Essays on Preference Formation

Inauguraldissertation

zur Erlangung des Grades eines Doktors
der Wirtschaftswissenschaften

durch

die Rechts- und Staatswissenschaftliche Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Cavit Görkem Destan

aus Salihli, Türkei

2023

Dekan:	Prof. Dr. Jürgen von Hagen
Erstreferent:	Prof. Dr. Florian Zimmermann
Zweitreferent:	Prof. Dr. Thomas Dohmen
Tag der mündlichen Prüfung:	7. August 2023

Acknowledgements

I would like to express my deepest gratitude to my advisors, Florian Zimmermann and Thomas Dohmen, for their continuous support and guidance throughout my Ph.D. journey. Florian Zimmermann has been a great mentor since the beginning of my studies, always reminding me of the scientific discipline and aiming for the best. Thomas Dohmen has been a wonderful co-author and a source of inspiration, supporting me in every aspect and never letting me down. I am truly fortunate to have such brilliant and dedicated advisors.

I am also very grateful to the great team of researchers at the University of Bonn Economics Department, who have contributed to my academic development and provided me with valuable feedback and suggestions. In particular, I would like to thank Sebastian Kube for his help whenever I needed it, Armin Falk and Lorenz Götte for their insights and encouragement, Holger Gerhardt for his extensive support and always keeping high standards, and Stephan Lauermann for his helpful suggestions. I am also grateful for my fellow Ph.D. students for sharing the journey with me.

I would like to acknowledge the supportive team at the Bonn Graduate School of Economics and the Institute for Applied Microeconomics, especially Ilona Krupp, Simone Jost and Ozlem Icpinar, for their administrative assistance and never-ending support. I also thank all research assistants for their support.

I would also like to thank my hosts at Boston, USA: Lucas Coffman, Frank Schilbach and Mehmet Ekmekci, for their invaluable feedback and making it possible for me to be a part of the academic community in the new world. I learned a lot from them and enjoyed my time there.

This dissertation would not have been possible without the support of Bonn Graduate School of Economics, Collaborative Research Center Transregio 224, and the Institute for Applied Microeconomics at the University of Bonn. I also acknowledge the financial support of the DAAD.

Last but not least, I would like to thank my family and friends for their love and support throughout this challenging journey. My mother, Hülya Destan, and my father, Zafer Destan, have always been there for me and believed in me. I dedicate this dissertation to them. I would like to thank Mustafa Kaba for being a family in Bonn and making research even more fun for me.

Most importantly, I am indebted to my dear friend, colleague, co-author, wife, inspiration, and everything, Günnur Ege Bilgin. Thank you for turning my life into heaven.

Cavit Görkem Destan
Bonn, 15. May 2023

Contents

List of Figures

List of Tables

Introduction

Preferences, as the back bone of economic theory, have been studied extensively. Almost all studies in the field analyze the economic consequences of different individual preferences in various settings. Yet, the detailed investigation of the formation of preferences is a relatively recent endeavour. For instance, risk aversion and other non-classical preferences are known for a long time (e.g. Kahneman and Tversky (1979)). Nonetheless, the relation between cognitive ability, patience, and risk preferences was studied decades later by Dohmen et al. (2010). Around the same time, I remember taking the Econ 101 course as a freshman and the professor's claim about economists' disinterest in the sources of individual preferences. He stated that in economics, we take preferences as given and do not question why someone likes apples or oranges. However, many economists were apparently intrigued by the question of how the preferences emerge. Thus, the research on preference formation has been accelerated in the last decade.

In addition to the biological or other personal factors shaping preferences, each decision environment also has an effect on individual choices. Many studies show that instead of having fixed preferences, a person might choose drastically distinct alternatives in slightly different settings. For example, Tversky and Kahneman (1974) demonstrated the effect of anchoring and Tversky and Kahneman (1981) showed the importance of framing, decades ago. More recently, cognitive factors such as attention, complexity, motivation, and memory are commonly used to explain the choice processes of people in various decisions.[1] With this dissertation, I aim to contribute to that literature by explaining the effect of decision environment on individual choices in three different contexts.

Chapter 1 consists of an analysis of the choices between starting an action and staying passive. Experimental results shows that people have an intrinsic tendency to become active even when their material payoffs are maximized by staying passive and any action is costly. This tendency is called *active participation bias* and its potential consequences are discussed in detail. In Chapter 2, a model of electoral competition is provided. The model explains how *salience bias* of the voters can cause extreme policies even when the voters have moderate preferences. Chapter

[1] Bordalo, Gennaioli, and Shleifer (2020), Oprea (2020), and Zimmermann (2020) are just a few examples.

also includes possible implications and survey evidence confirming the theoretical predictions. Finally, Chapter 3 presents an experimental investigation of learning patterns with state-dependent preferences. I show that individuals cannot differentiate the effect of the consumption state from the quality of the good if the state is not easily observed. This systematic error is called *attribution bias*, and I explain its underlying mechanism along with potential solutions.

References

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2020. "Memory, Attention, and Choice." *Quarterly Journal of Economics* 135 (3). https://doi.org/10.1093/qje/qjaa007. [1]

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2010. "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review* 100 (3): 1238–60. https://doi.org/10.1257/AER.100.3.1238. [1]

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47 (2): 263. https://doi.org/10.2307/1914185. [1]

Oprea, Ryan. 2020. "What Makes a Rule Complex?" *American Economic Review* 110 (12): 3913–51. https://doi.org/10.1257/AER.20191717. [1]

Tversky, Amos, and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185 (4157): 1124–31. https://doi.org/10.1126/SCIENCE.185.4157.1124. [1]

Tversky, Amos, and Daniel Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science* 211 (4481): 453–58. https://doi.org/10.1126/SCIENCE.7455683. [1]

Zimmermann, Florian. 2020. "The Dynamics of Motivated Beliefs." *American Economic Review* 110 (2): 337–61. https://doi.org/10.1257/AER.20180728. [1]

Chapter 1

Active Participation Bias[*]

Joint with Thomas Dohmen

1.1 Introduction

The human tendency to prefer activity over non-activity is a widely observed phenomenon. While activity might be optimal in many situations, we also observe individuals become active in many different contexts, even where inactivity would yield higher payoffs. A CEO, for example, might invest in a seemingly unpromising project, in the absence of alternative projects simply to avoid being seen as lazy, incompetent, or too risk-averse by shareholders. There could be various other motives such as social-image or self-image concerns, (false) beliefs, social preferences, risk preferences, or comprehension mistakes that contribute to harmful activity in different settings. Yet, the existence and prevalence of an intrinsic preference for activity that overrides the payoff-maximizing option of inactivity have not been rigorously examined or established.

