MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„ Search in turbulent environment “

'The effect of a dynamic environment on individual search behavior'

verfasst von / submitted by

Mario-Patrick Drazil, BSc

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

Master of Science (MSc)

Wien, 2021 / Vienna 2021

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
Mitbetreut von / Co-Supervisor:	

Acknowledgements

At this point I would like to thank all those who have contributed to the success of this master thesis through their professional and personal support. In particular, my supervisor and Valentina Richter for the content assistance. My special thanks go to my family, especially my parents, who made my studies possible and supported me in all of my decisions.

I would also like to thank my working colleague Simona Macovei for the proofreading. Finally, I would like to thank everyone who was there for me while this work was being carried out, especially my girlfriend Anna Wegscheider.

Abstract

The cornerstone of this work lies in the fact that people have to make both conscious and unconscious decisions on a daily basis. Hence the strong urge to examine all aspects that lead to decision-making. Search behavior is one of the main pillars in the decision-making process. This is understood as the relationship between the search for alternative options and the use of existing information. The aim of this thesis is to uncover a possible difference in two different environments in search behavior. On the one hand, there is a greatly simplified environment, the so-called static environment, in which the reward for each decision is always located in the same corridor. On the other hand, there is a more realistic approach to the reality, the so-called dynamic environment, in which the reward of every decision can change completely with a certain probability. It is assumed that the dynamic environment influences search behavior, so that there is an increase in the search for new alternatives. In close connection with an increase in search behavior, there is also a decrease in the conviction of how good a decision was. The heart of this work is a multi-armed bandit experiment for each of the environments. Within it, the participants were asked to choose among five different decisions in each round within a prescribed number of rounds. The aim for the participants was to collect as much money as possible within 100 rounds. It was found that the participants in the dynamic environment were significantly more likely to engage in exploratory behavior and, at the same time, their conviction towards the individual poor was weakened. Final discussions refer to possible reasons why the participants changed their behavior.

Table of Contents

1. Introduction

What am I going to eat today? What should I wear today? These are choices that most of us face every day but it´s not only these everyday decisions that define us as fundamental decision makers. Every action we consciously or unconsciously take is due to a decision. It can be said that many intangibles are involved in a decision-making process. Most of these intangibles are valued by everyone in their personal subconscious. This combined with the multitude of different intangibles makes every human special. The great unknown that all kinds of sciences try to understand is how people evaluate intangibles and how the trade-off between these intangibles looks like. Understanding this would help to predict human decisions (Saaty, 2008).

In order to ensure we choose the best possible alternative, we need to gather all kinds of information related to the decision. Collecting information about something and converting it into knowledge is one of the most powerful intangibles. In making many decisions, people have to repeatedly choose between uncertain alternatives that they can only learn about through experimentation. The theory of search provides a useful account of how good current choices are and if better alternatives exists. The concepts of exploitation and exploration are central to the search process. In the case of the food choice, should someone stick to their current diet which is already well known and therefore exploit the current knowledge or try something new and explore other alternatives. All kinds of different literature suggest ways to find the optimal balance between exploiting current knowledge and exploring new alternatives. An optimal balance is important because if people do not experiment enough, they can lose considerable welfare by missing out on better alternatives. On the other hand, if people experiment too much, they might fail to find better choices which ends in losing welfare because of not choosing the best possible alternative often enough (Anderson, 2000).

In particular, the field of strategic management has an interest in understanding how the trade-off between exploitation and exploration looks like under all kinds of different circumstances. Predicting strategic decisions made by managers and understanding which factors do have an impact on these decisions would be very beneficial to every company. In general, strategic decisions play an important role in the life of a company because they have a major impact on the future of many people. Therefore, it is important to examine different circumstances that influence the exchange between exploitation and exploration. One of these circumstances is the environment in which the decision has to be made. Despite the importance of this topic, most of the existing literature has only examined exploration versus exploitation in a stable, never changing world. A more realistic approach is a world where shocks can have a huge impact on the environment. A shock is an event that can have positive or negative effects on the environment. The central question to address now is how an external shock influences the search process of people. More precisely how does the trade-off between exploration and exploitation look like in a dynamic environment.

In this master thesis I use the between-subjects approach and therefore consider two different types of environment. First, the static environment in which basically everything remains the same at all times and second, the dynamic environment. In this case people are influenced in their search process by external shocks which lead to a change in the underlying payout structure. With the help of a multi-armed bandit I want to provide useful information about the targeted question. A multi-armed bandit gives me the opportunity to replicate the decision-making process because it is a useful modeling structure that captures the trade-off between exploration and exploitation. A bandit consists of multiple arms where each of it has a unique payoff. Generally speaking, each arm represents a different choice which can be picked by the participants in every round. In each period, agents can then choose to stick to their current arm or gather more information about the other arms by selecting them.

The first section of this paper deals with a detailed analysis of the previous literature on the topic. A special focus is on the three main components of this work – Multi-armed bandit, search strategies and environment. Following that, the developed experiment with all components is described in more detail. The last section consists of a detailed discussion of the results, followed by a conclusion including an outlook on future projects.

2. Literature review

Since the beginning of the 20th century studies have provided important information about human decision-making and human behavior. In his book Simon (1957) defines the terms social and rational as a collection of mathematical essays on rational human behavior within a social setting. In the beginning of his academic work Simon (1957) examines the decision-making process from an economic point of view underlaid with the help of mathematical models. In the second section of the book he examines the subject from a psychological point of view. This is just one example of how many different fields namely economics, psychology, mathematics, sociology, and management are involved in the subject of human decision-making and human behavior.

The main focus of this thesis is located in the area of strategic management and therefore the majority of the used literature is based on this field. An already well-known problem in strategic management is the finite Markov decision process. To put it in other words, the Markov decision process is a classical formalization of sequential decision making. The decision maker is called agent and every decision made has direct impact on the immediate and future rewards. This is due to the influence of actions on subsequent situations or states, which are considered in a Markov decision process. Precisely this consideration is the biggest difference to the multi-armed bandit used in this work. That means in a bandit problem only the value of each action is estimated while in a Markov decision process the value of each action in each state is estimated (Sutton & Barto, 2017). These are two well-known examples of models which can be used to frame the problem of decision-making. The preferred model used in this work is that of the multi-armed bandit. Factors found to be influencing human decision-making have been explored in several studies. For my thesis I analyzed the literature for the three main parts that are related to my question of interest. These are namely the multi-armed bandit, search strategies used in bandits and the environment. Going now deeper into all topics related to my work, the logical next step is to analyze the literature about multi-armed bandits.

2.1. Multi-armed bandit

The first related topic constraining the multi-armed bandit, has been addressed from many major perspectives. The initial simple idea of a bandit origins from a slot machine and was first described by Robbins (1952). It is a machine with multiple arms where each arm has his own reward. The focus is only on the selected action and therefore only the selected arm returns a reward. In the next round someone can either pull the same arm again, which means exploiting current knowledge or explore another arm. In general, there are N possible actions. The reward depends only on the present action and is characterized by an unknown distribution which is fixed for each action. The bandit goes over T rounds. In theory a bandit is primarily used as a modelling structure that captures the trade-off between exploration and exploitation. To put it in other words, a bandit is a model which can be used to calculate the optimal balance between exploiting current knowledge and exploring new alternatives. Without a reasonable balance the agent may fail to discover that one alternative has a higher average return than the others or the agent may fail to pull the best arm often enough, if he is not exploiting properly (J. C. Gittins, 1979; Sledge & Príncipe, 2018; Sutton & Barto, 1998). According to Hoelzemann & Klein (2017) the vast majority of the strategic-experimentation literature has its focus on examining the decision between a safe and a risky arm. The safe arm yields a fixed known amount whereas the risky arm can either be better or worse. This can only be deduced by choosing the risky arm. In their experimental setup Hoelzemann & Klein (2017) allowed players to observe each other's actions. This led to a free rider problem where participants are less likely to choose the risky arm than in the single player case. Keller et al. (2005) did a similar experiment and they suggest that in the absence of a breakthrough, the level of experimentation is fully influenced by the initial belief of the agents about the arms. Slivkins (2019) mentioned that in a full feedback bandit where decisions made by individuals can be observed by others, free riding strongly influences the agents towards exploiting. He suggests that agents prefer to consume exploration done by others than exploring themselves. As a result, society as a whole may suffer from insufficient exploration.

Previous research has established that the major differentiating factors for bandits are on one hand the number of arms and on the other hand the underlying reward structure. In his paper Slivkins (2019) mentioned a couple of the most important types of bandits used in the literature. The basic form of a bandit is the stochastic bandit where all the rewards are independent and identically distributed. Following Slivkins (2019) there are three essential assumptions for these kind of bandits. The first one is that the agent can only observe the reward from the selected action. The second assumption is that each action is based on the underlying reward distribution. Every time this action is chosen, the reward is sampled independently from the initially unknown distribution. The last assumption contains a restriction for per-round rewards because they are bounded. What the underlying reward distribution looks like depends on the characteristics which the bandit should reflect. A few examples are the normal distribution, beta distribution or simply a Bernoulli distribution with only 1 or 0 as a reward. Another type of bandit mentioned by Slivkins (2019) is the Bayesian bandit. In this case the Bayesian assumption is added into the stochastic bandit. This means that the problem instance is drawn initially from some known prior distribution. The goal now is to optimize in expectation over this distribution and permanently update the prior distribution given the new data. Examples for distributions which are used for this type are Bernoulli or unit-variance Gaussian. Another type is the Lipschitz bandit, which has information about similarity between arms. This means that similar arms have similar expected rewards. Nevertheless, to be considered as a Lipschitz bandit, they have to fulfill the Lipschitz condition. If the reward is dependent on the action and a context, then the bandit is called a contextual bandit. The definition of a context includes many different things such as features of the environment, major events, the set of feasible actions and additional information about each action. A detailed overview of different bandit problems can be found in Bergemann & Välimäki (2006). Considering the possibility that unplayed arms can evolve further, Whittle (1988) described this case as a restless bandit. Acuna & Schrater (2007) pointed out in their paper that for this type of bandits only approximate solutions are known to exist.

An experiment where students from the University of Amsterdam had to solve two different bandit problems was conducted by Banks et al. (1997). In the first bandit experiment they provided one certain and one uncertain arm to the students. In the second bandit experiment they changed the settings so that both arms were unknown. They were able to show that in both experiments the optimal decision rule depends on the decision maker’s updated beliefs if the uncertain arm is good or bad. Keller et al. (2005) were able to support these findings with their experiment. They suggest that in the absence of a breakthrough, the amount of experimentation depends only on the initial beliefs about the arms. Experiments that are not based on computer simulations or mathematical calculations are difficult to find in the field of multi-armed bandits - a practical example in which real people were used is the Gans et al. (2007) paper. In their experiment subjects had to play three two-armed Bernoulli bandit problems where the success probabilities were different for each problem. The goal was to compare the most common artificial models of calculating the optimal balance between exploration and exploitation with the performance of real humans. They suggest that their findings are positive for researchers despite the fact that the learning process seems to be overvalued in theory. In general, it seems that the tractable models preferred in management research can capture the essential aspects of individuals more than sufficient.

2.2. Search strategies and learning in bandits

Extensive research has shown that learning and searching for alternative decisions are important factors in human behavior. Learning is a term that accompanies us human beings for a lifetime. A considerable amount of literature has been published on the topic of learning. These studies propose many different theories and subdivisions on this subject. A relatively new area of learning that is often associated with the training of artificial intelligence is reinforcement learning. In their book, Sutton & Barto (2017) take up the concept of reinforcement learning by referring to a basic interaction. The mentioned interaction is learning by interacting with our environment. By doing so we collect a lot of information about cause, effect and consequences of our actions as well as what to do in order to achieve goals.

If one compares reinforcement learning with other learning methods mentioned in the literature, a big difference becomes clear. Reinforcement learning means to evaluate the action taken rather than instructing by giving correct actions. This fact provokes the urge to explore because by evaluating only the action taken, it creates some kind of uncertainty as to whether the chosen action was the best. On the other hand for example, instructive feedback which is used in the supervised learning, indicates the correct action to take (Sutton & Barto, 2017). The bandit problem can be very helpful to create an environment in which reinforcement learning can be examined. For my thesis it is important to understand the context of reinforcement learning and a multi-armed bandit. The subjects evaluate their own actions, which is equal to choosing an arm and by doing so they learn after some time which arm yields the highest pay off. March (2003) recognizes that experience can also be a poor teacher. This is due on the one hand to limitations of human cognition and on the other hand to the complexity of history. In addition, some circumstances hinder learning because adaptive processes are subject to systematic error. These circumstances are errors in interpreting history, its own myopia, its tendency to eliminate the variation that it requires and the intricate relations among different levels of a nested and conflictual learning system.

When people make decisions, they often follow a specific strategy. Following Anderson (2000) a search strategy in a multi-armed bandit is a series of history-dependent arm selections. The two terms which are linked with search strategies are exploration and exploitation and studies of search strategies reveal the importance of these two. March (2003) even goes as far as to say that maintaining a balance between exploitation and exploration is a fundamental requirement for intelligent adaption; in other words, maintaining a balance between things that are already known and things that might come to be known. The key aspect in the multi-armed bandit problem is to specify an optimal strategy for dividing resources between exploiting existing knowledge about each arm and exploring other unknown arms to take the chance that these may be the better choice.

Data from several studies suggests that humans are more biased towards exploitation rather than exploration. This means that people do not explore given alternatives enough to the extent that the theory suggests. In his study, Anderson (2000) mentioned two possible explanation given in the search literature in order to be able to explain the suboptimal search bias. On the one hand there is the risk aversion of people and on the other hand the unobservable search cost. In his study, Anderson (2000) took a different approach to explain why people don't search enough. He follows the approach that people and firms underappreciate the value of information in bandit problems. Furthermore, he conducted three laboratory experiments with each focusing on a separate explanation for undervaluation of information in bandits. In the first experiment he controlled for mean-conditional risk and search costs, but people still did not explore enough. This means search costs and risk aversion are not primary factors to explain less search. The second experiment had its focus on testing if agents are hyperbolic discounters who tempt to maximize their current payoff rather than the future payoff. With his third experiment Anderson was then able to find a new explanation for underexperimentation. Summing up all of his experiments, he suggests that ambiguity aversion is a likely cause of suboptimal search in bandits. March (2003) argues that the payoffs for exploiting current knowledge are more certain and attainable compared to exploring new alternatives and Anderson (2000) affirms this by calling such behaviors hyperbolic discounting where agents behave myopically and pick the arm with the current highest expected value. Therefore, adaptive processes are more biased towards exploitation. An imbalance in the trade-off between exploitation and exploration causes two traps for organizations and individuals. First, too much exploration lowers the gain of experience in one field. Second, too much exploitation lowers the possibility to gain new experience and the finding of new better alternatives. In his article, March (2003) also provides a recipe to counteract the bias against searching. So should the stimulation of altruism and self-control help to compensate the adaptive bias. As the exploration increases, a negative point that arises is the increased fare due to exploring being costly. It is now well established from a variety of studies, that in a simple framework where all possible options to choose are identical, the search continues until a reward is greater than some reservation price. Such a reservation price is the hypothetical value where the maximum reward is exactly equal to the expected net gain of choosing one more option (Weitzman, 1979).

