & George, J. M. (2003). Work Interrupted: A Closer Look at the Role of Interruptions in Organizational Life. *The Academy of Management Review*, 28(3), 494–507.

Keum, D. D., & See, K. E. (2017). The influence of hierarchy on idea generation and selection in the innovation process. *Organization Science*, 28(4), 653–669.

Lant, T. K. (1992). Aspiration level adaptation: An empirical exploration. *Management Science*, 38(5), 623–644.

Leroy, S., & Glomb, T. M. (2018). Tasks Interrupted: How Anticipating Time Pressure on Resumption of an Interrupted Task Causes Attention Residue and Low Performance on Interrupting Tasks and How a "Ready-to-Resume" Plan Mitigates the Effects. *Organization Science*, 29(3), 380-397.

Levinthal, D. A., & March, J. G. (1981). A model of adaptive organizational search. *Journal of Economic Behavior & Organization*, 2(4), 307–333.

Levinthal, D. A., & March, J. G. (1993). The myopia of learning. *Strategic Management Journal*, 14(S2), 95–112.

Levinthal, D. A. (1997). Adaptation on rugged landscapes. *Management Science*, 43(7), 934-950.

Levinthal, D. A., & Rerup, C. (2006). Crossing an Apparent Chasm: Bridging Mindful and Less-Mindful Perspectives on Organizational Learning. *Organization Science*, 17(4), 502–513.

Lim, J., & Kwok, K. (2016). The Effects of Varying Break Length on Attention and Time on Task. *Human Factors*, 58(3), 472–481.

March, J.G. (1988). Variable risk preferences and adaptive aspirations. *Journal of Economic Behavior & Organization*, 9(1), 5–24.

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, 2(1), 71-87.

Mihm, J., Loch, C. H., Wilkinson, D., & Huberman, B. A. (2010). Hierarchical structure and search in complex organizations. *Management Science*, 56(5), 831–848.

Milliken, F. J., Morrison, E. W., & Hewlin, P. F. (2003). An exploratory study of employee silence: Issues that employees don't communicate upward and why. *Journal of Management Studies*, 40(6), 1453–1476.

Morrison, See, K. E., & Pan, C. (2015). An Approach-Inhibition Model of Employee Silence: The Joint Effects of Personal Sense of Power and Target Openness. *Personnel Psychology*, 68(3), 547–580.

Newell, A. (1990). Unified theories of cognition. Cambridge: Harvard Univ. Pr.

Okhuysen, G. A. (2001). Structuring change: Familiarity and formal interventions in problem-solving groups. *Academy of Management Journal*, 44, 794-808.

Piao, M., & Zajac, E. J. (2016). How exploitation impedes and impels exploration: Theory and evidence. *Strategic Management Journal*, 37(7), 1431–1447.

Pichler, S. (2012). The social context of performance appraisal and appraisal reactions: A meta- analysis. *Human Resource Management*, 51(5), 709–732.

Porter, M. E. (1991). Towards a dynamic theory of strategy. *Strategic Management Journal*, 12(2), 95–117.

Posen, H. E., & Levinthal, D. A. (2012). Chasing a moving target: Exploitation and exploration in dynamic environments. *Management Science*, 58(3), 587–601.

Puranik, H., Koopman, J., & Vough, H. C. (2021). Excuse me, do you have a minute? An exploration of the dark- and bright-side effects of daily work interruptions for employee well-being. *Journal of Applied Psychology*, 106(12), 1867–1884.

Reitzig, M., & Maciejovsky, B. (2015). Corporate hierarchy and vertical information flow inside the firm - A behavioral view. *Strategic Management Journal*, 36(13), 1979-1999.

Richter, V., Janjic, R., Klapper, H. J., Keck, S., & Reitzig, M. (2020). Authority and Search in Organizations. *Academy of Management Proceedings*, (1), 13448.

Rivkin, J. W., (2000). Imitation of Complex Strategies. *Management Science*, 46(6), 824–844.