Therefore, in this paper, we address three questions: Is becoming active the predominant choice in environments where the payoff-maximizing strategy is staying passive? If so, does active participation bias prevail, even if social motives, risk preferences, and mistakes in comprehension are controlled for? And finally, how prevalent is an intrinsic preference to become active even when the payoff-maximizing action is to stay passive?

To address these questions, we devised a series of experiments in which participants engage in a sequential search game in pairs. In these search games, the

[*] I am thankful to Florian Zimmermann and Armin Falk for many discussions, and to Sebastian Kube, Lucas Coffman, Frank Schilbach, Paul Heidhues, Thomas Graeber, and Luca Henkel for helpful comments. Funding by the German Research Foundation (DFG) through CRC TR 224 is gratefully acknowledged. This study is preregistered under Open Science Framework: https://osf.io/d2pwt . The ethical approval by German Association for Experimental Economic Research e.V.: No. I3jdLuqv

optimal strategy for one player is to remain passive. During the search phase, both players can independently draw offers by paying a fee. However, only their own draw is visible to them, while the search actions or outcomes of the other player remain unknown. At the end of the search phase, both players receive a payoff based on the higher of the two maximum draws of players, minus their own search costs. It was made clear to both players that they draw numbers from different uniform distributions, with one player (high-type) drawing from an interval shifted to the right on the number line compared to the interval of the other player (low-type). In the unique equilibrium of the game, high-type players are expected to search until they reach a number that exceeds a reservation value, while low-type players were expected to not search at all. However, low-types might be tempted to search for various possible reasons, such as equity concerns, social image concerns, (false) beliefs about the high-type partner's search behavior, curiosity, or comprehension errors. To eliminate these factors, we also conducted a treatment (computer treatment) in which human low-type players were paired with a computer bot that was programmed to search optimally as a high-type player, and low-type players are informed about that.

The choices of participants in the setting of our controlled experiments reveal three main results. Firstly, active participation bias is prevalent as almost all low-type players (97 percent) search at least once. Notably, a large majority are aware of the optimal strategy for the low-type players, as participants answer when asked about the optimal behavior of the low-type players that they should not search. Moreover, 95% of the low-type players who answer this question correctly searched nonetheless. This strongly suggests that participants intentionally chose to be active even if this is costly.

Second, active participation bias remains even when other factors are eliminated in the computer treatment. This intrinsic preference for active participation is evident for a large majority (85 percent) of participants. Compared to the 97% active search in the baseline, we can infer that the motives that may stem from human interaction, such as social image concerns or trust in other group members can only account for a 12 percentage points difference in the search activity while 85% are unexplained by social motives. Moreover, we collect measures of cognitive ability, creativity, personality traits, time and risk preferences. None of these character traits and preferences are associated with the active participation choice in any of the treatments. To further rule out that wrong beliefs, distrust, or experimenter demand effects drive these results, we designed a third treatment in which we offer low-type participants who are paired with a high-type computer bot a free outside option that guarantees a fixed amount higher than the expected payoff that could be obtained if the low-type would play his search subgame alone. In other words, even if the computer would not search at all or would not search optimally, the low type who accepts the free offer could not be better off by becoming active in the search game. Surprisingly, only 31% of the participants accept the outside option and the remaining 69% reject the option to play the game themselves. These results indicate that people derive an intrinsic utility from being active regardless of the outcome.[1] Thus, the tendency to be active seems to be a distinct character trait.

Third, when examining the behavior of these low-types who become active and engage in search, we observe another interesting pattern. Search behavior of low-types is consistent with the optimal search strategy that would apply if the search game was played in isolation, suggesting that they become narrow-minded, focusing on the search game they started but apparently not taking into account the entire structure of the game anymore, similar to narrow framing.[2] This finding is important as narrow framing bias can worsen the harmful consequences of active participation bias. For example, consider the example of a CEO who initiates a project even though it would have been better not to invest in it. Narrow framing might prevent her from abandoning it, even if better investment options emerge, leading her to keep investing more time and energy in the non-optimal project after it is initiated.

In order to test whether the interaction between active participation bias and narrow framing bias is systematic, we designed another set of experiments that follow the structure of our first set of experiments but alter the search game. In one experiment, we parameterized the search game for the low-type player such that the optimal stopping strategy in the isolated search game entails a higher reservation value, i.e. longer search, and therefore even further deviation from optimal non-activity. In another experiment, we altered the structure of the search games such that it would be optimal in the isolated search game of the low-type player to draw only once. We find strong evidence for low-type players to adhere to the optimal search strategy in the sub-game, which robustly corroborates that the harmful consequences of activity bias are moderated by narrow framing.

Moreover, we conducted a follow-up survey to show that active participation bias is a consistent behavior in various contexts and is also relevant in real-life settings. In the survey, participants are asked about active participation behavior in different domains such as investment, political engagement, education, time management, and medical choices. The answers of a person to different questions are correlated with each other and they predict their active search behavior in the main experiment, indicating the external validity of the lab results.

Active participation bias is first observed in a lab experiment by Lei, Noussair, and Plott (2001) in which participants make investments in the stock market too

[1] This treatment can also be seen as counter-evidence against experimenter demand effects since some people might believe that researchers expect participants to accept their offer. However, we do not see a considerable effect as the vast majority reject the offer.

[2] Barberis, Huang, and Thaler (2006) used the term narrow framing to describe that tendency to evaluate decisions in isolation without considering other related outcomes.

often, thereby generating bubbles. In a treatment condition with an alternative task available, bubbles are less likely to occur as trading activity is reduced to optimal levels since some of the individuals do not participate in the asset market. While Lei and coauthors coin the term *active participation bias* to explain the tendency to become active, they do not investigate the mechanism further. Since then, the active participation bias has been proposed as an explanation for several experimental results.[3] However, non of these studies rigorously scrutinized the various mechanism that might drive active participation. This is the first study that shows the prevalence of an innate preference to be active, apart from other biases and preferences, as a driver of active participation bias.

We believe that we can design work environments more efficiently on the grounds of the insights on active participation bias. For example, we can remind decision-makers of other related factors and suggest alternative strategies after they begin to engage in a task. Hence, vast and unnecessary costs can be avoided if managers and workers themselves are aware of this bias and act accordingly.

Similarly, we can design experiments more accurately by taking the pervasive tendency to be active into consideration. Otherwise, it would be impossible to distinguish a preference for a particular action from the general tendency to be active. For instance, one can reduce the noise and increase the precision of the experimental results by providing subjects with more than one way of becoming active. Alternatively, we can frame alternative options such that they do not allude to passivity.