All of the above papers focus on explaining search strategies based on psychological factors. Nevertheless, the biggest part by far in the search literature contains explanations based on mathematical models which are primarily used in multi-armed bandits. The aim of these mathematical models is to predict human behavior and in a broader sense, human decision-making. An additional point of these models is that they are also used to solve search problems. This means they can be used as a guideline to allocate given resources to either exploiting or exploring. However, in many decisions there is no opportunity of using such mathematical tools and therefore they are particularly well received in the search theory. Baumann (2015) explained in his article a simple and well-known model which is helpful in modelling organizational structures. The model he mentioned is the NK model, first introduced by Stuart Kauffman. In an organization many decisions are made on an individual level. The question to address is how dependent these decisions from each other are. The K-value is used as an expression to indicate how independent the decisions are from each other. This means that a value of 0 is equal to a fully independence of all decisions. The best strategy for this case would then be to solve each decision individually. As the K-value increases, decisions are more and more influenced by one another, which means that even small decisions affect overall performance. Especially when decision makers face sets of interdependent choices the NK model is proved to be a valuable tool.

Gittins (1979) was the first to find a functional mathematical solution to a simple multi-armed bandit problem. He calculated the Gittins index for each arm and formulated then a simple strategy: always play the arm with the highest Gittins index. Calculating this index is similar to the idea of a reservation price for each arm presented by Weitzman (1979). Over the last few decades, Gittins' original idea was often part of various work in the field of multi-armed bandits. So proved Weber (1992) the optimality of the Gittins index and in his second edition, J. Gittins et al. (2011) further developed this idea. Other well-known strategies for balancing exploration and exploitation are the ε-greedy methods. The basic idea of these strategies is to choose at any point in time the alternative with the highest expected payoff. In ε-greedy methods only a small amount of time is used to randomly pick an alternative in order to ensure exploration. This means that the chance of choosing the worst and best alternative is equal. A more precise method is the Upper Confidence Bond, where the choice is made deterministically by favoring the alternatives that have so far received fewer samples. A special improvement of the ε-greedy methods is the soft-max choice rule. Here is the random choice based on a soft-max distribution (Auer et al., 2002; Sutton & Barto, 1998, 2017).

Previous research has established that Thompson sampling is a proper way to solve multi-armed bandit problems. In other words, Thompson sampling is a heuristic which helps to solve the optimal allocation between exploring and exploiting. This happens through the fact that optimal arms are explored more frequently than arms that might be suboptimal. For example, if an arm has 30% chance of being the best arm, then the chance that this arm will be explored next is also 30%. In the beginning Thompson sampling randomly chose alternatives and after each round of playing it updated his beliefs about each arm. As a result, as the rounds continue, the best arms are played more and more often. This is a great advantage because with an increasing sample size among the best arms, it is easier to distinguish between them to find the absolute best arm. Nevertheless, there is still a random factor in Thompson sampling which guarantees exploring of assumed suboptimal arms (Scott, 2015).

2.3. Environment

When speaking of the environment around the decisions, it encompasses a large number of different parameters that interact in a non-simple way (Simon, 1962). A simple definition of all involved parts in a decision process is provided by Sutton & Barto (1998). They called the decision, maker and learner, agent and everything outside the agent is called the environment. Baumann (2015) painted a picture of complex adaptive systems within firms by saying that organizations often face interdependent choices. Agents in this decision processes are often handicapped by bounded rationality or limited capabilities for searching and making decisions. Therefore, a complex environment includes a multitude of interdependent decisions. This complexity of the system is also an advantage for organizations because it prevents competitors to imitate the value streams. Following Siggelkow & Rivkin (2005) an environment is called turbulent or dynamic if firms assigning of decisions to performance outcomes changes frequently or in ways that are difficult to predict. This also supports the definition of Baumann (2015) by assigning companies whose decisions are independent of one another to a simple environment.

One important parameter about the environment is the underlying reward structure, not specifically how the structure looks like. This would fit more within the topic of the multi-armed bandits. It entails a scenario where the landscape undergoes shocks in periodic intervals. An environment in which the reward distribution stays the same all the time is called a static environment. If the landscape undergoes shocks in periodic intervals, then the environment is called turbulent or dynamic. A shock means that the payout is re-selected for each arm. This replaces the original reward distribution with a completely new one (Siggelkow & Rivkin, 2005). To the best of my knowledge, there are very few studies in the field of strategic management that have examined in more detail the search process in turbulent environments. One of those studies is from Posen & Levinthal (2012), where they conducted three different simulations. Their simulations consisted of a 10-arm bandit with 25000 organizations as participants. They simulated 500 periods for each setting starting with initial beliefs of 0.5. In the first experiment with a stable world they could find support for the findings of March (1991). This means the best performance was at a moderate level of exploration which helped balancing exploration and exploitation. In their second experiment they added neutral turbulences, which means that if a shock happens, they redraw the rewards from the same initial beta distribution, to their environment. The last experiment was then with non-neutral shocks which changes the mean payoff value of each arm. After analyzing all their data, they suggest that at low levels of turbulence the appropriate answer of organizations is to increase exploration. This is due to the possibility of gaining more knowledge with decreasing opportunity costs. One event that is caused by shocks is the decrease in the life span of knowledge which results in a reduction in future value of existing knowledge. Therefore, as the level of turbulence increases the best solution is to shift the balance between exploration and exploitation towards exploitation.

2.4. Research gap

Although there is a great deal of literature on the subject of bandit problems, general solutions for optimal distribution of resources between exploring and exploiting only exists in the simplest of situations. Due to the complexity of this area and the multitude of different options, it is difficult to determine optimal strategies (March, 2003). If one looks at the relationship in the existing literature between mathematical models and their comparison with real people, one recognizes a strong overhang towards the models. In general, there is a lack of real experiments in bandit literature, especially on complex topics. As Baumann (2015) suggested, developing models with agents whose behavior is based on constrained rationality is somewhat more realistic.

The economic search literature has dealt extensively with the situation where the environment remains unchanged all the time. In reality, people and companies face a constantly changing world. It is therefore a far more realistic approach to examine the trade-off between exploring and exploiting in such an environment (Baumann, 2015). As far as I know there is only one paper which examines the trade-off in a dynamic environment, namely the work of Posen & Levinthal (2012). Due to the fact that their results are based on computer simulations, the research gap mentioned above can be guessed. In an experiment with real people there is still much uncertainty about the relationship between exploring and exploiting in a dynamic environment. You can adapt mathematical models to the conditions of a dynamic environment, but the question of how well they perform compared to humans remains open. The bandit problem is one in which a certain uncertainty always plays a role. Nevertheless, it is often possible to learn enough about the static environment to make the best possible decisions. In a dynamic environment, however, there is always the possibility of a sudden shock which lowers the learning effect. In this context, an unanswered question remains still: how people change their trade-off between exploring and exploiting in the dynamic environment.

3. Research design

This thesis intends to determine the extent to which and whether a dynamic environment influences the search process of people. Therefore, the main aim of this study is to investigate the different trade-offs between exploring and exploiting in the static and dynamic environment.

There is a large number of published studies (e.g., Anderson, 2000; March, 2003; Sutton & Barto, 1998; Weitzman, 1979) that describe a human tendency towards exploiting current rather than exploring new knowledge. For this tendency there are some rationally explainable processes that can be taken from the respective studies. Posen & Levinthal (2012) performed a series of computer simulations to show the effect of a dynamic environment on the optimal search strategy. They were able to show that at low levels of shocks the best response is to experiment more. This is due to the positive net benefit of gaining more knowledge rather than losing knowledge because of the lower lifetime. Accordingly, after a while, a static environment leads to a certain saturation of knowledge about the possible decisions. Therefore, there is little incentive for people to deviate from what appears to be the best decision. In a dynamic environment however, the process of saturation is at least pushed back further. This can be explained by the fact that the occurrence of a shock changes the environment. A dynamic environment is therefore synonymous with the ability to generate more knowledge. However, one must not ignore the fact that in a dynamic environment you can never be sure whether everything will be the same in the next round. Therefore, the subsequent working hypothesis is proposed:

Working hypothesis 1: The trade-off between exploitation and exploration is favored towards exploitation, less so in a dynamic environment.

If one follows the view of the current literature on the subject of search, the belief of the people about how good or bad the respective decisions are is playing a major role. In their experimental analysis of the bandit problem, Banks et al. (1997) identify the importance of the decision maker's updated beliefs concerning whether the uncertain arm is good or bad. They suggest that the optimal strategy for the agent depends only on his current beliefs. According to Keller et al. (2005), the number of experiments depends only on the initial beliefs about the arms and the idea of how long it is possible to experiment. As noted by Posen & Levinthal (2012), the strength of the opinion reflects how strongly a company or person believes that an alternative is superior to other alternatives. As belief levels decrease, organizations endogenously increase the extent to which their decisions are exploratory. Therefore, the subsequent working hypothesis is proposed at the individual level:

Working hypothesis 2: The average level of belief is higher in a static than in a dynamic environment.

4. Experimental design

4.1. Overview

96 persons were asked to participate in one of two simple multi-armed bandit games, which means that this study uses the between-subjects approach. The between-subjects method is one of the more practical ways of observing the effect on a dependent variable when changing the independent variable. Therefore, two multi-armed bandit games were created using a Python-based program called oTree (Chen et al., 2016). The two bandits were then uploaded to the Heroku cloud platform. After creating a session on Heroku participants were sent an individual link to be able to play the assigned bandit game. The link was available for every internet-enabled device. Another advantage of using the between-subjects method is that it avoids the problem of knowledge transfer between the two multi-armed bandit games; the reason being, each test participant is assigned to only one of the two bandit games.

On the first page participants were given introductions (Appendix: Figure 20) about how the game works. The introduction for both experiments was almost identical except for the addition that in a dynamic environment the expected value and the variance of the two arms can change with a 50% probability (Appendix: Figure 21). Accordingly, it was already clear to the participants after reading the introduction whether they were in an experiment with a dynamic or static environment. In addition, they were asked about their age and gender. The purpose of the study was not explicitly disclosed but it was pointed out that the goal is to achieve the best possible performance within 100 rounds. Participants were also specifically informed about the question of beliefs that comes up every 5 rounds. This is done to minimize misunderstandings and misinterpretations. On the second page the partakers were asked to choose either Arm A, Arm B, Arm C, Arm D or Arm E (Appendix: Figure 22). After selecting an arm, they were able to move on to the next page where they were asked how strongly they are convinced that they have chosen the best arm with the information already collected (Appendix: Figure 23). In order to prevent the participants from falling into a kind of automatism after just a few rounds, this page only appears every 5 rounds. The fourth page is called the results page. This page is for the purpose of illustrating the options chosen and what the arm paid out (Appendix: Figure 24). The participants then returned to the second page to choose an arm for the next round. Taking into account Miller's (1956) findings that humans can only remember about 7 things, a round history table is provided in the lower part of the second page. The running payoff is located above the round history table for a better overview (Appendix: Figure 25).

The specific objective of this experimental setting was to investigate a possible difference in the probability of exploring of people when facing a changing environment. Therefore, the probability of exploring, i.e. the dependent variable, form the first experiment was compared with that from the second experiment. The independent variable for both experiments was therefore the environment - the static environment for the first experiment and consequently the dynamic environment for the second one. Data from several studies suggest that the main trigger for exploring or exploiting is the belief about the given choices. Understanding the link between the beliefs about the given options and the search probability in a dynamic setting will aid in understanding possible changes. In the following sections I describe in detail more insights about the participants and how the model is structured.

4.2. Participants

The experiment was sent to a total number of 96 people living in Austria. 48 people were randomly assigned to each of the two experimental settings. From the original number of test persons, 70 people remained in the end. This shrinkage is due to the fact that either the person did not fully complete the experiment, or it was absolutely not understood. Thus, I have collected data on 100 choices times 35 participants = 3500 choices in the static environment and 100 choices times 35 participants = 3500 choices in the dynamic environment. All participants took part in the experiment voluntarily.

4.3. Model

4.3.1. Explorative and exploitative behavior

I define the trade-off between exploring and exploiting of people as my variable of interest and thus my dependent variable. As mentioned in the previous chapters the search strategy is the trade-off between people exploiting their decisions or exploring for new knowledge. In view of the statement by Posen & Levinthal (2012) that agents have to choose from a variety of decisions in each period and can either choose to continue with the same choice made in the previous period- or choose a different option, I measure the probability of exploration. It will be measured in unique choices. In this case unique means that the subject chooses an arm which is not the same as the one selected immediately before. After counting all unique choices, I calculate the probability of exploring by dividing the number of unique choices by the total number of rounds. In order to be able to make a statement about the exploitation of the knowledge, I simply applied the following formula: 1 – probability of exploring. In the existing literature there are now and then slightly different definitions of what exploring actually is. According to March (1991), exploring also refers to the willingness to forgo the possibility of engaging in an action that one considers to be superior. Conversely, this means consciously choosing decisions that you are not so strongly convinced of. Posen & Levinthal (2012) contradict these partly different definitions by noting that they also examined other metrics for exploration in their investigations. However, nothing changes in the qualitative results with alternative measures considered. However, this does not necessarily mean that it is also valid for my experiment. Therefore, I also measure the frequency with which the individual arms were played, which allows me to make a statement about the exploratory behavior without being dependent on how the test subjects came to their result. This means that the playing frequency of the arm is a round-independent measure and focuses exclusively on how often the participants have played a particular arm.

4.3.2. Environment

The environment in which the participants have to make their decisions reflects the independent variable for this study. In the first experiment I consider the case of a static environment. This means that each arm has the same individually selected expected value and variance for the entire 100 rounds. If the subject gathers enough information through experimenting it will find the best possible arm fast. The static environment allows to fully rely on collected information which helps to create an easy search strategy.

In the second experiment I consider the case of a dynamic environment. An environment is considered as dynamic if the landscape undergoes exogenous shocks. Therefore, two different expected values and variances are assigned to each arm. However, the second value differs from the first in every respect. If a shock occurs, a payout is drawn for the arm based on the second values. Following this definition, the underlying reward distribution for each arm changes frequently. The selected probability of shock occurring is 50%. This specific selection of the probability enables the participants to be indifferent between exploring and exploiting.