Rivkin, J. W., & Siggelkow, N. (2003). Balancing Search and Stability: Interdependencies Among Elements of Organizational Design. *Management Science*, 49(3), 290–311.

Schillebeeckx, S. J. D., Yimin L., & Gerard G. (2019). When do expert teams fail to create impactful inventions? *The Journal of Management Studies*, 56(6), 1073-1104.

Siggelkow, N., & Rivkin, J. W. (2005). Speed and search: Designing organizations for turbulence and complexity. *Organization Science*, 16(2), 101–122.

Simon, H. A. (1957). Models of Man. New York. NY: John Wiley & Sons.

Simon, H. A. (1979). Rational Decision Making in Business Organizations. *The American Economic Review*, 69(4), 493–513.

Speier, C., Valacich, J.S., & Vessey, I. (1999). The influence of task interruption on individual decision making: An information overload perspective. *Decision Sciences*, 30(2), 337–360.

Sonnentag, S., Reinecke, L., Mata, J., & Vorderer, P. (2018). Feeling interrupted—Being responsive: How online messages relate to affect at work. *Journal of Organizational Behavior*, 39, 369–383.

Sutton, R. S., & Barto, A. G. (2018). Introduction to reinforcement learning (Second edition). Cambridge: MIT Press.

Townley, B. (1993). Performance appraisal and the emergence of management. *Journal of Management Studies*, 30(2), 221-238.

Wajcman, J., & Rose, E. (2011). Constant connectivity: Rethinking interruptions at work. *Organization Studies*, 32, 941–961.

Appendix

Appendix A: List of Tables and Figures

Appendix B: Correlation Matrix

Matrix of correlations

	Avg Search Distance	Appraisal	Positively Framed Appraisal	Negatively Framed Appraisal	Female	Age
Avg Search Distance	1					
Appraisal	0.0296	1				
Positively Framed Appraisal	-0.083	0.5	1			
Negatively Framed Appraisal	0.1126	0.5	-0.5	1		
Female	0.0315	0.1181	0.0945	0.0236	1	
Age	0.0503	-0.3385	-0.0682	-0.2703	0.012	1

Table A1: Correlation Matrix

Appendix C: Robustness Checks

Dependent variable: Search Distance	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Appraisal	0.066			
	(0.237)			
	[0.780]			
Positively Framed Appraisal		-0.057		-0.251
		(0.245)		(0.246)
		[0.815]		[0.308]
Negatively Framed Appraisal			0.097	
			(0.137)	
			[0.480]	
Payoff Feedback	-3.672	-4.663	-4.077	-2.491
	(0.827)	(1.080)	(1.105)	(0.842)
	[0.000]	[0.000]	[0.000]	[0.003]
Round Number	-0.005	-0.007	-0.003	-0.006
	(0.008)	(0.010)	(0.011)	(0.009)
	[0.508]	[0.459]	[0.813]	[0.530]
Constant	3.487	4.066	3.649	3.038
	(0.502)	(0.635)	(0.626)	(0.479)
	[0.000]	[0.000]	[0.000]	[0.000]
R-squared between	0.434	0.458	0.389	0.303
R-squared overall	0.076	0.105	0.0798	0.0555
Number of observations	2,340	1,560	1,560	1,560
Number of total searchers	60	40	40	40
Excluded category		Negatively Framed Appraisal	Positively Framed Appraisal	Control Group

Note: Robust standard errors in parentheses; p-values in brackets.

Table A2: Ordinary least squares model with random effects at the round and individual levels

Dependent variable: Search Distance	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Appraisal	0.122			
	(0.281)			
	[0.665]			
Positively Framed Appraisal		0.006		-0.235
		(0.285)		(0.290)
		[0.982]		[0.417]
Negatively Framed Appraisal			0.121	
			(0.165)	
			[0.463]	
Payoff Feedback	-4.456	-5.596	-4.688	-3.323
	(0.879)	(1.207)	(1.146)	(0.856)
	[0.000]	[0.000]	[0.000]	[0.000]
Round Number	-0.011	-0.019	-0.004	-0.011
	(0.008)	(0.011)	(0.009)	(0.010)
	[0.157]	[0.079]	[0.692]	[0.246]
Constant	4.061	4.887	3.994	3.654
	(0.599)	(0.761)	(0.748)	(0.563)
	[0.000]	[0.000]	[0.000]	[0.000]
R-squared between	0.443	0.470	0.400	0.406
R-squared overall	0.122	0.158	0.123	0.0977
Number of observations	1,800	1,200	1,200	1,200
Number of total searchers	60	40	40	40
Excluded category		Negatively Framed Appraisal	Positively Framed Appraisal	Control Group