In Section 1.2 of the paper, we summarize the related literature and explain connections to our research. In section 1.3, we discuss the potential mechanisms of the bias and describe our hypotheses. The experimental design is explained in Section 1.4 and the results are presented in Section 1.5. We demonstrate additional treatments and robustness checks in Section 1.6. Our conclusion is in Section 1.7.

1.2 Literature

Experimental settings are often subject to the influence of active participation bias, where participants have to take a specific action to avoid passively waiting. For example, in auction experiments, participants may have no alternative task available if they choose not to bid. This design feature can have nontrivial effects on participants' behavior in diverse settings, including contests, auctions, and many more. That is mainly because participants seem to deliberately avoid strategies that involve staying passive in these settings even if it is costly for them not to

[3] See Dechenaux, Kovenock, and Sheremeta (2015) for the detailed investigation of contests, all-pay auctions, and tournament experiments.

stay passive. This behavior induces an often observed pattern in which participants deviate from the equilibrium prediction, specifically when the optimal strategy involves staying passive, e.g., choosing not to bid, not to provide effort, or not to take action.

Lei, Noussair, and Plott (2001) found that non-optimal trading activity was frequent in their double-sided asset market experiments. In fact, they observed excess purchasing activity in all but one treatment condition. In this latter condition, participants could engage in an alternative task. When an alternative task was offered, many participants ceased trading in the asset market and solely focused on the alternative task. Consequently, they hypothesized that individuals engaged in trading merely because there were no other available activities. They called this behavior, for the first time, *active participation bias*.

We observe active participation bias in numerous other contexts. For example, in all-pay auction experiments, it is often observed that participants bid positive amounts even when the Nash equilibrium prediction is zero. Lugovskyy, Puzzello, and Tucker (2010) show that almost all of the bids are positive, which leads to overdissipation. The overdissipation is eventually eliminated only in the treatment where negative bids are possible. They argue that active participation bias is effective in this context since individuals do not need to bid positive amounts to feel active in the negative-bid setting.

Similarly, in contest experiments, participants exert costly effort even when the optimal strategy is zero effort. Sheremeta (2010) shows that 40% of subjects exerted costly effort in contests for a prize of zero. Moreover, in the group contest framework, Abbink et al. (2010) and Sheremeta (2011) show that almost all players exert positive effort where the theoretical prediction is zero.

Likewise, Goerg, Kube, and Radbruch (2019) conducted an experiment in which participants performed a real-effort task under different payment schemes. Across all treatment conditions, individuals exerted higher levels of effort if they were unable to leave the lab upon completion. Furthermore, when participants were prohibited from using the internet while waiting, their effort levels increased even more. This indicates that individuals exert costly effort to evade idleness.

Moreover, Carpenter, Liati, and Vickery (2010) show that many people send positive amounts in a two-way dictator game experiment in which two players simultaneously choose how much to donate to the other player out of their endowments. They argue that altruistic motives are not relevant since both players start with the same endowment, and participants donate because they are “ready to play”. They also show that impulsivity (measured by an ADHD questionnaire) is positively correlated with donating positive amounts.

Active participation bias is also prevalent in the war of attrition experiments. Under full information, the theoretical prediction in these settings is stopping immediately. Both Bilodeau, Childs, and Mestelman (2004), and Hörisch and

Kirchkamp (2010) show that stopping at time zero is rarely observed, contrary to standard theory. Individuals tend to avoid choosing null actions in general.

Furthermore, active participation bias may also apply to some behaviors observed outside the lab. For example, Bar-Eli et al. (2007) noted that elite goalkeepers in Israel's professional football league did not stay in the middle during penalty kicks as often as they optimally would have done, considering the probability of kickers shooting to the middle. The goalkeepers could have saved more penalties by increasing their probability of staying passive, but in-depth interviews revealed that they chose to jump to a corner because staying in the middle was perceived as incompetence by fans.

Another related phenomenon is the fear of missing out (FOMO). Recent studies have shown that people are more likely to engage in an activity when many others are engaged as well, and they feel lonely or depressed when they cannot join popular endeavors. For instance, Zeelenberg and Pieters (2004) found that using a postcode-style lottery, where everyone with a ticket in a postcode area wins if their postal code is drawn, increases participation drastically compared to a state-wide lottery. Likewise, Kuhn et al. (2011) demonstrated that neighbors of car winners in a lottery are more likely to buy a new car. Hence, people tend to take an action when it is prevalent in society.

Additionally, the abundant literature on the demand for agency can also be related to active participation bias. *Agency* or *authority* is generally defined as the ability to influence the outcome. Although individuals who exhibit active participation bias in the studies mentioned above may not always affect the outcome (e.g., by making low bids in auctions or putting in minimal effort in contests), they alter their individual payoffs. They may be willing to incur some cost to have a personal impact on the game.

Lastly, we can mention the literature on boredom as it can be one of the main variables in active participation decisions. The alternative to active participation is often waiting and it may provoke boredom. Killingsworth and Gilbert (2010) show that being idle causes unhappiness since people tend to think about what potential activities they could be involved in. Likewise, Wilson et al. (2014) conduct an experiment where subjects have to wait for 15 minutes in an empty room and can choose to give electric shocks to themselves. During the 15-minutes "just thinking" phase, 43% of the participants give themselves at least one electric shock.

All the findings mentioned above suggest that people are inclined to be active instead of staying passive in numerous domains. This study provides a common explanation that can clarify and connect the diverse evidence.

1.3 Possible Mechanisms

Several factors might drive people to take an action even when the payoff maximizing option is staying passive. Firstly, social image concerns can be at play. In situations in which multiple players interact, costly actions might be used to signal to others that one is willing to actively contribute to group members' payoff. Such kind of signaling is particularly relevant in situations in which free-riding or laziness would lead to inactivity.

A second potential channel is fairness concerns. If the reward of a group task is shared by everyone but some players do not pay any cost, that naturally causes some inequality among group members. The fairness literature documents (e.g., Forsythe et al. (1994)) that many people are willing to sacrifice some of their earnings to reduce the inequality in their group. Similarly, some individuals can engage in a costly activity to achieve a fairer distribution of payoffs net of costs. Naturally, fairness can also be a virtue that agents want to signal to other parties. That part of fairness concern can be seen as a form of the social image concerns described above.