4.3.3. Subject´s beliefs

The experimental design involves the measurement of the subject’s beliefs about the different arms. Using a 5-point Likert scale, participants were asked to rate how strongly or weakly they believed that they have chosen the best arm in this round. The question started after the first selection and continued every 5 rounds. The participants were made aware of the meaning of this question in the introduction. An additional effect of this measurement is to identify possible misunderstandings of the participants which helps to increase the quality of my data.

4.3.4. Additional settings

As of today, literature suggest risk aversion as one factor why people do not search enough. For this reason, I am controlling for risk, so that there is no risk advantage in switching arms. More about that in chapter 4.3.5 the characteristics of the arms. Another important factor are the search costs because they can lead to insufficient exploring. An easy way to manage it is to not charge people if they decide to explore new arms (Pratt et al., 1979). At the beginning, the participants do not know what payout they can expect for the individual 5 arms. Nevertheless, the participants are given some important information about the arms and the environment. This information includes whether the environment changes the payout of the arms or whether it remains the same over the entire duration of the game. It is also made clear that the arms are subject to a ranking. So, there is a best, second best, third best, fourth best, and worst arm. Unlike many other experiments with multi-armed bandits, each participant goes through exactly the same number of 100 rounds.

Both experiments were created with oTree and Python and then implemented on a website using the Heroku app and HTML coding (Chen et al., 2016). Each participant was given a separate link to the website. Therefore, everyone could only participate exactly once. The different arms were represented with simple buttons. At the decision page subjects saw the balance of their total winnings so far and a summary table (Miller, 1956). All participants took part in the experiments voluntarily and were not paid. Only the knowledge of whether one was an optimal strategist was made available to each participant.

4.3.5. The characteristics of the arms

Both experiments consist of 5 different arms where each payoff is independently drawn from a normal distribution. The individual arms differ in that each has its own expected value and variance. For the purpose of risk control in the static environment, the mean-variance ratio is the same for each arm. To ensure that participants can develop an optimal strategy, the 95%-confidence intervals for each arm do not cross.

People often make decisions that are not understandable to others at first glance. Regardless of the assumption that everyone is operating to maximize utility, the lack of understanding lies in the different evaluations of the physical dimensions. For example, service quality, speed, accuracy, and much more are assessed differently by each individual. Therefore, I operationalize differences in perceived value with money. This means that the value of each arm is expressed in a dollar value which helps to induce participants to take their task seriously. A monetary value is more tangible than a hypothetical value without a unit (Friedman & Sunder, 1994; Gans et al., 2007).

5. Results

5.1. Descriptive Statistics

Of the 35 participants who completed the static environment 20 are males and 15 females with an average age of 39,45 years. Their performance ranged between $100715,97 and $133264,02, with an average of $121796,87. Of the 35 participants who completed the dynamic environment 15 are males and 20 females with an average age of 42,4 years. Their performance ranged between $90597,34 and $114911,06, with an average of $103416,30. With regard to the further course of this chapter, the main focus is on the differences in the search behavior of the participants between the dynamic and static experiment.

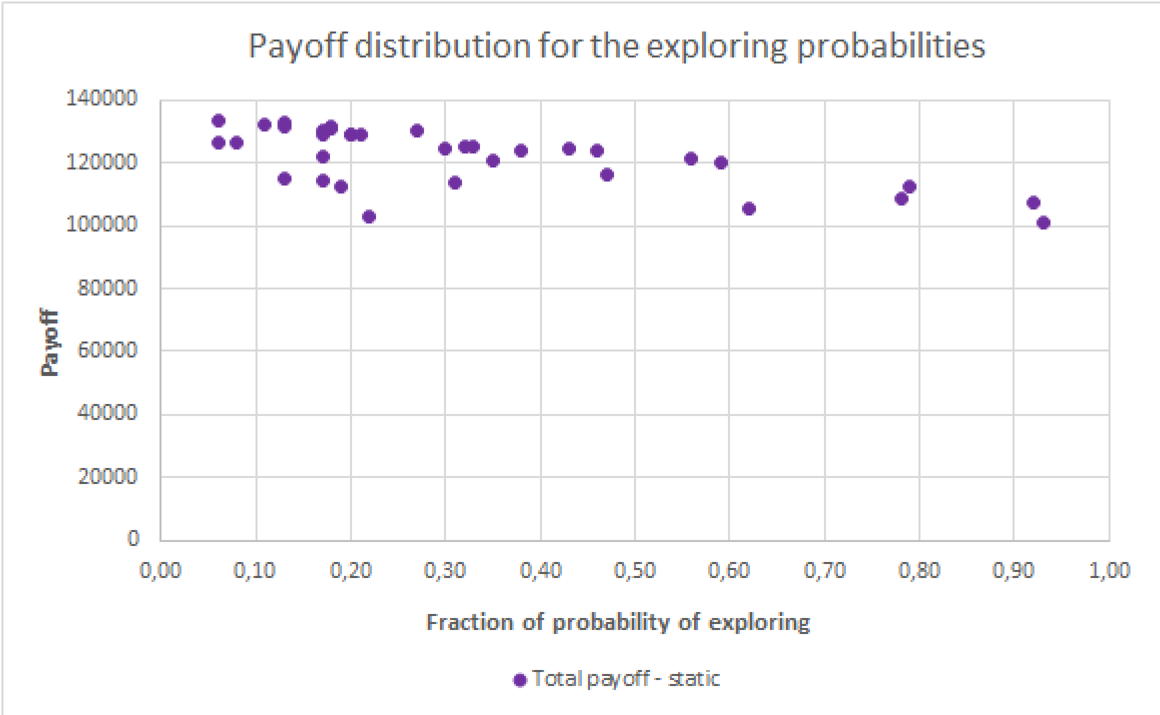

Figure 1: Payoff distribution for the exploring probabilities - static

Figure 1 presents the payoff from all the participants in the static experiment, considering their likelihood of exploratory behavior after 100 rounds. Closer inspection of the figure shows that the best results can be found at a low level of exploring. The best performance was at a 6% probability of changing the arm from the choice in the previous round. The results obtained from the dynamic experiment are shown below in Figure 2. Similar to the results from the static experiment the best results are located at a low probability of exploring. Hence the best performance was with a probability of exploration of 11%.

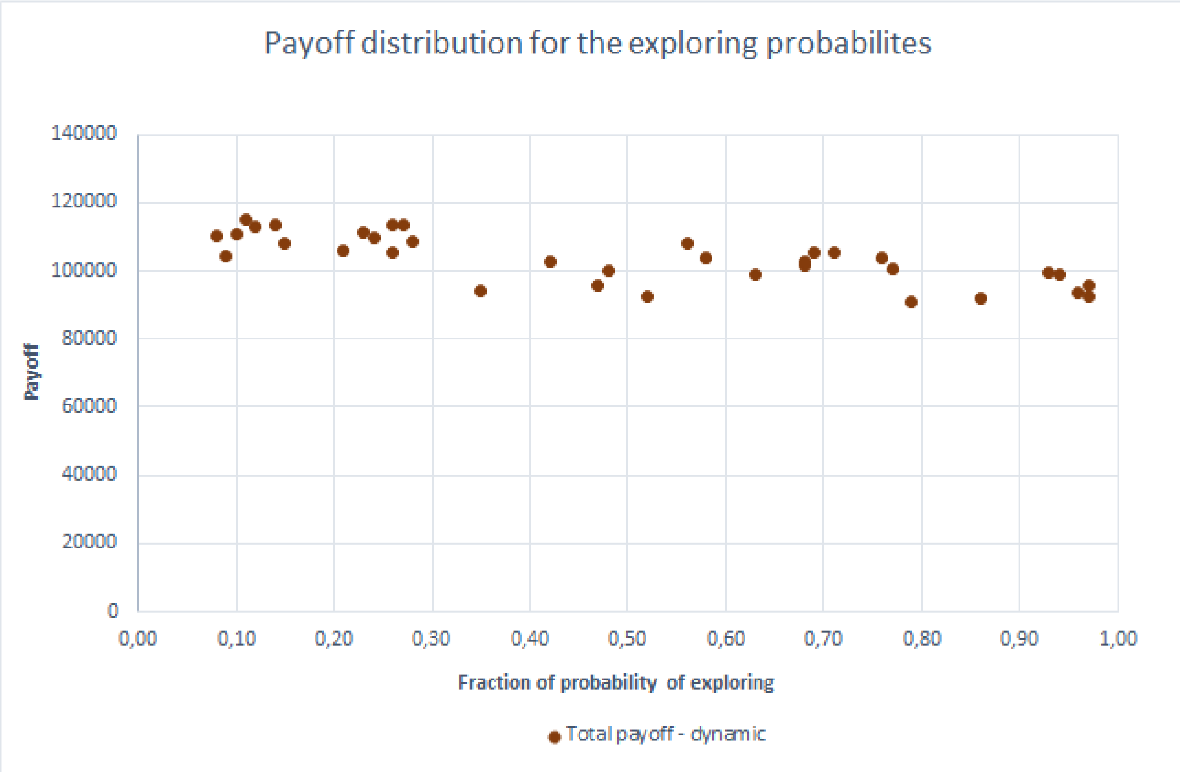

Figure 2: Payoff distribution for the exploring probabilities - dynamic

5.2. Results of working hypothesis 1

The purpose of hypothesis 1 was to predict a difference in the probability of exploring between the static and dynamic environment. Changes in the probability of exploring were compared using a nonparametric test, namely the Mann-Whitney test. Figure 3 presents the results of the experimental data on the probability of exploring. As seen in Figure 3 below, the minimum probability of exploring in the static environment was 6% and the maximum 95% with an overall mean value of 33%. The participants in the dynamic environment explored with a mean probability of 49,6%, the lowest probability being 8% and the highest 98%. The Mann-Whitney test revealed a significant difference between the probability of exploring in the static and dynamic environment ($p = 0,026$). This result is significant at the $p = 0,05$ level.

Summary statistics:

Variable	Observations	Obs. with missing	Obs. without	Minimum	Maximum	Mean	Std. deviation
Probability exploring - static	35	0	35	0,060	0,950	0,330	0,245
Probability exploring - dynamic	35	0	35	0,080	0,980	0,496	0,300

Mann-Whitney test / Two-tailed test:

U	422,500
U (standardized)	-2,227
Expected value	612,500
Variance (U)	7243,605
p-value (Two-tailed)	0,026
alpha	0,05

An approximation has been used to compute the p-value.
The continuity correction has been applied.

Test interpretation:
H0: The difference of location between the samples is equal to 0.
Ha: The difference of location between the samples is different from 0.
As the computed p-value is lower than the significance level alpha=0,05, one should reject the null hypothesis H0, and accept the alternative hypothesis Ha.

Ties have been detected in the data and the appropriate corrections have been applied.

Figure 3: Mann-Whitney test probability of exploring

As shown in Figure 4, a clear difference in the probability of exploring between the two conditions can also be seen graphically. The most interesting aspect of Figure 4 is that in the static environment, almost 28 participants had a probability of exploring below 50%. In contrast, in the dynamic environment, it is only just under half of the total 35 participants.

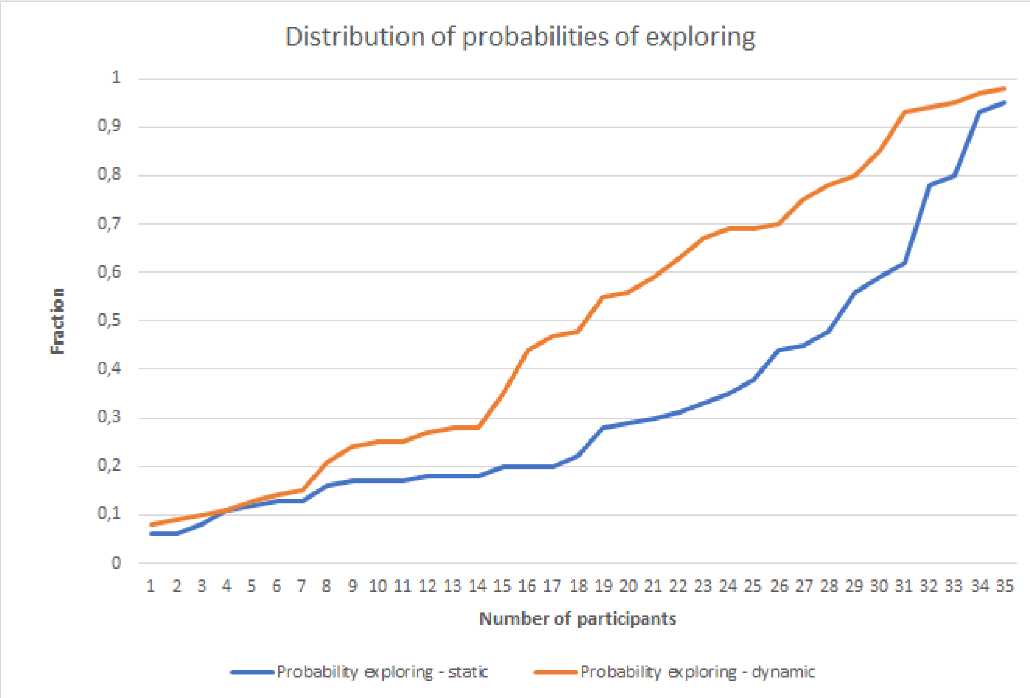

Figure 4: Distribution of probabilities of exploring

In order to evaluate the strength of the difference between the two mean values, I then calculated the effect strength using the results from the Mann-Whitney test according to Cohen (1988). The calculated effect strength $z = 0{,}2662$ which means that the effect falls into the category of a small effect ($z < 0{,}3$).