Note: Robust standard errors in parentheses; p-values in brackets.

Table A3: Ordinary least squares model with random effects at the round and individual levels for rounds > 10

Appendix D: Experimental Instructions

Welcome to the Alien Game!

Imagine you are transported into an alternate universe and you have to work for an **alien boss** in an **alien company.**

You have been selected for a very special task, as you will try to create **artistic illustrations for alien customers**.

You know nothing about your alien customers, let alone their taste in art, however, you will try to earn **as much payoff as possible** by selling your artistic illustrations to them.

In order to do so, you are provided with **10 symbols**, which you can combine **freely** and without any restrictions into an artistic illustration. You can select **NONE, SOME, or ALL symbols** to create an artistic illustration.

You submit your illustration to your alien customers by clicking the **Next-button**. If you **DO NOT klick Next**, the program will not recognize your intended submission. After your submission, your alien customers evaluate your creation and reward you with a **payoff** specific to your particular submission.

Please make sure to click the Next-button before the time runs out. Otherwise your selection will not be submitted.

This is how your workshop looks like:

- Tick the corresponding **boxes** to select a symbol.
- Click the **Next-button** to submit a combination before the time runs out.
- You will see your artwork's **layout** and **payoff** after clicking the Next-button.

This is your output:

- On the bottom left, you see your **selected symbols, your round payoff, your total payoff**, and later you will see your **average payoff of the last 5 rounds.** All payoffs are in Euro.

- On the bottom right, you see how your **artistic illustration** looks like for your **alien customers.**

You have a total of **40 rounds** to create artistic illustrations for your customers. Each round's payoff cumulates to your **total payoff**. You want to earn as much payoff as possible.

But first you have **5 trial rounds** to familiarize yourself with the game.

Remember:

- Combine the symbols however you like.
- Click the Next-button before the time runs out.
- Try to earn as much payoff as possible.

TRIAL ROUNDS 1-5

[for treatment groups only]

So far, so good! But remember your **alien boss?**

Imagine that your alien boss regularly wants to know what you are doing. Your alien boss is very interested in your search process and your development in the **alien company.**

So after every 5th round of alien art creation, your alien boss will observe your **average payoff** of the last 5 rounds.

You are now done with the trial. One last tip:

The aliens in the next part are new customers, therefore you can't rely on your experiences from the trial.

It is relatively straightforward how to discover what your alien customers like or don't like. They tend to have **"favourite" symbols** which are valued higher than others and constructing artwork around such symbols is a good idea. Clients don't care that much about "overall" looks of your work but prefer to have certain symbols in the piece of art they purchase.

The payoffs of the trial are not part of the overall payoff of the game!

This is all you need to know! Now have fun playing the **Alien Game**!

(Note: This customer-specific information is meant to inform participants about the respective complexity of the combinatorial search task which is determined by the underlying NK-landscape. This type of customer "favours" single design features over facetted combinations in an artistic illustration, thus indicating a low level of interdependence between choice variables which is characterised by an NK-landscape where K=2.)