Thirdly, beliefs and risk preferences can potentially lead to active participation bias. One player might understand that she must stay passive in the equilibrium play. However, she might act differently if she believes that her partner would make a mistake and not follow his optimal strategy. Thus, active participation can arise from insurance motives.

Fourthly, active participation could result from mistakes in decision-making or in implementation. Since negative activity is not possible in many games by definition, we would observe any error as active participation. For example, in most auction or trade experiments, you can either bid zero or a positive amount. In those cases, even small incidents such as clicking a button would be an indication of activity. Therefore, one-sided errors can be perceived as an overall tendency to be active.

Lastly, intrinsic utility from playing the game can be a source of active participation. Participants may be aware of the optimal strategies but still, deliberately play actively for pleasure or fun. Similarly, curiosity can cause them to take costly action to discover the potential outcomes of their actions.

Since active participation bias can come about through one or several of these mechanisms, we conduct an experiment that allows us to control these factors. Lab experiments provide an ideal setting to investigate whether active participation bias arises due to intrinsic utility from being active as they make it possible to eliminate the influence of other factors that potentially induce activity. In our experiment, we rule out the influence of some factors by design and observe how active participation is affected.

1.3.1 Hypotheses

We test two hypotheses using the data from our experiment. The first one is about the existence of active participation bias resulting from an intrinsic motive for being active. We conjecture that people have a tendency to become active instead of doing nothing even when the payoff-maximizing action is staying passive.

The second hypothesis is about one mechanism that drives active participation. We conjecture that social image and fairness concerns are other important drivers of active participation. We will investigate whether people take costly actions to signal their merit to other players and to reduce inequality within their group. More formally:

H1: People choose to be active even when it is costly and does not increase payoffs.

H2: Fairness and social image concerns increase the tendency to become active.

1.4 Experimental Design

The experiment was programmed in oTree (see Chen, Schonger, and Wickens (2016)) and conducted online during a Zoom meeting. We used an online procedure instead of an offline lab procedure because of COVID-19 restrictions at the time of implementation. In order to mimic a setting as similar to a lab experiment as possible, and to ensure data quality, we used the Zoom procedure. Li et al. (2021) show that using Zoom for online experiments improves data quality at the lab level and generates comparable data.

Procedural rules of the experiment: Subjects were invited to a Zoom meeting and they entered the waiting room upon arrival. Participants were then sequentially and individually admitted to the main room of the meeting, where they received instructions to keep their cameras turned on during the whole experiment and were briefly informed about procedures during the experiment on Zoom before they were moved to individual breakout rooms. Hence, participants never met each other in the Zoom meeting. After all participants had been transferred to their personal waiting rooms, the link to the experiment was shared and the experiment began. Being in separate breakout rooms mimics the cubicles in a lab while preserving anonymity among participants. Participants could use the "ask for help" button when they had a question or a problem.

Main Search Task: In our search task, two players are randomly matched to form a group. Each of the two group members separately plays a sequential search game in which each group member can search for a number that is uniformly distributed on a closed interval. Searching entails drawing a number from a known distribution by paying a fee of 1 point. After learning the result of the draw, she can either draw another offer by paying the search cost again or she can stop searching. The search phase lasts for 10 minutes and each draw takes 10 seconds. During this phase, each player can draw as many offers as she wants and as time permits. It is also possible not to draw any offers. However, it is not possible to end the search phase earlier. Players would have to wait for the search phase to finish when they stop searching.

At the end of the search phase, the highest offer each player received determines their individual performance. This value is announced to the other group member and the highest individual performance in a group determines group performance. The payoff for each group member equals the group performance, i.e., the highest number drawn by any of the two group members, minus the individual search costs. In other words, the highest offer in a group is treated as a public good whereas search costs are paid individually.

Each group consists of one high-type and one low-type player. The offers of high-type players are drawn uniformly from the interval $[35, 100]$ and the offers of low-type players are drawn uniformly from $[0, 65]$. In equilibrium, the high-type player searches until reaching the threshold of 89 and the low-type player does not search at all. The proof of the equilibrium and the theoretical analysis under different utility structures are provided in Section 1.A.1 in Appendix. Our main focus will be on low-type players since their optimal choice is to stay inactive but they can choose to engage in the search task to become active. That behavior reveals active participation bias.

We used a version of a sequential group search game in order to assess active participation bias and its underlying drivers because this game has three major advantages for testing our hypotheses. Firstly, the task does not preclude factors that have been conjectured to explain active participation when staying passive is optimal. For example, the group setting could trigger social motives, such as social image or equity concerns. Also, the potential risk of the partner's performance and the random nature of a draw makes risk preferences relevant.

Secondly, the rich set of features of the search games enables us to modify them to adapt to our testing purposes. For instance, we can change the group and information structure or the search parameters in order to test the effect of different factors.

Lastly, we know that individuals are generally capable of understanding and optimally playing search games. Many studies such as Schotter and Braunstein (1981) and Hey (1982) showed decades ago that the search behavior of the participants is usually close to the optimal strategy. Hence, search games are a natural starting point to see the behavioral effects of some modifications to the game.

Timeline and Other Tasks: The timeline of the experiment is as follows.

(1) Instructions and comprehension questions
(2) Demo task
(3) Main task: group search game
(4) Questionnaire about the search game
(5) The revelation of search outcomes in the group
(6) Survey on demographics and preferences
(7) Raven's matrices
(8) Remote associates test

Chronological sequence After having read the instructions and correctly having answered the comprehension questions about the game structure, participants played a demo search game. The setup of the demo was exactly like the main task and each participant drew one offer.[4] We used this demo task for two purposes. Firstly, players can get familiarized with the task. Secondly, and more importantly, we eliminate the curiosity motive with the demo draw. Hence, we can induce that players do not draw offers in the main task just to see the game flow.

We also implemented an attention check mechanism to ensure that participants did not engage in other activities during the search phase. Participants were informed that a pop-up window would appear on their screen at a certain point in time and that they would have to click a button within a 5 seconds time frame in order to confirm that they were paying attention.[5] If they failed to click the button, they lost all the earnings from the task and only received the participation fee of 3 Euros.

After the search phase and before the revelation of group outcomes, we asked participants a few questions about the main task to better understand their strategy and decision rules. These were not incentivized. In particular, we asked them to describe the decision rule they followed in the task, and then, we asked them to state their beliefs about the highest offer of their partners.

Later, they filled out an extensive survey including demographic questions as well as subjective risk, patience, general trust, math ability, and other personality traits. Subsequently, they could solve a maximum of 8 Raven's matrices with visual puzzles in 2 minutes. We use this measure as a proxy for cognitive ability or IQ. Finally, they solved a 10-question remote associates test in 3 minutes. The score on this test is used as an indicator of creativity. We used a between-subject design

[4] The offer they draw is fixed to their expected offers (68 for high-types and 33 for low-types) to avoid any anchoring.