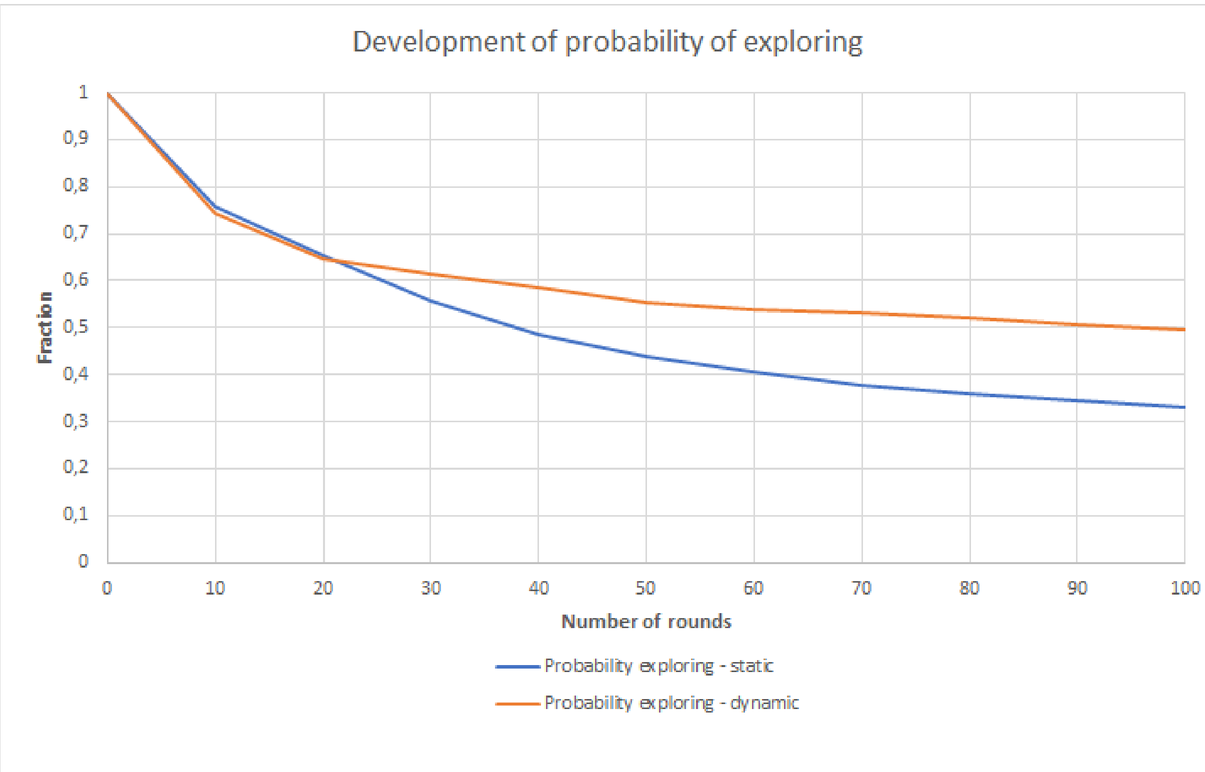

Figure 5: Development of probability of exploring

The first set of analyses examined the impact on the mean values. The following Figure 5 illustrates the development of the probability of exploring over the entire duration of the experiment.

What is striking about Figure 5 is that in the first 20 rounds of the experiment no difference in the probability of exploring between the static and dynamic environment exists (-probability of exploring = 65%). However, in both cases the probability curve equals a sloping curve, which becomes flatter as the experiment continues. The outstanding point in the graph is around round 20 because from that point on the probability of exploring decreases faster in the static environment. Further analysis showed the differences in the search probability within a 10-arm period. These differences between the static and dynamic environment are highlighted in Figure 6 and Figure 7. Within rounds number 1 – 10 and 11 – 20 no significant difference between the two environments was evident. Interestingly, the probability of exploring in the static environment significantly shrunk until round 30 was over. The difference between rounds number 31 – 40 and 91 – 100 was not significant (p = 0,075).

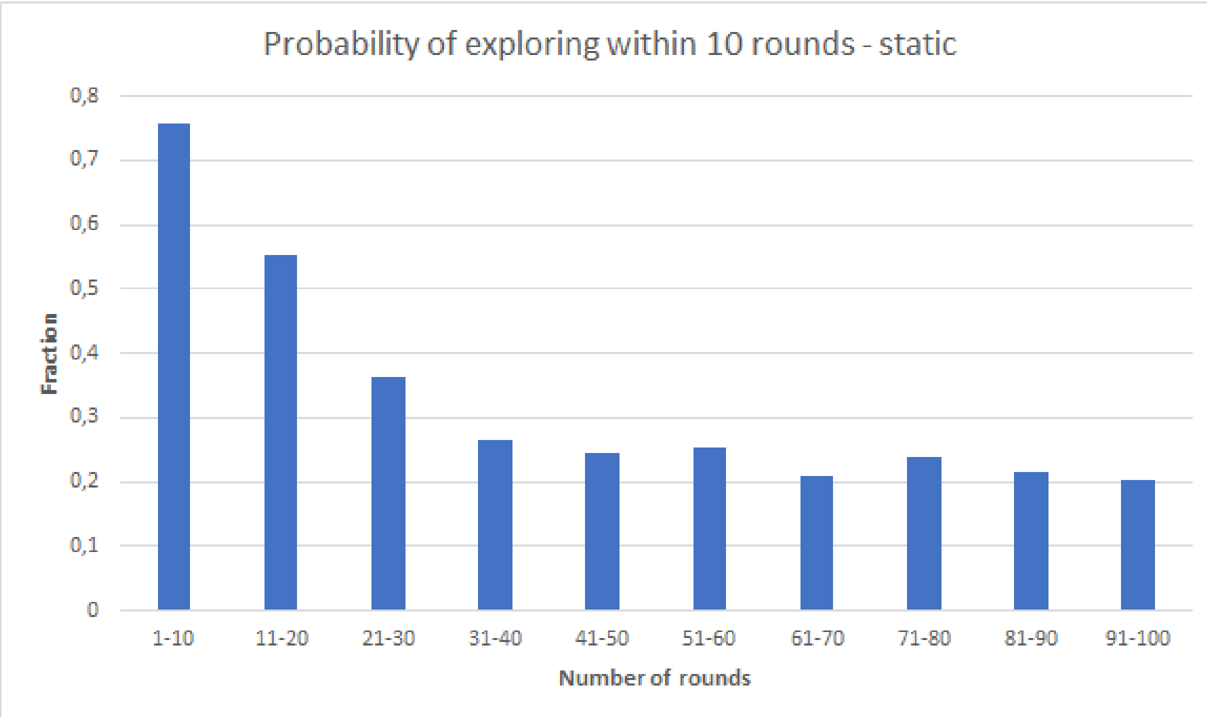

Figure 6: Probability of exploring within 10 rounds - static

This indicates that there no longer was any significant difference between all subsequent 10-arm periods.

A different result showed a closer examination of the dynamic environment. Already after round 10 there was no significant difference between the individual 10-arm periods. The probability of exploring

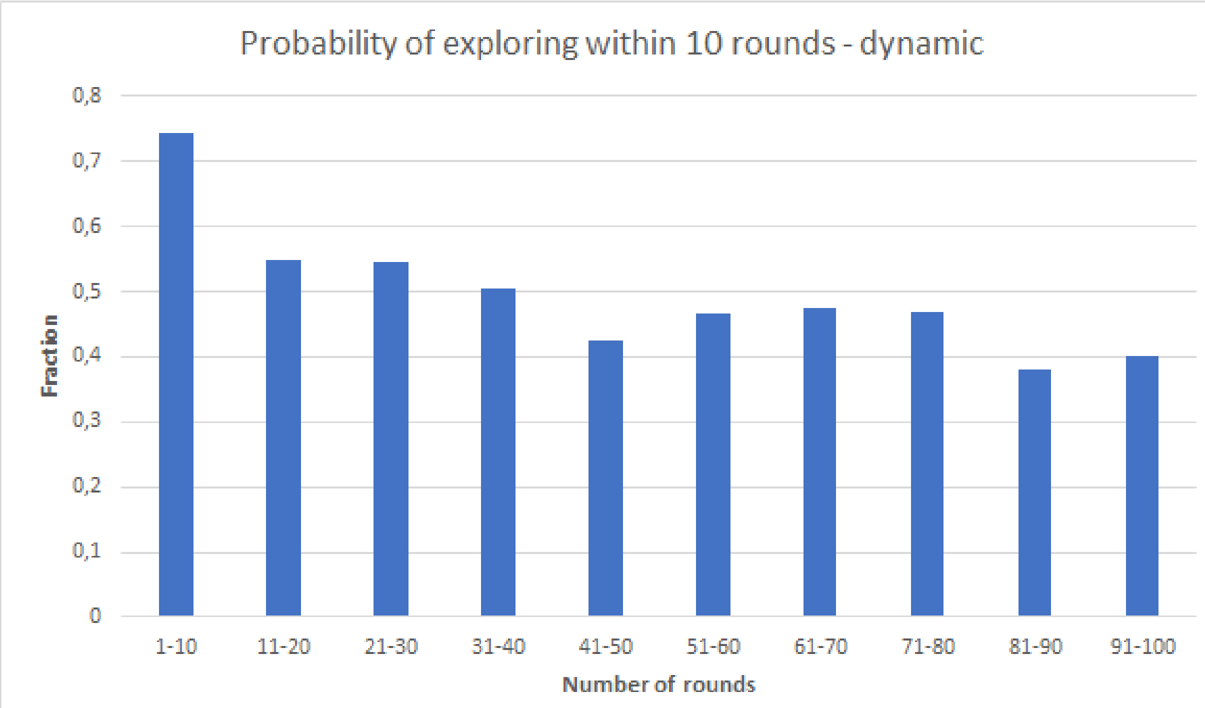

Figure 7: Probability of exploring within 10 rounds - dynamic

decreased slightly, but the difference between rounds 11 – 20 and 91 – 100 was not significant (p = 0,102).

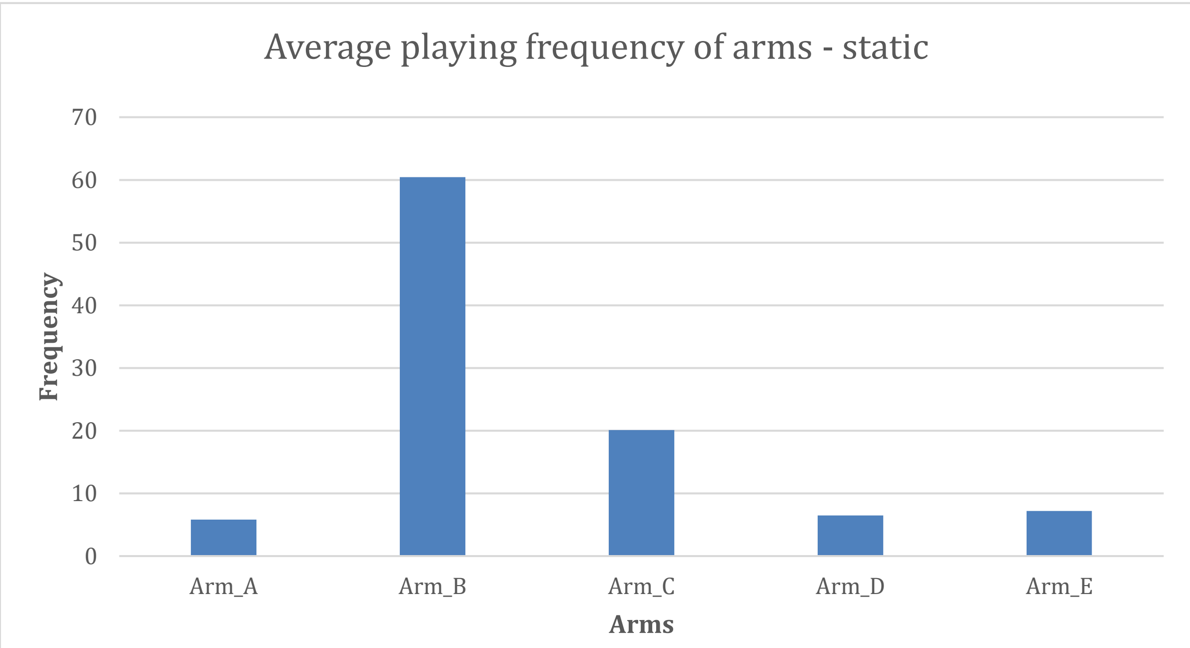

Figure 8: Average playing frequency of arms - static

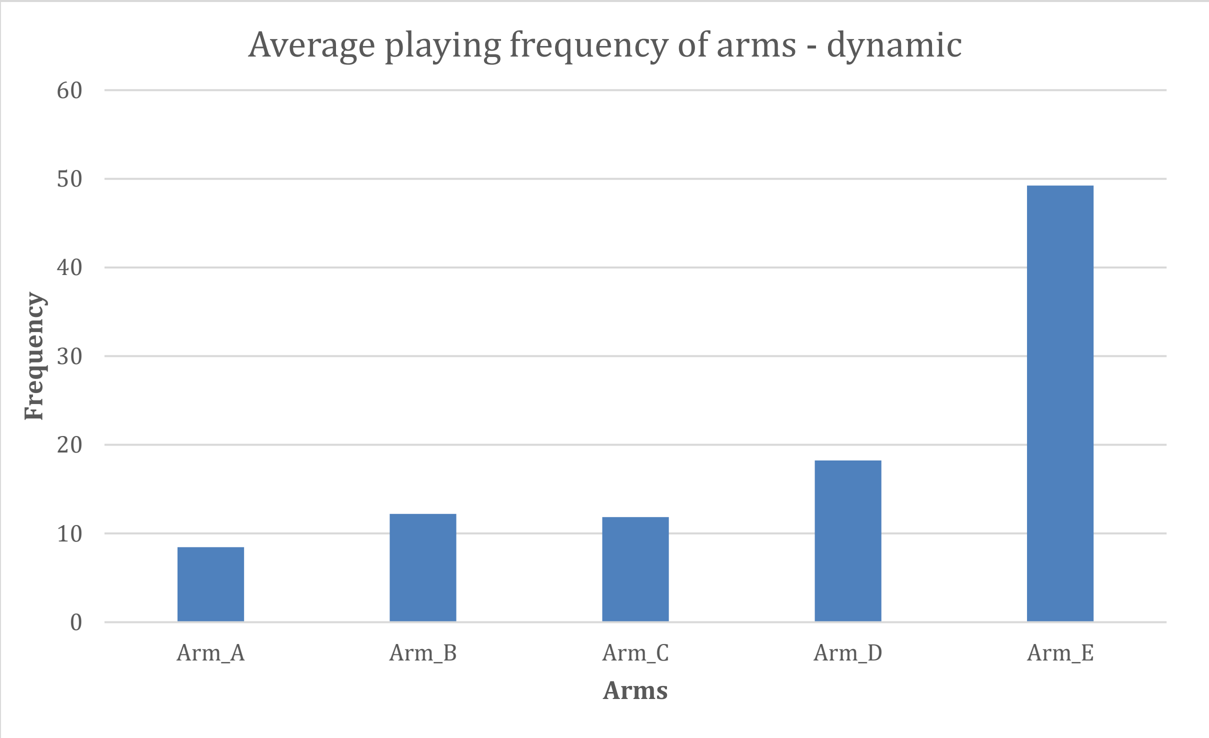

Figure 9: Average playing frequency of arms - dynamic

It can be seen from the data in Figure 8 and Figure 9 that the participants in both environments played one arm significantly more frequently than the other possible arms. What stands out in the tables is not only the fact that participants in the static environment played the apparently two best arms more frequently but also the circumstance that the assumed weaker arms were played more frequently in the dynamic case. Even though the strategy in both environments is top heavy it seems participants played different arms more frequently in the dynamic environment.