START OF MAIN GAME

Appendix E: Schematic screenshots

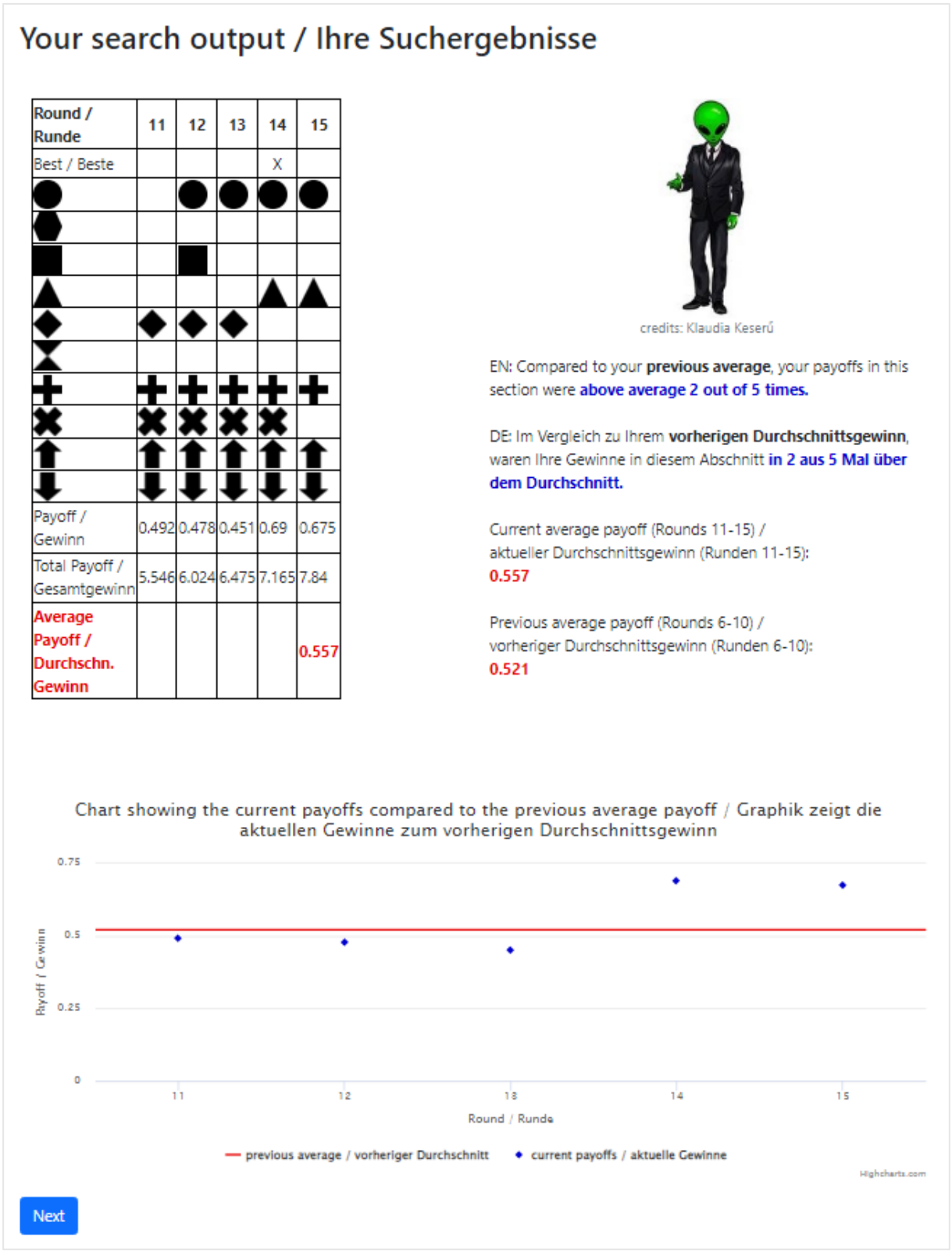


Figure A1: Managerial intervention: positively framed appraisal

Your search output / Ihre Suchergebnisse

Round / Runde	11	12	13	14	15
Best / Beste				X	
●		●	●	●	●
⬢					
■		■			
▲				▲	▲
◆	◆	◆	◆		
⧗					
✚	✚	✚	✚	✚	✚
✖	✖	✖	✖	✖	
⬆	⬆	⬆	⬆	⬆	⬆
⬇	⬇	⬇	⬇	⬇	⬇
Payoff / Gewinn	0.492	0.478	0.451	0.69	0.675
Total Payoff / Gesamtgewinn	5.064	5.542	5.993	6.683	7.358
Average Payoff / Durchschn. Gewinn					**0.557**