[5] The pop-up screen was set to appear just 5 seconds before the end of the search phase in order to avoid any impact on the search behavior afterward, but the exact timing was not revealed to participants in the instructions.

such that participants are randomly placed in one of the two treatments described below.

1.4.1 Social (Baseline) Treatment

In this treatment, participants play the search game as described above. In this setting, we can expect all the possible channels mentioned above to be effective. For instance, low-type players may try to get a high offer to signal to their partners that they are hard-working, capable, or fair. Alternatively, they may search to insure themselves against the case their partner makes a mistake and not search at all.

1.4.2 Computer Treatment

In this treatment, high-type players are replaced with a computer algorithm that plays an optimal strategy with a threshold of 90. Low-type players are informed that a computer bot undertakes the role of the high type and it plays the optimal strategy. However, the explicit strategy of the computer is not described.[6]

In this setting, we would expect all of the social, fairness, and risk concerns to be irrelevant. Low-type players do not have any motive to signal something, reduce inequality, or insure themselves against possible mistakes. Hence, by comparing the behavior in this treatment to the baseline, we can clearly see the effect of all the other factors combined.

1.4.3 Procedural Details

The experiment was conducted with the participants from the BonnEconLab subject pool. A total of 217 people participated and 80 of them were in the computer treatment. Out of 137 people in the social treatment, 69 people were low-type players and 68 people were high-type players.[7] The average duration of the experiment was 45 minutes and the average payment was 16.4 Euros.

1.5 Experimental Results

The experiment is conducted with the participants from the BonnEconLab subject pool. A total of 217 people participated and 80 of them were in the computer treatment. Out of 137 people in the social treatment, 69 people were low-type

[6] To make the computer bot comparable to human players, we imposed the same time limit (10 min.) and draw time (10 sec.). Hence, the computer can draw up to 60 offers and this is announced to participants. This constraint was never binding in our data (see Section 1.A.1 in Appendix).

[7] A few people had connection issues and left the experiment during the sessions. In those cases, their partners continued. Hence, the number of types does not match.

players and 68 people were high-type players[8]. The average duration of the experiment was 45 minutes and the average payment was 16.4 Euros.

The first observation is that our attention manipulation is successful. Only 14 participants (8 in social, 6 in computer treatment) failed the attention check in the search phase. We can deduce that almost all of the participants have been waiting for the pop-up screen even when their search activity is finished. We discard participants who failed the attention check from the analysis since their behavior is not comparable to the behavior of other participants[9].

The first main result of the experiment is that the tendency to become active is almost universal. Nearly all players searched actively in the baseline treatment such that only 2 low-type players out of 67 (3%) have not drawn any offer in the social treatment. Although it is payoff maximizing to stay passive for low-type players, almost all of them chose to draw at least one offer to participate the task actively. As Figure 1.1 shows, most of the low-type players drew more than one offer, and the average number of draws of searchers was 4.1.

Another important remark is that participants do not keep drawing offers indefinitely. Since each draw takes around 10 seconds, the total duration of the search activity is approximately one minute for most of the participants. Hence, the waiting time for searchers (≈9 min.) is not very different from the waiting time of non-searchers (10 min.).

The second main result is that active participation bias cannot be explained by risk preferences, fairness, or social motives. Only 14.86% of participants in the computer treatment have not drawn any offer. Despite the 11.86 percentage point increase from the 3% not searching in the social treatment, 85% of the low-type players still chose to search even when their partner is a computer algorithm that plays the optimal strategy. Thus, all of the human factors such as virtue signaling, altruism, or trust can only account for a small fraction of active participation behavior. Similar to the social treatment, most of the participants drew more than one offer and the average number of draws was 3.3.

Moreover, active participation bias is not explained by a mistake caused by the inability to solve the problem correctly, as answers to the bonus question at the end of the main task reveal. The question asks participants about the optimal strategy for the low-type players and 62.4% of the participants correctly indicate that it is to stay passive[10]. Nonetheless, 86.4% of the participants with the correct answer drew at least one offer even though they know the optimal strategy (see Table 1.A.4 in Appendix for details). Hence, we can deduce that the tendency to be active is not a simple error but a deliberate choice for a majority of participants.

[8] A few people had connection issues and left the experiment during the sessions. In those cases, their partners continued. Hence, the number of types does not match.

[9] This selection criterion was also included in the preregistration.

[10] There is no difference between the two treatments (see Table 1.A.3 in Appendix).

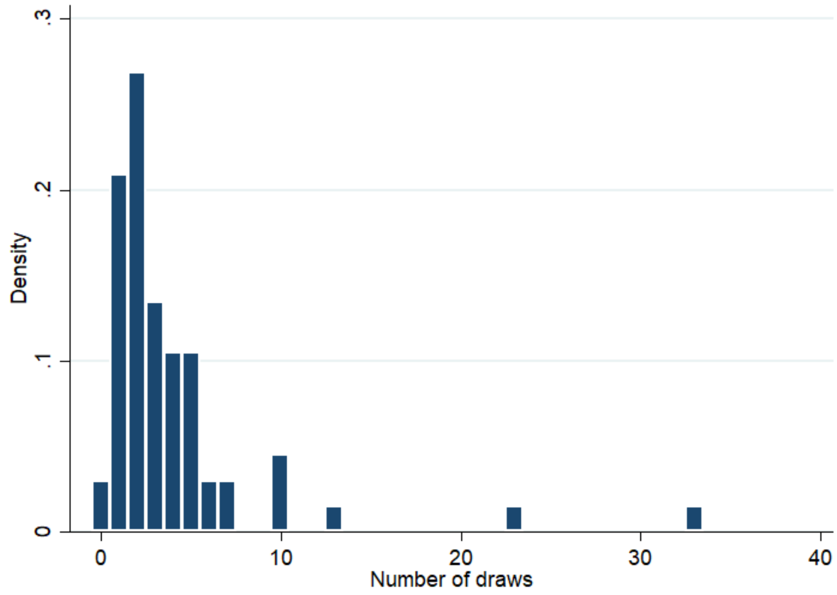

Figure 1.1. Histogram of number of draws by low-type players in the social treatment.