5.3. Results of working hypothesis 2

In order to delve even deeper into the possible differences in the search probability of the two environments, Hypothesis 2 was brought into being. Participants were asked to rate their current belief that their previous selection was the best possible. Precisely this development of beliefs in the course of the experiment and its connection with the search probability will be examined in the following section. To compare the difference in the mean values of the participants' beliefs between the static and dynamic environment a Mann-Whitney test was carried out. Figure 10 below illustrates the results.

Summary statistics:

Variable	Observations	Obs. with missing	Obs. without	Minimum	Maximum	Mean	Std. deviation
Average beliefs - Static	35	0	35	1,619	4,762	3,810	0,702
Average beliefs - Dynamic	35	0	35	1,000	4,905	3,124	0,973

Mann-Whitney test / Two-tailed test:

U	884
U (standardized)	3,184
Expected value	612,500
Variance (U)	7242,210
p-value (Two-tailed)	0,001
alpha	0,05

An approximation has been used to compute the p-value.
The continuity correction has been applied.

Test interpretation:
H0: The difference of location between the samples is equal to 0.
Ha: The difference of location between the samples is different from 0.
As the computed p-value is lower than the significance level alpha=0,05, one should reject the null hypothesis H0, and accept the alternative hypothesis Ha.

Ties have been detected in the data and the appropriate corrections have been applied.

Figure 10: Mann-Whitney test average beliefs

As shown in the summary statistics, the average beliefs of the participants in the static environment ranges from a minimum of 1,619 to a maximum of 4,762 with a mean value of 3,810. In the dynamic environment the value ranges from a minimum of 1 to a maximum of 4,905 with a mean value of 3,124. The results reflect in Figure 10, indicate a significant difference between the two environments ($p = 0{,}001$). This result is significant at the $p = 0{,}05$ level.

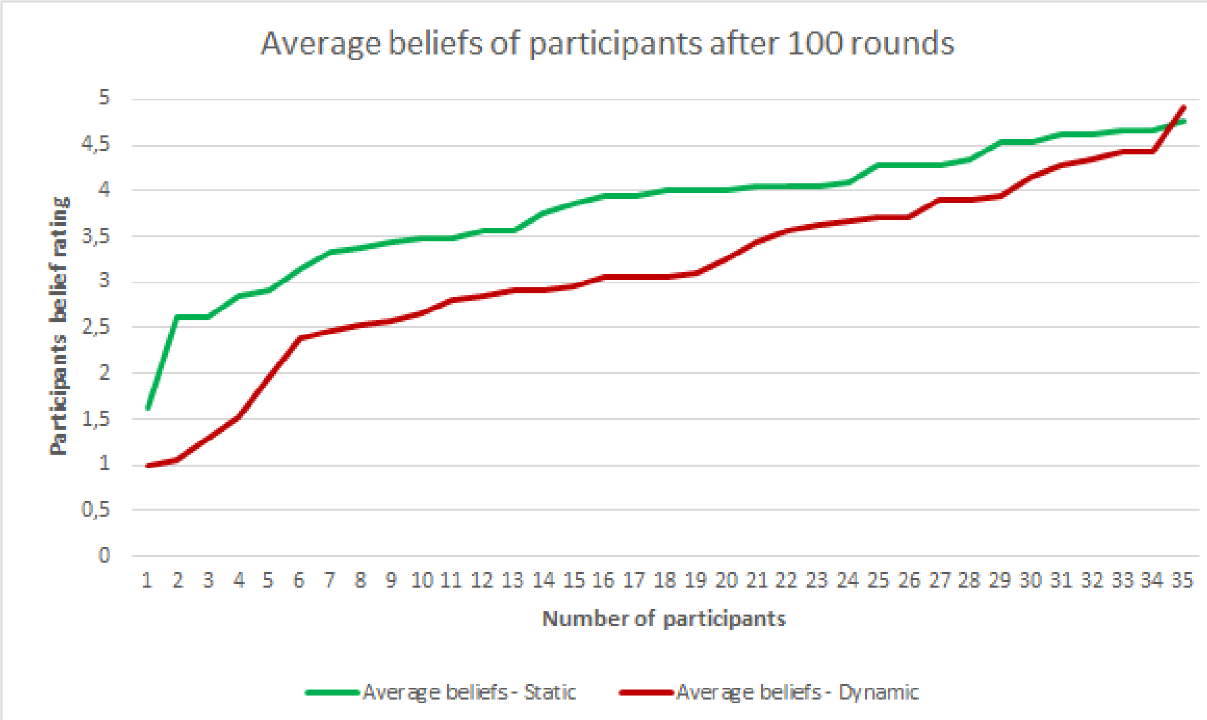

Figure 11: Average beliefs of participants after 100 rounds

Graphically, a clear difference between the two environments can also be observed, as demonstrated in Figure 11. What is interesting about the data in this figure is that in the static environment only 5 participants were below a mediocre level of beliefs, whereas in the dynamic environment there are almost 16 people below this level. The strength of the effect on participant's beliefs was assessed using a test suggested by Cohen (1988). The calculated effect strength z = 0,38 which means that the effect falls into the category of a medium effect (z = 0,3 – 0,5). Figure 12 compares the average beliefs of the participants after playing a certain number of rounds. This figure is quite revealing in several ways. First, similar to Figure 5 the interesting round number is 20. In the static environment the highest increase in beliefs was between the period of playing 10 rounds and 20 rounds. After just 20 rounds, the participants strongly believed that they had chosen the best possible arm in the previous round. However, in the dynamic environment there is only a small change in the beliefs similar to the probability of exploring. Second, the two highest belief ratings for both environments were reached after playing 90 rounds. The slightly higher value after 10 rounds for the dynamic environment is not significantly different from the one from the static environment (p = 0,442). Finally, it can be said that the evaluations of the different beliefs more or less increase as the experiment continues, more so in the static environment.

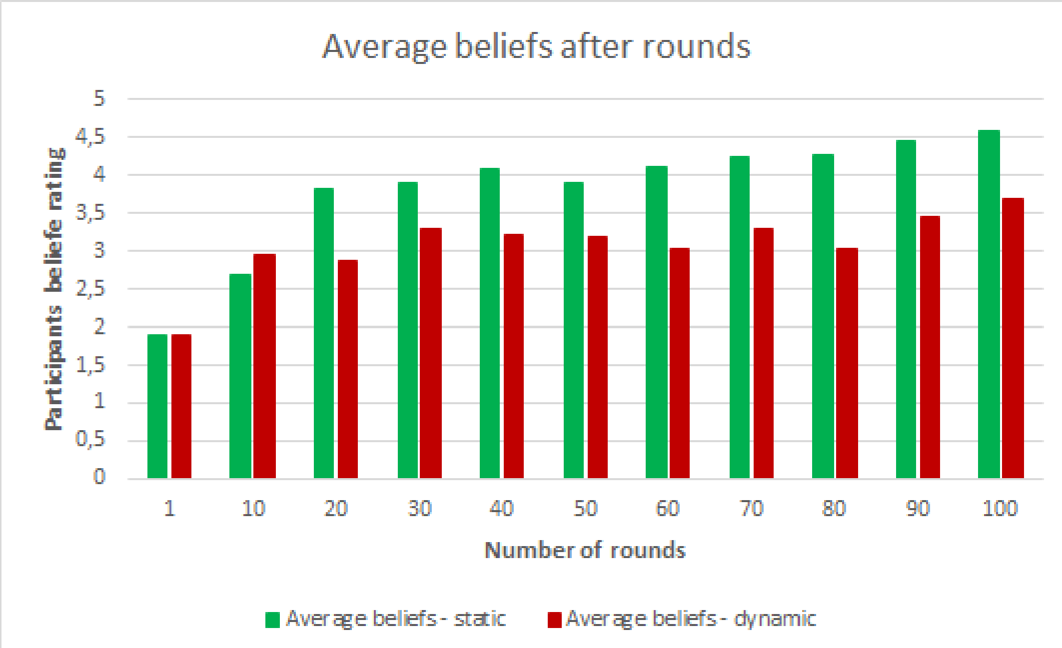

Figure 12: Development average beliefs over 100 rounds

In order to compare the course of the beliefs and the probability of exploring, Figure 13 and Figure 14 were created for the static and for the dynamic environment respectively. On the left axis is the probability of exploring and on the right axis the belief rating. A negative correlation was found between

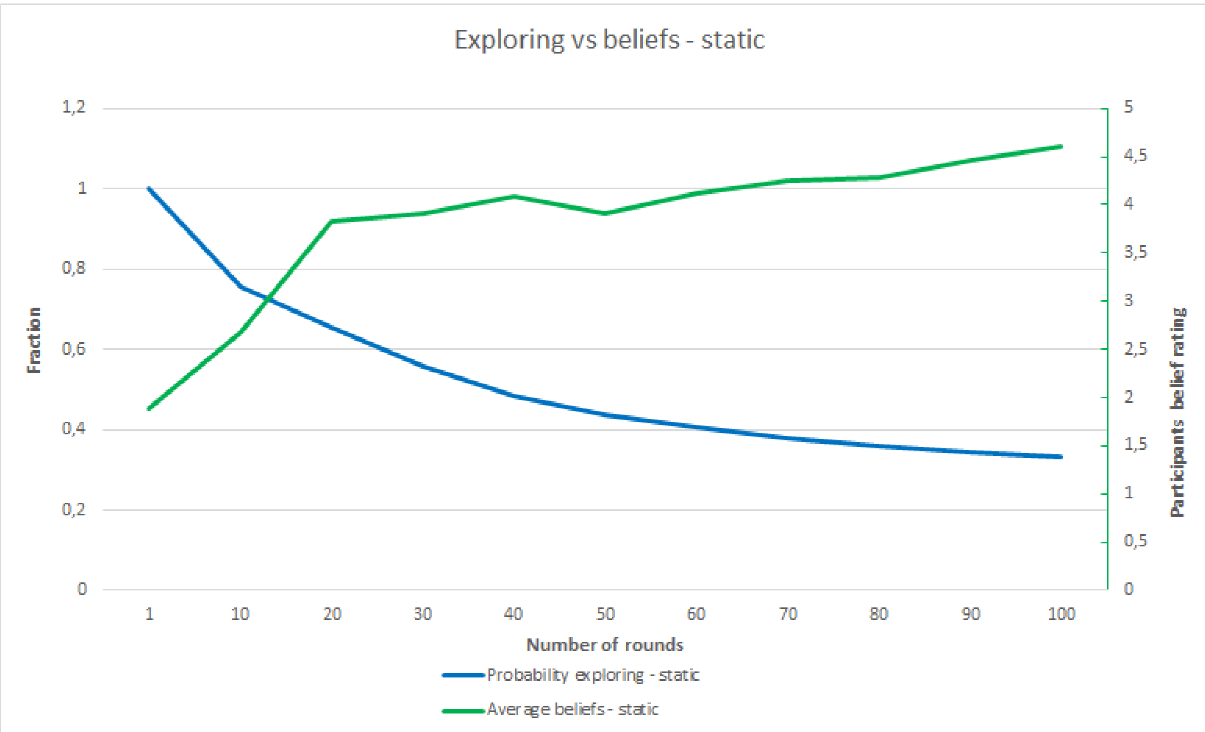

Figure 13: Exploring vs beliefs - static

the beliefs and the probability of exploring. With successive increase in the intensity of beliefs, the probability of exploring decreased.

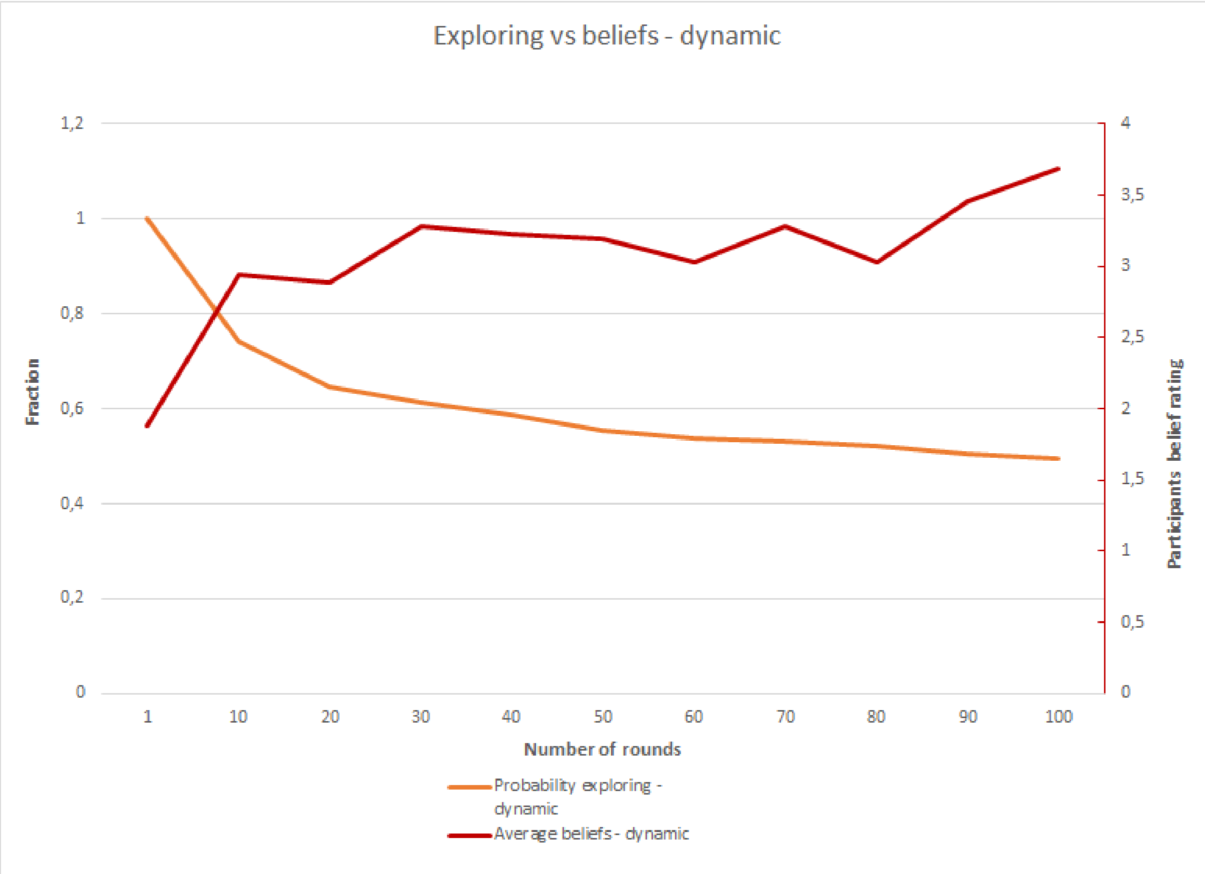

Figure 14: Exploring vs beliefs - dynamic

In order to subsequently check the role of beliefs as mediator, a mediator analysis was carried out. Therefore, various regressions were performed which ultimately follow the logic described by Baron & Kenny (1986). In order for an effect to be mediated by a third variable, according to the approach of Baron & Kenny (1986), there must be an effect at all. Therefore, the first step is to check whether the independent variable has a significant effect on the dependent variable. Figure 15 presents the results obtained from the first regression. The outcome is not surprising because it essentially supports the findings from the first hypothesis. So is the coefficient between the probability of exploring and the environment negative (coefficient = -0,165). This means that the probability of exploring decreases when changing the environment from the dynamic to the static case. The value of the change is equal to the difference of the two mean values. Finally, this result is significant at the p = 0,05 level (p = 0,014).