credits: Klaudia Keserű

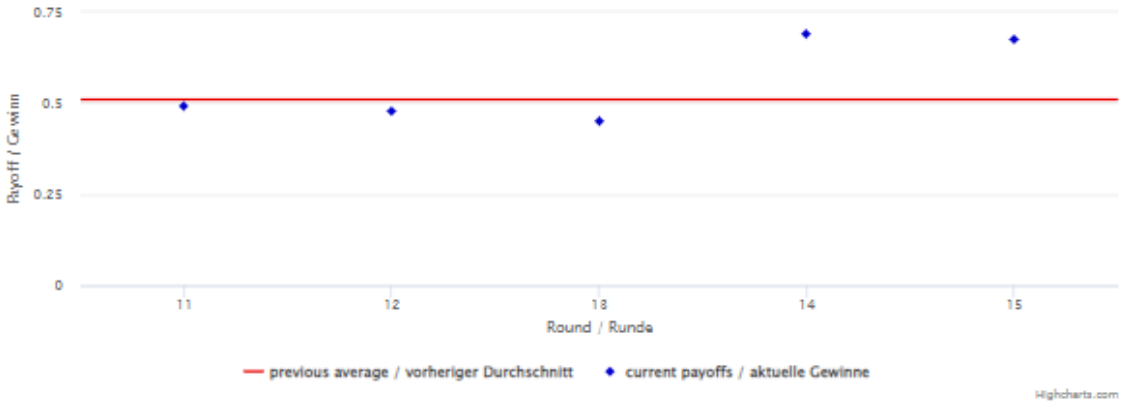
EN: Compared to your **previous average**, your payoffs in this section were **below average 3 out of 5 times.**

DE: Im Vergleich zu Ihrem **vorherigen Durchschnittsgewinn**, waren Ihre Gewinne in diesem Abschnitt **in 3 aus 5 Mal unter dem Durchschnitt.**

Current average payoff (Rounds 11-15) / aktueller Durchschnittsgewinn (Runden 11-15):
0.557

Previous average payoff (Rounds 6-10) / vorheriger Durchschnittsgewinn (Runden 6-10):
0.51

Chart showing the current payoffs compared to the previous average payoff / Graphik zeigt die aktuellen Gewinne zum vorherigen Durchschnittsgewinn