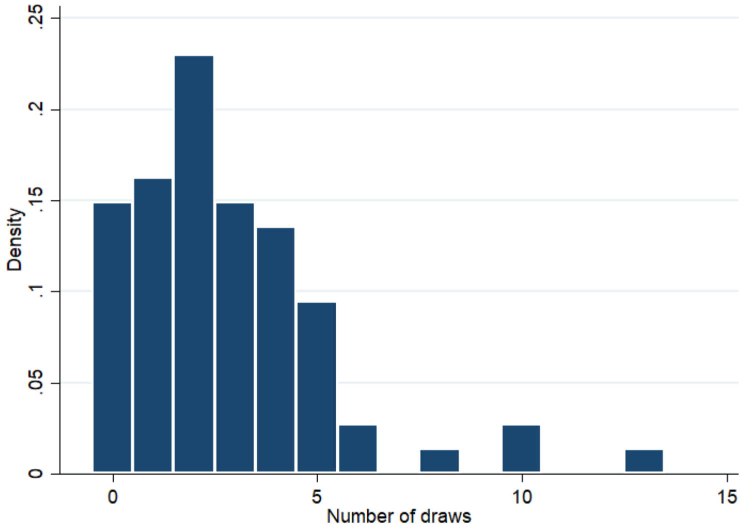

Figure 1.2. Histogram of number of draws by low-type players in the computer treatment.

We can also rule out that participation is driven by risk-taking behavior. We mentioned beliefs and risk preferences as potential mechanisms in Section 1.3. However, low-type players' beliefs about the high-types are too high to explain their search behavior. The average beliefs are 76.7 and 77.1 in the social and the computer treatments, respectively[11], which are much higher than the maximum possible offer for low-types. Moreover, as Table 1.1 shows, neither beliefs about the partner's performance nor risk preferences have a significant effect on search.

Table 1.1. OLS regressions of the indicator for not searching on belief and risk aversion measures.

	(1)	(2)
	nosearch	nosearch
computer treatment	0.119**	0.120**
	(0.0482)	(0.0478)
belief	-0.00113	
	(0.00226)	
risk aversion		0.0174
		(0.0105)
constant	0.116	-0.0572
	(0.177)	(0.0630)
N	141	141

Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Furthermore, the choice of active participation is also independent of other skill and preference measures. As Table 1.2 shows, Raven's test score, remote associates test score, self-reported math ability, general trust in strangers, and patience have no significant effect on search behavior. In addition, the other demographic and personality variables are not correlated with active search, either[12]. Therefore, we can confirm that the active participation bias is not a by-product of other preferences or biases and results from an intrinsic tendency to be active.

The third and final result of the experiment is about the impact of the active participation bias. If people choose to be active but stop immediately after a brief activity, that would not have huge consequences. However, we see the opposite phenomenon. Low-type players try to have a good offer and keep searching until they reach a satisfactory level. In other words, they ignore the broader game and focus on their own task as if they are playing alone.

We can understand their search strategy by looking at a few indicators. Firstly, participants generally reject small offers and only accept high levels. We can investigate their strategy more precisely using a probabilistic regression to find the threshold level that fits the data best. As exhibited in Table 1.3, we estimate the threshold of 52.34 for the social treatment and 53.15 for the computer treatment. Note that the optimal threshold for low-type players, if they play alone, would

[11] There is not a significant treatment difference in beliefs.

[12] Only neuroticism (i.e., emotional stability) seems to have a significant impact on search decisions but this effect vanishes in other treatments. See Appendix for a detailed analysis.

Table 1.2. OLS regressions of the indicator for not searching on the skill and preference measures.

	(1)	(2)	(3)	(4)	(5)
	nosearch	nosearch	nosearch	nosearch	nosearch
computer treatment	0.120**	0.118**	0.120**	0.117**	0.114**
	(0.0499)	(0.0482)	(0.0481)	(0.0484)	(0.0487)
Raven's score	-0.00171				
	(0.0186)				
rat score		0.00999			
		(0.0158)			
math ability			0.00763		
			(0.00860)		
trust				0.00436	
				(0.00918)	
patience					0.00859
					(0.0126)
constant	0.0371	0.0175	-0.0131	0.0106	-0.0324
	(0.0858)	(0.0401)	(0.0597)	(0.0534)	(0.0977)
N	141	141	141	141	141

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

be 54. Thus, we can easily conclude that those who become active follow a strategy similar to the optimal strategy of the search task in which they search alone. Furthermore, participants exhibit similar search behaviors in both treatments, as you can clearly see in Figure 1.3. The distributions of accepted offers are almost identical in the two treatments.

In addition, we can look at the recall behavior as an indication of a threshold strategy. Since a threshold strategy requires searching until a certain value, it never generates a stop after a low offer and a recall of a higher past offer. When we look at the 128 low-type players who drew at least one offer, only 27.3% of them recalled a past offer[13]. The behavior of the remaining 72.7% is in line with a threshold strategy.

Furthermore, the results from the question about the decision rule they followed in the search task also support the finding that they try to optimize their individual search task. Overall, 76% of the low-type players mentioned that they tried to maximize their own outcome and kept searching if the expected net return from a draw is positive[14]. On top of that, only 7% mentioned that they searched just for fun. These results provide additional evidence for focusing on the individ-

[13] No significant treatment difference, see Table 1.A.8 in Appendix.

[14] Two research assistants digitized the verbal answers separately. We combined their data such that an answer mentions a certain decision rule if at least one of the two assistants indicates so.

Table 1.3. Probit regression to estimate the threshold levels used in two treatments.

	(1)	(2)
	social	computer
threshold	52.34^{***}	53.15^{***}
	(0.14)	(0.14)
N	265	207

Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

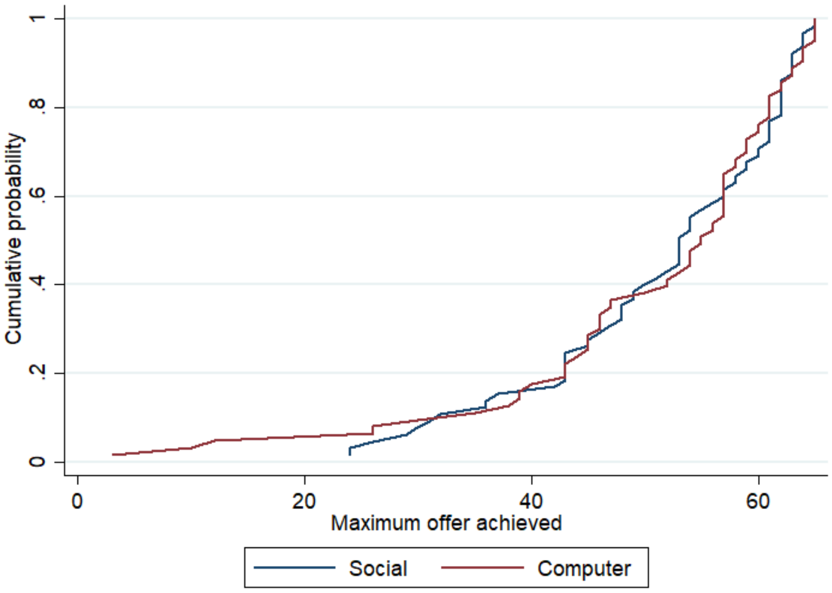

Figure 1.3. Cumulative distribution functions of achieved maximum offers in two treatments.

ual task and optimizing only that. Low-type players ignore the broad game and play as if they are alone.