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0,292908453							
R Square	0,085795362							
Adjusted R Square	0,072351176							
Standard Error	0,273945956							
Observations	70							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	0,478915714	0,478915714	6,381595921	0,013864558			
Residual	68	5,103154286	0,075046387					
Total	69	5,58207						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	0,495714286	0,046305318	10,70534244	0,000000000	0,403313462	0,588115109	0,403313462	0,588115109
Static=1/Dynamic=0	-0,165428571	0,065485609	-2,526182084	0,013864558	-0,296103069	-0,034754074	-0,296103069	-0,034754074

Figure 15: Regression - probability of exploring vs environment

The second step is to check the effect of the environment on the participants' beliefs. This is another prerequisite for a possible mediation. As shown in Figure 16, the coefficient between the beliefs and the environment is positive (coefficient = 0,686). This means that a change in the environment from the dynamic to the static case increases the average participants' beliefs by 0,686. This result is significant at the p = 0,05 level (p = 0,001).

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0,37941912							
R Square	0,143958869							
Adjusted R Square	0,131370029							
Standard Error	0,848273439							
Observations	70							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	8,228571429	8,228571429	11,43543544	0,001198094			
Residual	68	48,93061224	0,719567827					
Total	69	57,15918367						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	3,123809524	0,143384381	21,7862608	2,20876E-32	2,837690446	3,409928602	2,837690446	3,409928602
Static=1/Dynamic=0	0,685714286	0,202776137	3,381632067	0,001198094	0,281080805	1,090347766	0,281080805	1,090347766

Figure 16: Regression - beliefs vs environment

The third step is to check whether the effect of the beliefs is significant on the probability of exploring.

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0,54066112							
R Square	0,292314446							
Adjusted R Square	0,281907306							
Standard Error	0,241025697							
Observations	70							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	1,631719702	1,631719702	28,08787356	1,35458E-06			
Residual	68	3,950350298	0,058093387					
Total	69	5,58207						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	0,998722429	0,114210772	8,744555439	9,77287E-13	0,770818381	1,226626476	0,770818381	1,226626476
Average beliefs - Total	-0,168958393	0,031880149	-5,299799389	0,00000135	-0,232574239	-0,105342546	-0,232574239	-0,105342546

Figure 17: Regression - probability of exploring vs beliefs

Figure 17 shows a negative coefficient (coefficient = -0,169) which means that if the participants' beliefs increase by one unit, the probability of exploring shrinks by 16%. This effect is significant at the p = 0,05 level (p = 0,000). The last step is to enter the mediator into the model and check how the effect of the environment on the probability of exploring changes. The result of the regression is presented in

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0,548920558							
R Square	0,301313779							
Adjusted R Square	0,280457474							
Standard Error	0,241268891							
Observations	70							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	2	1,681954605	0,840977303	14,44713132	6,0737E-06			
Residual	67	3,900115395	0,058210678					
Total	69	5,58207						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	0,98553151	0,115204439	8,554631367	2,41092E-12	0,755582487	1,215480534	0,755582487	1,215480534
Average beliefs - Total	-0,156801246	0,034491414	-4,546094994	0,000023501	-0,225646397	-0,087956095	-0,225646397	-0,087956095
Static=1/Dynamic=0	-0,057907717	0,062335434	-0,928969506	0,356239742	-0,182329764	0,06651433	-0,182329764	0,06651433

Figure 18: Regression - Full model

Figure 18. The question of interest is whether the environment predicts the probability of exploring and whether the direct path would be mediated by the beliefs. The first three steps showed a significant relationship between all the variables. However, in the full model case the participants' beliefs still have a significant influence on the exploring probability (p = 0,000) but the environment no longer has a

significant impact on the probability of exploring (p = 0,3562). This suggests that the relationship between the environment and the probability of exploring is fully mediated by the participants' beliefs.

Input:			Test statistic:	Std. Error:	*p*-value:
a	0.685714286	Sobel test:	-2.71328871	0.0396275	0.0066619
b	-0.156801246	Aroian test:	-2.67199118	0.04023997	0.00754026
s_a	0.202776137	Goodman test:	-2.75656224	0.03900542	0.00584125
s_b	0.034491414	Reset all	Calculate		

Figure 19: Sobel test

In order to test if the resulting indirect effect in total is significant a simple Sobel test was performed. This evaluates the null hypothesis that the indirect effect in the population is zero. A significant Sobel test means that there is a significant indirect effect. As shown in Figure 19, the Sobel test is significant (p = 0,007) (Sobel, 1982).

6. Discussion

As mentioned in the literature review, very little is known about the association between the trade-off in the decision-making process and a dynamic environment. Prior studies that have noted the importance of the trade-off between exploiting existing and exploring new knowledge suggested the term of exploiting as more preferred by individuals in a static environment (Anderson, 2000; March, 2003; Sutton & Barto, 1998; Weitzman, 1979). The present study was designed to determine the effect of the dynamic environment on this trade-off. The first starting point in the discussion of the experimental results is that which deals with the different performances in the respective environments. The most important question that arises regarding performance is – What is the probability of exploring to perform best and how much do people deviate from these benchmarks in the static and the dynamic environment? A theoretical calculation of the best strategy can be found in the vast amount of literature available on the subject. What almost all of these theoretical approaches have in common is the starting point that the participants always act in a way that maximizes their benefits. Surprisingly, benefit-maximizing behavior was not fully observed given the data on the performance. It can be clearly seen that some participants achieved significantly different results despite the identical or almost identical probability of exploring. It is somewhat surprising that this irrational behavior was observed in both the static and dynamic experiment. The static experiment in particular is one that can be easily learned with sufficient collection of information. Particularly in the higher areas of the exploring probability, this phenomenon cannot be reduced to just a single wrongdoing. Rather, it is due to a combination of different factors such as overestimation, failure to understand the experiment or simply the limitations of the experimental design. This phenomenon can be explained at least in the low ranges of the exploring probability. In their paper, March (2003) found that people are exploring too little. This leads to a competency trap which can be seen in my experimental data. For example, I had three participants which explored at a level of 6 - 8 % in the static environment. However, only one of them had an excellent performance. The other two participants fared significantly worse than the best among them. This performance is the result of insufficient search because they were technically unable to find the best arm.

In order to reduce the influence of these different factors it was therefore necessary to divide the participants into 5 different groups. So, group 1 represents the best 7 participants, group 2 the second best 7 participants and so on. The best group in the static environment had a probability of exploring of 15% with a payoff of $131705,67. The best group in the dynamic environment had a probability of exploring of 18% with a payoff of $112733,26. The most important relevant finding was that neither in the static nor in the dynamic environment did the participants perform on average significantly close to the optimum. Nevertheless, there are clear indications that the dynamic environment significantly intensifies this suboptimal exploring behavior. On the one hand, significantly more participants are close to the optimum in the static experiment and on the other hand, the difference between the best search value and the mean value of all participants is significantly smaller.

A question that is omnipresent for the quality of the experiments involves determining whether the randomization of the participants was successful. This is particularly important in order to rule out that the relationships between the variables are not based on an incorrect randomization or other disruptive factors but rather arose due to the different interactions of the variables. On the one hand the demographic distribution between both experiments is similar and on the other hand there is some empirical evidence in the data. After evaluating the round-based probability of exploring, it was found that there was no difference in the probability of exploring between the static and dynamic environment within the first 20 rounds. It is therefore likely that differences in people's search strategies are equally distributed among the two experiments. Another source of empirical evidence was the average beliefs after rounds in Figure 11. At the beginning, the participants in both experiments were almost identically convinced of the respective decisions. Only with the continuation of the two experiments did the convictions develop to different degrees. It can therefore be assumed that the findings are based on the different interactions between the variables.

6.1. Discussion of working hypothesis 1

The research question in this study sought to determine the influence of a dynamic environment on the individual search behavior. The first cornerstone of this study therefore relates to the differences in the participants probability of exploring. Looking at the data from this research, one concludes that they broadly support the first hypothesis. It was shown that the participants in the static experiment showed an exploratory behavior with a probability of 33%. Conversely, this result means that the participants acted exploitive with a 67% probability. In accordance with the present results, previous assessments have demonstrated that people are more likely to behave in an exploitive manner. In the dynamic experiment, as suggested in Hypothesis 1, there is a significant change in search behavior. Here the data shows that the participants had an almost 50% probability of engaging in exploratory behavior. In general, therefore, it seems that the dynamic environment lowers the described exploitative bias more in favor of exploring new alternatives. It is encouraging to compare this data with that found by Posen & Levinthal (2012) who concluded that at low level of environmental shocks the best answer is to increase exploring. A note of caution is due here since it is not clear what is exactly meant under a low level of shocks. A change in the search probability can not only be seen by comparing the two mean values. Therefore, the data in this paper also provides information on how the search probability has developed over the course of the experiment. The results of this study indicate that in the first 20 rounds participants show no difference in their reaction on the environment. After round 20 the probability of exploring drops significantly faster in the static than in the dynamic environment. A possible explanation for this might be that after reaching round 20 participants start to understand the different circumstances in the environment. This result may be explained by the fact that the available information in the static environment do not change within the course of the experiment. Therefore, after gathering sufficient information, the participants could easily identify the best arm. Put simply, the participants have learned which arm is the best. If one recalls the characteristics of a dynamic environment, one can see that the information available can change constantly. It is precisely this fact that makes it impossible, or at least very difficult, to draw a learning effect from the information collected. As my data shows, the point at which one can see the result of the different learning effect comes around round 20. These results further support the idea of Posen & Levinthal (2012) who argued that the changes in the dynamic environment reduces the lifespan of generated knowledge and therefore its value. This makes it almost impossible to use a search strategy which relies on past information.

Differences can be determined not exclusively with regard to the entire course of the experiment. Therefore, differences between the static and dynamic environment became apparent when looking at the rounds in isolation. It turned out that the exploring probability of the participants in rounds 1 to 10 in both the static and dynamic experiment was around 75%. The probabilities in rounds 11 – 20 were again almost the same in both cases with a value of roughly 55%. From this point on, the previously mentioned difference in learning about the given environment is visible again. In the static environment participants explored within the rounds 21 – 30 with a probability of 36%. In contrast, the participants in the dynamic experiment explored within these rounds with a probability of 55%. It is interesting to note that after the 30th round, the probability of exploring within the 10-round timeframe doesn't significantly change much in the static environment. These findings suggest that participants were looking for information up to and including round 30. From this point on it seems as if no more new knowledge can be generated and the participants therefore complete the rest of the experiment with a fixed search probability of roughly 23%. The picture is different in the dynamic environment. After just 10 rounds, the participants no longer changed their search probabilities significantly. The values here were, for example, 55% for rounds 11 – 20 and 40% for rounds 91 – 100. Another possible explanation for this is the emerging blandness in relation to the static environment compared to the possibility of constantly generating new knowledge in the dynamic environment. It seems possible that after the 30th round the participants arguably collected all the given information about each individual arm which leaded to a consistent low probability of exploring. On the other hand, due to the constantly changing environment in the dynamic experiment, there is always the possibility of learning new things. These findings cannot be extrapolated to all participants. This is due to the different evaluation of people when it seems to be worth to collect new knowledge. For this reason, in my work I have set the probability with which an arm can change at 50%. At this level people should be indifferent between exploring new knowledge and exploiting existing knowledge. One interesting finding is that the mean probability of exploring in the dynamic environment is indeed at 50%. However, the probability of exploring might not fully reflect the true values of exploring. This is because of the way participants choose their arms is not taken into account. Two participants may have played the same frequency of the five arms, but still have a different probability of exploring. However, this possible source of error can be mitigated. On the one hand, the theory shows that exploratory behavior can largely be found at the beginning of a decision process. This behavior could also be seen in my results since the exploratory behavior in both experiments was higher in the first part of the experiment than in the second part. Additionally, there were only two participants in the dynamic environment and three in the static one which used a different strategy of selecting arms than the vast majority of participants by initially choosing the same arms four or five times behind each other before sticking to the best one. On the other hand, there are many indications of good randomization of the experiments, which in turn makes the different forms of arm selection equally likely for both experiments. Nonetheless, at the end I compared the different playing frequencies of the two experiments. It was shown that the participants less often played the best and second-best arm but more often the other arms in the dynamic environment. The selecting frequency is completely independent of the constellation how the arms were selected and supports the results from the search probability analysis. Another possible explanation for an increase in suboptimal behavior of people in a dynamic environment is that they use simple heuristics to apply strategies. The simple heuristic in the case of the multi-armed bandit was to act greedy. In other words, people followed the simple strategy of mainly choosing the arm which performed best in the previous round. In the dynamic environment the supposedly best arm was really the best arm in 50% of the time when no shock occurs.

As can be seen from my data participants have indeed played one specific arm with a frequency of 50 times out of 100 rounds. Whenever the selected arm no longer had the best outcome, most of the participants changed their decision in the following round. Additionally, the average probability of exploring in the dynamic environment is approximately 50% which supports this explanation. The characteristics of the dynamic environment are not present in the static environment which means that the best arm remains the best most of the time. As a result, participants in the static environment have less incentive to switch arms.

6.2. Discussion of working hypothesis 2

As mentioned in the literature review, the urge for knowledge is the main driving force behind exploratory behavior. Collecting information increases the level of knowledge about an upcoming decision. In the context of a multi-armed bandit, knowledge is synonymous with an increase or decrease in the beliefs of an arm. Therefore, a strong relationship between exploratory behavior and the beliefs about each arm has been reported in the literature. Basically, the second hypothesis serves the purpose of supporting and explaining the findings from the first hypothesis and to catch a glimpse of the different developments in the beliefs.