Next

Figure A2: Managerial intervention: negatively framed appraisal

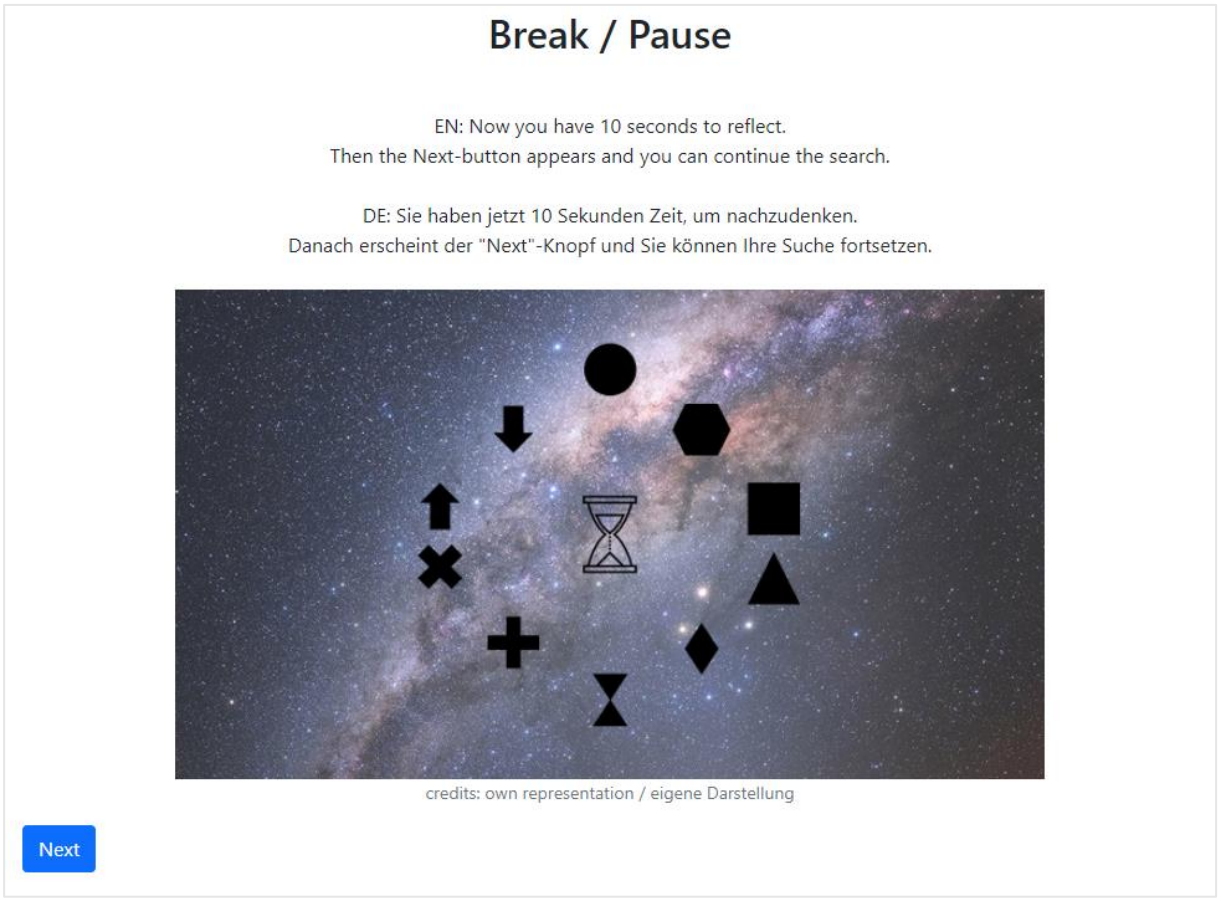


Figure A3: Control Group's regular interruption (no Managerial intervention)

Abstract

The purpose of this thesis is to examine the effect of managerial appraisal on subordinates' search behaviour compared to non-managerial interruptions during a search task on a complex landscape. It is proposed that interruptions initiated by a superior who provides managerial appraisal would increase the search distance, when compared to non-managerial interruptions, due to the desire of the subordinates to signal a high level of effort to their manager. A laboratory experiment based on the NK landscape was conducted, however, the results are unable to support the hypothesis. Nevertheless, this thesis contributes to the research field by emphasising that managerial practices can constitute an interruption of subordinate's work task. This new perspective can lead to additional insights and explanations when studying a manager's effect on search behaviour.

Zusammenfassung

Das Ziel dieser Arbeit ist es, den Einfluss von Managementbewertungen auf das Suchverhalten von MitarbeiterInnen, die eine komplexe Aufgabe auf einer Suchlandschaft ausüben, zu untersuchen. Es wird argumentiert, dass MitarbeiterInnen einen hohen Einsatz signalisieren möchten und dadurch Unterbrechungen, die von Vorgesetzten durch Leistungsbeurteilung ausgelöst werden, die Suchdistanz erhöhen würden, im Vergleich zu Unterbrechungen, die keine Managementpräsenz- oder Beurteilung beinhalten. Es wurde ein auf den NK-Algorithmus aufgebautes Experiment durchgeführt, allerdings konnten die Ergebnisse die Hypothese nicht unterstützen. Dennoch leistet diese Arbeit einen Beitrag zum Forschungsfeld indem aufgezeigt wird, dass Managementpraktiken auch eine Unterbrechung einer Arbeitsaufgabe der MitarbeiterInnen darstellen können. Diese neue Perspektive kann zu zusätzlichen Erkenntnissen und Erklärungen bei der Erforschung des Einflusses, den ManagerInnen auf das Suchverhalten haben, führen.