Low-type players try to play their task as well as possible, despite the fact that each additional draw reduces their payoffs further since it is costly and their performance would be overridden by the high-type player. This behavior can be described as narrow framing, as in Barberis, Huang, and Thaler (2006), such that individuals break the decision process into different parts and try to solve them separately. The interaction between active participation bias and narrow framing has not been demonstrated before. According to our findings, the behavior is like a 2-step process. Agents first decide whether they start searching or not. Then, they choose how to play the game. The tendency to be active drives people into an activity and then, narrow framing causes them to focus on the confined task once they become active. Hence, active participation bias is problematic when it lures people into a situation in which they should not be. Also, its initial impact gets much worse if solving the problem in the new situation leads to an even stronger departure from the optimum in the overall game. We run three additional treatments and a follow-up survey to better understand this pattern. We explain the designs and their results in the next section.

1.6 Additional Treatments and the Follow-up Survey

We designed three additional treatments to test the extent of active participation bias and its interaction with narrow framing. Specifically, we provide additional evidence as a robustness check of the two-step activity decision. Particularly, we investigate whether people have an intrinsic preference to play the game themselves (i.e., the source of active participation bias) and whether they keep narrowly focusing on their task under different settings (i.e., narrow framing). All three new treatments are based on the computer treatment with small modifications.

We also conducted a follow-up survey with the participants from the experiment to provide evidence for the external validity of our lab experiment. We simply test whether their search behavior in the experiment is associated with their decisions in the real life.

1.6.1 High Intensity Treatment

In this treatment, we check if people keep searching further when their task is slightly changed. The treatment is exactly like the computer treatment except for the distribution of offers. The intervals for the draws are changed to $[0,85]$ for low-type players and $[15,100]$ for high-type players[15]. In this case, the equilibrium remains similar to the computer treatment. In equilibrium, high-types search until 88, and low-types do not search. On the other hand, the optimal threshold for low-type players becomes 73 if they search as if they are alone. Therefore, by comparing the search behavior to the computer treatment (where the optimal threshold in the alone task was 54), we can test whether participants ignore their partner and try to play optimally alone in their own task.

The data confirms again that individuals choose to be active since the vast 92% of the participants drew some offers. Furthermore, the results from the high intensity treatment clearly show that low-type players ignore their partners and search as if they play alone. As you can see in Figure 1.4, most of the players stopped only after achieving a high enough offer. The estimated threshold using probabilistic regression is 73.52 (s.d.=0.13) which is extremely close to the threshold of 73 (depicted by a dashed red line in the graph), the optimal level in

[15] Remember that they were $[0,65]$ and $[35,100]$, respectively.

the alone search task. Hence, the search activity is intensified when the optimal strategy of the (non-optimal) alone subgame required so.

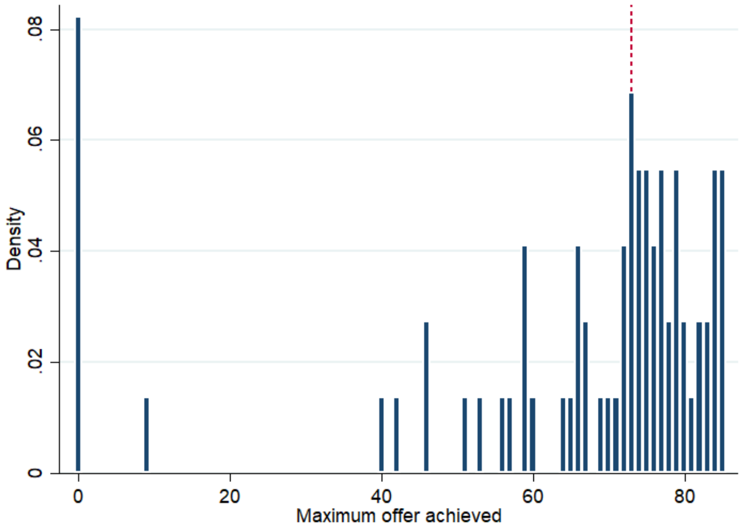

Figure 1.4. The histogram of maximum offers achieved in the high intensity treatment. The dashed red line represents the optimal threshold level of 73 in the alone search game.

1.6.2 Minmax Treatment

In this treatment, we test whether search duration reduces with a slight modification of the game. Particularly, the payoff structure is changed such that *the minimum* offer of each player determines their individual performance and *the maximum* of the individual performances determines the group performance (hence the name minmax). Therefore, the equilibrium becomes high-type players drawing only once and low-type players not drawing. Since the expected draw of a high-type is 67.5 (the mean of $U[35, 100]$), it is not optimal for a low-type to draw (from $U[0, 65]$) because they cannot improve the expected group reward. However, low-type players would also draw once if they narrowly focus on their own tasks and try to maximize their individual performance. Thus, by looking at their search behavior we can understand the scope of the narrow framing.

The findings in this treatment also support the active participation bias and narrow framing. As depicted in Figure 1.5, only one participant stayed passive (indicating active participation bias) and almost half of the participants drew only one offer (indicating narrow framing). Evidently, participants chose to be active and followed the optimal strategy of the alone search task once they begin to play.