The results of this study show that the average beliefs of the participants were significantly higher in the static than in the dynamic environment which supports my second hypothesis. In the static environment participants reported on average to strongly believe (3,81) that their previous made decision was the best. In contrast to that participants in the dynamic environment reported on average that they are only moderately (3,12) sure that their previous made decision was the best. One interesting finding is that in the static environment only 5 participants reported average beliefs below the mediocre level (3). On the other hand, this number of participants increased to around 16 people in the dynamic environment. The calculated effect size of 0,38 also illustrates a mean effect in the different development of beliefs within the two experiments. As can be seen from the data, in the beginning of the experiments the participants were more or less equally convinced of how good their choice was. This result is logical because initially the participants have no information about how well each arm is performing. The non-existent difference in the beginning between dynamic and static environment makes it clear again that all participants started under the same conditions. Only with continuous information gathering about the different arms did the respective beliefs of the participants develop differently. In the case of the static environment, it was shown that the participants were already strongly convinced after the 20^{th} round, that the decision they made immediately before was the best. These results are likely to be related to the findings from hypothesis 1 where the upcoming difference in the probability of exploring began to change after the 20^{th} round. In the further course of the static experiment, it becomes apparent that the less new information is gained or, in other words, with increasing exploitation, the conviction of the participants increases. Even in the dynamic experiment, the beliefs have similar development patterns as the corresponding probability of exploring. On the one hand, belief increases slightly but not necessarily significantly during the course of the experiment, on the other hand, belief changes only very slightly after the 10^{th} round. Expressed in numbers, the conviction of the participants in the dynamic environment always fluctuates around a moderate conviction. Due to the fact that the beliefs develop at the same level in the beginning but differ in strength as the experiments continue, this fact can be attributed to the dynamic environment. Regardless of how much information the participants collect due to the shock, it cannot be clearly stated which arm will be the best in the respective round. Comparing the conviction with the probability of exploring, the same similarities can be seen in both the static and the dynamic environment. This shows a possible indirect influence on the probability of exploring triggered by beliefs, which is consistent with the broader literature in the field of multi-armed bandits and decision making. So in order to prove a possible mediation, I followed the test scheme according to Baron & Kenny (1986). The results of the necessary regressions showed a significant direct effect of the environment on the probability of exploring (Figure), a significant effect of the environment on the beliefs (Figure) and a significant effect of the beliefs on the probability of exploring in the full model (Figure 18). Another important finding was that the direct effect between the environment and the probability of exploring was no longer significant in the full model. This would speak in favor of a full mediation effect of the beliefs. A note of caution is due here since this result could have arisen because of the different robustness of the variables. Due to the previously mentioned equality in the beginning of the experiments, this explanation can be ruled out. It can therefore be assumed that the environment shapes the participants beliefs. In order to prove whether the indirect effect as a whole is significant a Sobel-test was performed. The result of a p-value = 0,0067 showed a significant indirect effect.

7. Conclusion

Each of us make different decisions every day and have to find the optimal mix between exploring new alternatives and exploiting of existing knowledge. The aim of the present research was to examine the influence of a dynamic environment on the individual search behavior. This process is a central part of the decision-making process because it is responsible for obtaining information. It is generally assumed that people tend to explore too little. This thesis was undertaken to design a multi-armed bandit and evaluate the trade-off between exploring and exploiting knowledge in the dynamic environment. The multi-armed bandit is a suitable model for simulating regular independent decisions and the dynamic environment allows a more realistic replica of reality.

The quantitative research of this thesis tried to examine those ideas and models by performing two multi-armed bandit experiments. 70 participants were asked to perform only one of the two experiments. Therefore, 35 participants were randomly assigned to partake in the first experiment. The characteristics of the first experiment were a static environment, five different arms with a fixed expected value and mean-variance ratio and 100 rounds to play. The remaining 35 participated in the second experiment, in which they were exposed to a dynamic environment. This means that each arm can change the respective expected value in the next round with a probability of 50%. This fact serves to represent a possible ever-changing world. The goal for all participants in both experiments was to collect as much money as possible within 100 rounds. In order to get a more precise impression of how the convictions of the participants develop in the course of the two experiments, every five rounds they were asked how sure they were that the decision they made in that round was the best.

The research design covers two important areas of the search process. The first part deals with the possible different development of the exploring probabilities in the two environments. It is assumed that the participants explore with a higher probability in a dynamic than in a static environment. The second part serves to illustrate a possible further change during the decision whether to obtain new information or to use existing information. More precisely, it examines the extent to which the participants’ beliefs about the individual arms are influenced in the dynamic environment. It is assumed that the beliefs strongly influence the explore probability of the participants and therefore the average belief is greater in the static environment than in the dynamic one.

These experiments supported both hypothesis and could answer the research question. In the static environment, the collected information can be understood more easily because each arm is only assigned a certain range of values. In the dynamic environment, on the other hand, there is the possibility of more than one value range for each arm. This additional information is therefore more difficult to interpret, which in turn reduces the beliefs of the participants and, as a result, leads to an increase in the search probability. The relevance of the environment in which decisions are made is clearly supported by the current findings.

7.1. Practical implications

Overall, this study strengthens the idea that humans differ in the evaluation of information. This also implies that they do not always behave optimally and, in some cases, even irrationally. The results of the first experiment also make it clear that even when the task is apparently easy, some people are not able to understand all the impressions immediately. These findings suggest that in general the more information is available the greater the chance of an optimal behavior. Another implication for making decisions in practice is to eliminate as much ambiguity as possible and to create the most exact framework. Humans engage frequently in decision-making behavior in everyday life and therefore making good decisions has high impact in the quality of life. The findings of this research provide insights about human behavior in a changing world. This study suggests that uncertainty should be avoided by people who want to make optimal decisions and that people often follow simple heuristics to apply strategies.

7.2. Limitations & future directions

One source of weakness in this study which could have affected the measurements of the exploration probability and beliefs was the strong dependency on a good randomization of the participants. The data can only be appropriately evaluated if there is empirical evidence for the randomization. This is due to the fact that the representation of the exploring probability does not take into account the way in which the participants ultimately came to their result. It can therefore be the case that two participants played the same number of different arms but still have a completely different exploring probability.

An additional uncontrolled factor is the possibility that the participants' beliefs are not well represented. This is due, on the one hand, to the fact that it is a subjective measure and therefore leaves a lot of room for interpretation and on the other hand, the question was only limited to the directly selected arm. A further limitation is the usage of the mediation analysis described by Baron & Kenny (1986). According to some recent work, this method has some weaknesses which also apply to the Sobel test. Notwithstanding these limitations, the study suggests that people do explore with a higher probability as a response to a dynamic environment.

These findings provide the following insights for future research: people sometimes behave irrational even in a static setting, uncertainty coupled with additional information exacerbates this problem and people use simple heuristics to develop strategies. Further work is needed to fully understand the implications of irrational behavior in different settings. More information about the behavior in different degrees of a dynamic environment would help us to establish a greater degree of accuracy on this matter. Further research should focus on determining additional factors which might lead people to suboptimal behavior. Above all, it should not be overlooked that a large number of decisions are not made completely alone. Further work needs to examine more closely the links between social pressure and suboptimal behavior in terms of the exploring probability.

References

Acuna, D., & Schrater, P. (2007). Bayesian Modeling of Human Sequential Decision-Making on the Multi-Armed Bandit Problem. *Bernoulli*, 2065–2070. http://www-users.cs.umn.edu/~acuna/downloads/papers/pp199-acuna.pdf

Anderson, C. M. (2000). *Behavioral Models of Strategie in Multi-Armed Bandit Problems.pdf*. California Institute of Technology.

Auer, P., Cesa-Bianchi, N., & Fischer, P. (2002). Finite-time Analysis of the Multiarmed Bandit Problem. *Machine Learning*, *47*, 235–256. https://doi.org/10.1023/A:1013689704352

Banks, J., Olson, M., & Porter, D. (1997). An Experimental Analysis of the Bandit Problem. *Economic Theory*, *10*(1), 55–77. http://www.jstor.org/stable/25055024

Baron, R., & Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182. https://doi.org/10.1037//0022-3514.51.6.1173

Baumann, O. (2015). Models of complex adaptive systems in strategy and organization research. *Mind & Society*, *14*. https://doi.org/10.1007/s11299-015-0168-x

Bergemann, D., & Välimäki, J. (2006). *Bandit Problems* (Cowles Foundation Discussion Papers, Issue 1551). Cowles Foundation for Research in Economics, Yale University. https://econpapers.repec.org/RePEc:cwl:cwldpp:1551

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. https://doi.org/https://doi.org/10.1016/j.jbef.2015.12.001

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. ed.). Erlbaum.

Friedman, D., & Sunder, S. (1994). *Experimental Methods: A Primer for Economists*. Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9781139174176

Gans, N., Knox, G., & Croson, R. (2007). Simple models of discrete choice and their performance in bandit experiments. *Manufacturing and Service Operations Management*, *9*(4), 383–408. https://doi.org/10.1287/msom.1060.0130

Gittins, J. C. (1979). Bandit Processes and Dynamic Allocation Indices. *Journal of the Royal Statistical Society: Series B (Methodological)*, *41*(2), 148–164. https://doi.org/10.1111/j.2517-6161.1979.tb01068.x

Gittins, J., Glazebrook, K., & Weber, R. (2011). Multi-Armed Bandit Allocation Indices: 2nd Edition. In *Multi-Armed Bandit Allocation Indices: 2nd Edition*. https://doi.org/10.1002/9780470980033

Hoelzemann, J., & Klein, N. (2017). *Bandits in the Lab*. 1–44. johanneshoelzemann.com

Keller, G., Rady, S., & Cripps, M. (2005). Strategic Experimentation with Exponential Bandits. *Econometrica*, *73*(1), 39–68.

March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, *2*(1), 71–87. http://www.jstor.org/stable/2634940

March, J. G. (2003). Understanding Organisational Adaptation. *Society and Economy*, *25*(1), 1–10. https://doi.org/10.1556/socec.25.2003.1.1

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. In *Psychological Review* (Vol. 63, Issue 2, pp. 81–97). American Psychological Association. https://doi.org/10.1037/h0043158

Posen, H. E., & Levinthal, D. A. (2012). Chasing a moving target: Exploitation and exploration in dynamic environments. *Management Science, 58*(3), 587–601. https://doi.org/10.1287/mnsc.1110.1420

Pratt, J. W., Wise, D. A., & Zeckhauser, R. (1979). Price Differences in Almost Competitive

Markets. *The Quarterly Journal of Economics*, *93*(2), 189–211. https://doi.org/10.2307/1883191
Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.*, *58*(5), 527–535. https://projecteuclid.org:443/euclid.bams/1183517370
Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *Int. J. Services Sciences*, *15*(4), 83–98. https://doi.org/10.1016/0305-0483(87)90016-8
Scott, S. L. (2015). Multi-armed bandit experiments in the online service economy. *Applied Stochastic Models in Business and Industry*, *31*(1), 37–45. https://doi.org/10.1002/asmb.2104
Siggelkow, N., & Rivkin, J. W. (2005). Speed and search: Designing organizations for turbulence and complexity. *Organization Science*, *16*(2), 101–122. https://doi.org/10.1287/orsc.1050.0116
Simon, H. A. (1957). Models of man; social and rational. In *Models of man; social and rational.* Wiley.
Simon, H. A. (1962). The Architecture of Complexity. *The Roots of Logistics*, *106*(6), 335–361. https://doi.org/10.1007/978-3-642-27922-5_23
Sledge, I. J., & Príncipe, J. C. (2018). An analysis of the value of information when exploring stochastic, discrete multi-armed bandits. *Entropy*, *20*(3), 1–35. https://doi.org/10.3390/e20030155
Slivkins, A. (2019). Introduction to multi-armed bandits. *Foundations and Trends in Machine Learning*, *12*(1–2), 1–286. https://doi.org/10.1561/2200000068
Sobel, M. E. (1982). Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models. *Sociological Methodology*, *13*, 290–312. https://doi.org/10.2307/270723
Sutton, S. R., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT Press. https://books.google.at/books?id=CAFR6IBF4xYC
Sutton, S. R., & Barto, G. B. (2017). Reinforcement Learning: An Introduction, by Sutton, R.S. and Barto, A.G. In *The MIT Press* (Second edi, Vol. 3, Issue 9). https://doi.org/10.1016/s1364-6613(99)01331-5
Weber, R. (1992). On the Gittins Index for Multiarmed Bandits. *Ann. Appl. Probab.*, *2*(4), 1024–1033. https://doi.org/10.1214/aoap/1177005588
Weitzman, M. L. (1979). Optimal Search for the Best Alternative. *Econometrica*, *47*(3), 641. https://doi.org/10.2307/1910412
Whittle, P. (1988). Restless Bandits: Activity Allocation in a Changing World. *Journal of Applied Probability*, *25*, 287–298. https://doi.org/10.2307/3214163

Appendix

List of Figures

You are about to play a simple multi-armed bandit game. This game consists of 5 different arms where every arm pays a different amount, but they all have the same mean-variance distribution. The payouts of the individual arms are redrawn after each round. The expected value and the variance of the 5 arms are different but always remain the same for the respective arms. In other words, any arm gives a different payout and this payout for each arm fluctuates within a certain range. So there is a best, second best, third best, fourth best, and fifth best arm.

The goal is to develop a strategy to maximize your payout over a period of 100 rounds.

In the first round you will be asked how convinced you are that you have chosen the best arm. This question serves only to illustrate your level of knowledge about the experiment. So think carefully like yours conviction, in the first rounds where you hardly have any information about the respective arms, can look like. This Question then follows every 5 rounds.

Before each round you can review the "Round History" table below to see your arms chosen so far and their winnings. There is no time limit and there are no wrong answers!

Deutsch:

Sie sind dabei ein einfaches mehrarmiges Banditen Spiel zu spielen. Dieses Spiel besteht aus 5 unterschiedlichen Armen, die jeweils einen unterschiedlichen Betrag auszahlen aber alle die gleiche Mittelwert-zu-Varianz Verteilung besitzen. Die Auszahlungen der einzelnen Arme werden nach jeder Runde neu gezogen. Der Erwartungswert und die Varianz der 5 Arme sind unterschiedlich bleiben aber für die jeweiligen Arme immer gleich. Mit anderen Worten, jeder Arm gibt einen unterschiedlichen Gewinn und dieser Gewinn schwankt innerhalb einer bestimmten Schwankungsbreite. Es gibt daher einen besten, zweitbesten, drittbesten, viertbesten und fünftbesten Arm.

Das Ziel ist es eine Strategie zu entwickeln um Ihre Auszahlung, in einem Zeitraum von 100 Runden, zu maximieren.

In der ersten Runde werden Sie gefragt wie überzeugt Sie sind den besten Arm gewählt zu haben. Diese Frage dient lediglich zur Veranschaulichung Ihres Wissenstandes über das Experiment. Überlegen Sie daher gut wie Ihre Überzeugung, in den ersten Runden wo Sie kaum Informationen über die jeweiligen Arme haben, aussehen kann. Diese Frage folgt anschließend alle 5 Runden.

Vor jeder Runde können Sie unterhalb die "Round History" Tabelle begutachten um Ihre bislang gewählten Arme und deren Gewinne zu sehen. Es gibt kein Zeitlimit und es gibt keine falschen Antworten!

What is your age?

What is your gender?

- Male
- Female

Next

Figure 20: Overview static environment

You are about to play a simple multi-armed bandit game. This game consists of 5 different arms where every arm pays a different amount, but they all have the same mean-variance distribution. The payouts of the individual arms are redrawn after each round. The expected value and the variance of the 5 arms are different but always remain the same for the respective arms. In other words, any arm gives a different payout and this payout for each arm fluctuates within a certain range. So there is a best, second best, third best, fourth best, and fifth best arm.

The goal is to develop a strategy to maximize your payout over a period of 100 rounds. Attention! There is a 50% chance that the payout for the chosen arm will change completely in the current round.

In the first round you will be asked how convinced you are that you have chosen the best arm. This question serves only to illustrate your level of knowledge about the experiment. So think carefully like yours conviction, in the first rounds where you hardly have any information about the respective arms, can look like. This Question then follows every 5 rounds.

Before each round you can review the "Round History" table below to see your arms chosen so far and their winnings. There is no time limit and there are no wrong answers!

Deutsch:

Sie sind dabei ein einfaches mehrarmiges Banditen Spiel zu spielen. Dieses Spiel besteht aus 5 unterschiedlichen Armen, die jeweils einen unterschiedlichen Betrag auszahlen aber alle die gleiche Mittelwert-zu-Varianz Verteilung besitzen. Die Auszahlungen der einzelnen Arme werden nach jeder Runde neu gezogen. Der Erwartungswert und die Varianz der 5 Arme sind unterschiedlich bleiben aber für die jeweiligen Arme immer gleich. Mit anderen Worten, jeder Arm gibt einen unterschiedlichen Gewinn und dieser Gewinn schwankt innerhalb einer bestimmten Schwankungsbreite. Es gibt daher einen besten, zweitbesten, drittbesten, viertbesten und fünftbesten Arm.

Das Ziel ist es eine Strategie zu entwickeln um Ihre Auszahlung, in einem Zeitraum von 100 Runden, zu maximieren. Achtung! Es besteht eine Wahrscheinlichkeit von 50%, dass sich die Auszahlung, für den gewählten Arm in dieser Runde, vollständig ändert.

In der ersten Runde werden Sie gefragt wie überzeugt Sie sind den besten Arm gewählt zu haben. Diese Frage dient lediglich zur Veranschaulichung Ihres Wissenstandes über das Experiment. Überlegen Sie daher gut wie Ihre Überzeugung, in den ersten Runden wo Sie kaum Informationen über die jeweiligen Arme haben, aussehen kann. Diese Frage folgt anschließend alle 5 Runden.

Vor jeder Runde können Sie unterhalb die "Round History" Tabelle begutachten um Ihre bislang gewählten Arme und deren Gewinne zu sehen. Es gibt kein Zeitlimit und es gibt keine falschen Antworten!

What is your age?

What is your gender?

- Male
- Female

Next

Figure 21: Overview dynamic environment

Your choice in round number: 1 out of 100 rounds.

Ihre Wahl in Runde: 1 von insgesamt 100 Runden.

Arm A Arm B Arm C Arm D Arm E

Round history

Running total payoff: **$0.00**

Round	Choice	Payoff	Belief

Debug info

vars_for_template

previous_decision	[]
total_payoff	Currency(0.00)

Basic info

ID in group	1
Group	1201
Round number	1
Participant	P1
Participant label	
Session code	to03odmq

Figure 22: Second page overview

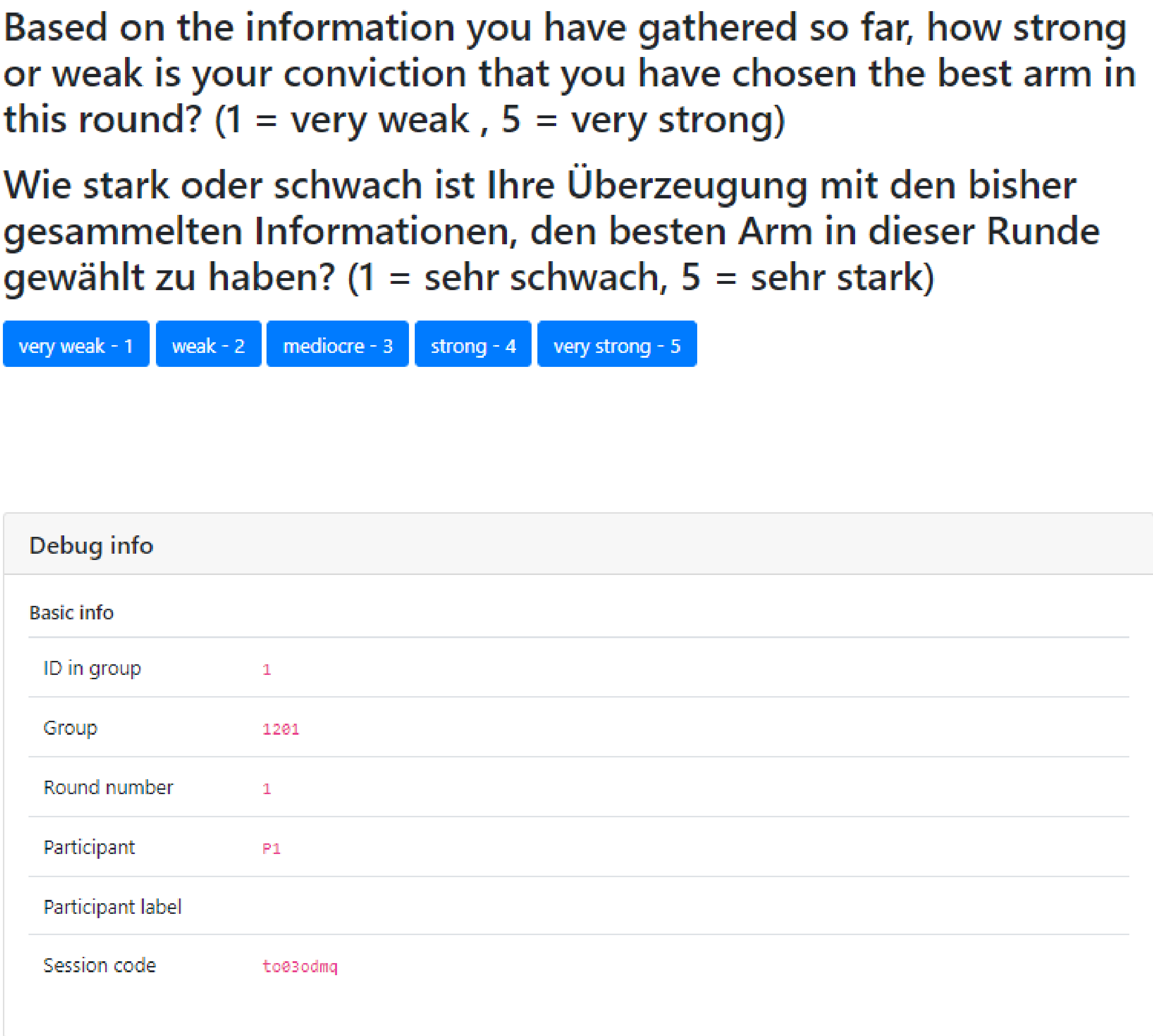

Figure 23: Beliefs page overview

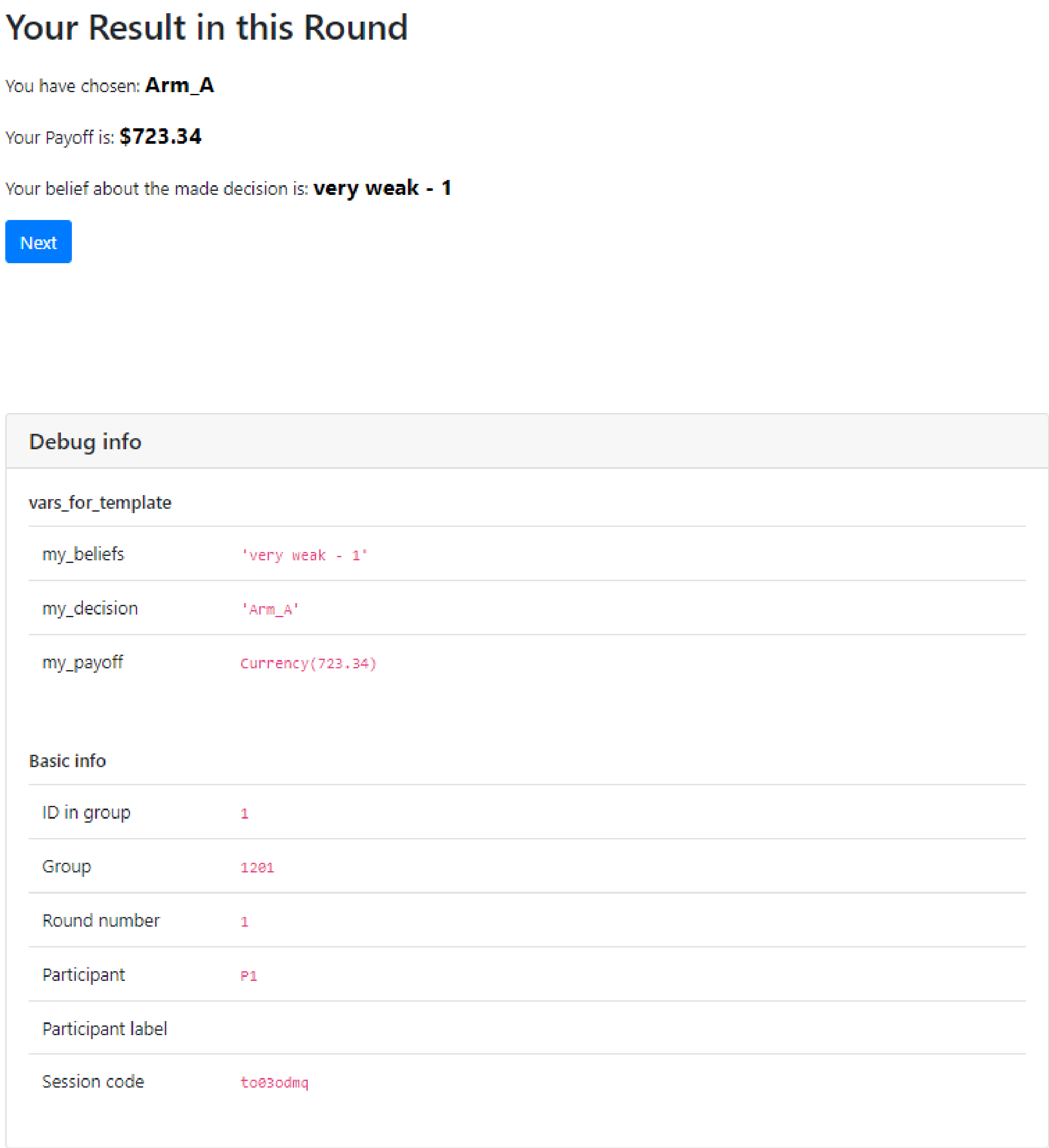

Figure 24: Result page overview

Your choice in round number: 11 out of 100 rounds.

Ihre Wahl in Runde: 11 von insgesamt 100 Runden.

Arm A Arm B Arm C Arm D Arm E

Round history

Running total payoff: **$10049.88**

Round	Choice	Payoff	Belief
1	Arm_A	$723.34	very weak - 1
2	Arm_C	$1310.95	None
3	Arm_B	$1427.61	None
4	Arm_D	$789.80	None
5	Arm_E	$1087.90	very weak - 1
6	Arm_A	$838.67	None
7	Arm_B	$1179.79	None
8	Arm_C	$1110.64	None
9	Arm_D	$775.64	None
10	Arm_E	$805.54	weak - 2

Debug info

vars_for_template

previous_decision	`[, , , , , , , , , ]`
total_payoff	`Currency(10049.88)`

Figure 25: Second page running total payoff and round history table

Zusammenfassung

Der Grundstein dieser Arbeit liegt in der Tatsache, dass Menschen täglich sowohl bewusste als auch unbewusste Entscheidungen treffen müssen. Daher rührt der starke Drang sämtliche Aspekte, die zur Entscheidungsfindung führen zu durchleuchten. Einer der Grundpfeiler im Entscheidungsprozess ist das Suchverhalten. Darunter versteht man wie das Verhältnis zwischen der Suche nach alternativen Möglichkeiten und der Verwertung bereits vorhandener Informationen aussieht. Das Ziel dieser Arbeit ist, einen möglichen Unterschied in zwei verschiedenen Umgebungen im besagten Suchverhalten aufzudecken. Auf der einen Seite steht eine stark vereinfachte Umwelt die sogenannte statische Umgebung, bei der die Belohnung für jede Entscheidung immer im selben Korridor angesiedelt ist. Auf der anderen Seite steht ein realistischerer Ansatz der Umwelt, die sogenannte dynamische Umgebung, bei der sich die Belohnung jeder Entscheidung mit einer gewissen Wahrscheinlichkeit vollständig ändern kann. Man geht davon aus, dass die dynamische Umgebung das Suchverhalten dahingegen beeinflusst, sodass es zu einer Erhöhung in der Suche nach neuen Alternativen kommt. In enger Verbindung mit einer Erhöhung des Suchverhaltens steht gleichzeitig auch eine Senkung der Überzeugung wie gut eine Entscheidung war im Raum. Das Herzstück dieser Arbeit ist ein mehrarmiges Banditen-Experiment für jeweils eine der Umgebungen. Darin wurden die Teilnehmer aufgefordert, innerhalb einer vorgeschriebenen Anzahl an Runden, in jeder Runde unter fünf verschiedenen Entscheidungen zu wählen. Ziel für die Teilnehmer war es die größtmögliche Summe an Geld, innerhalb von 100 Runden, zu sammeln. Nach Auswertung der Daten zeigte sich, dass die Teilnehmer im dynamischen Umfeld mit einer signifikant höheren Wahrscheinlichkeit ein erforschendes Verhalten an den Tag legten und gleichzeitig auch die Überzeugung gegenüber den einzelnen Armen abgeschwächt war. Abschließende Diskussionen beziehen sich auf mögliche Begründungen warum die Teilnehmer ihr Verhalten verändert haben.