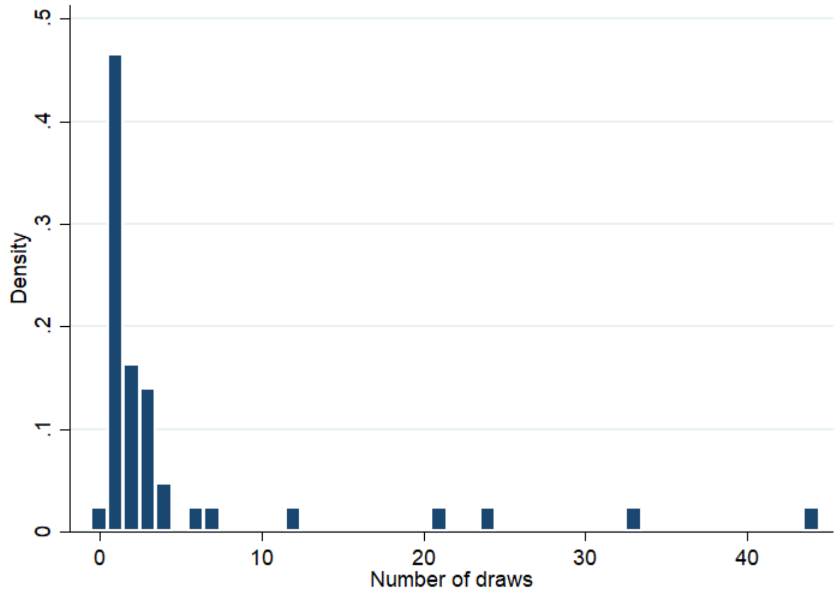

Figure 1.5. Histogram of the number of draws in the minmax treatment.

1.6.3 Option Treatment

In this treatment, we test whether participants have an intrinsic preference to play the game themselves or they participate only for a certain outcome. We do that be offering participants an outside option just before the search game begins. If they accept the outside option, their individual performance will be set as 55 with a total cost of 0. Then, they will wait for 10 minutes for the search phase to end. If they reject the outside option, they will play the search game as in the computer treatment.

Level 55 is chosen to exceed the expected value of playing the search game alone for the low-types. In other words, a low-type player cannot aim for an expected outcome above 55 by playing herself. Hence, we can deduce that they get an intrinsic utility from playing the game themselves if they reject the option and sacrifice some payoff by paying the search costs. Thus, we can clearly see the source of the active participation bias. Participants do not search only for the outcome but they also prefer to engage in the process of search.

On the other hand, this treatment can also be seen as a test of the experimenter demand effect. It is plausible to think that subjects might believe that researchers expect them to accept the offer if they present it. Hence, they might choose the outside option because they feel obliged to.

The result of the experiment supports our previous finding that people have an intrinsic preference to be active. Only 31% of the participants accepted the outside option whereas the remaining 69% rejected the option to search by themselves. Even under the strongest conditions (i.e., the generous outside option and the potential experimenter demand effect), the active participation bias is still prevalent. A vast majority of low-types chose to search actively.

When we look at the characteristics of those who rejected the option, we see again that active participation bias is independent of other individual traits such as risk preferences, cognitive abilities, or other preferences and skills. The regression results show that none of the ability and preference measures is correlated with the choice of accepting the outside option (see Appendix). Additionally, we can also induce that it is a deliberate choice and not only a mistake caused by some misunderstanding. Half of the participants who answered the bonus question correctly rejected the outside option[16]. Hence, their choice reveals that they are willing to pay the search costs in order to be active.

Lastly, we can look at the search behavior of the participants who rejected the outside option. The estimated threshold for searchers is 53.2 which is indistinguishable from the computer treatment or the optimal alone threshold. Furthermore, as shown in Figure 1.6, many people stopped searching with a maximum offer much below the outside option offer of 55. Thus, we can see that they do not reject the outside option to achieve an outcome higher than 55[17]. Instead, they prefer to search actively even if they pay the search costs and attain a worse offer.

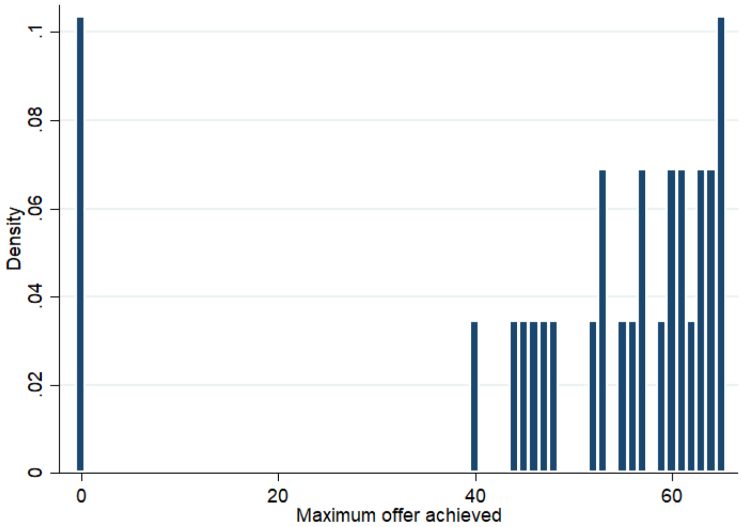

Figure 1.6. Histogram of maximum offers achieved in the outside option treatment

As a consequence of the further evidence in additional treatments, we can confidently deduce that people have an intrinsic tendency to become active instead

[16] See Table 1.A.11 in Appendix for details.
[17] 62% of the players achieved a task outcome (max. offer - costs) of 55 or below.

of staying passive even when it is costly. Moreover, they ignore the other parts of the problem and focus only on their own tasks once they become active.

1.6.4 Follow-up Survey

We employ a follow-up survey to inspect the relevance of active participation bias in real-life situations. In the survey, we asked participants from the experiment to indicate how much certain behavior of active participation applies to them in various daily choices such as consumption, investment, and time management. The survey includes 7 questions related to active participation bias, 7 questions about narrow framing, 1 question on the importance of internal motivation, and 2 vignette questions. In the vignette, we depict a situation in which one person chooses whether to take costly action when there is no alternative task available and the expected gain from the action is negative. We also ask whether that person should keep engaging in that activity further.

The answer to the vignette question about starting an activity is predicted by the active participation bias index[18] and the question about engaging further in that activity is predicted by the narrow framing index. Moreover, both the active participation bias index and the related vignette question are correlated with their search behavior in the main experiment.

These results show that active participation behavior is persistent across different domains which supports the external validity of our experiment. You can see Section 1.A.4 in the Appendix for the exact questions and the detailed analysis.

1.7 Conclusion

In this study, we show that active participation bias is a widespread phenomenon. People intrinsically prefer to be active and they are willing to pay considerable costs to avoid idleness. If there is only one activity available, people engage in it even when they know it is not optimal. On top of that, they narrow-mindedly focus on their individual task after they start so that they ignore the other options.

This insight can help us design our work environments in a more thoughtful way. Executives can be inclined to take ineffective actions if they are not presented with alternative options. Similarly, if an alternative side task is not provided to employees, they can involve in futile or even harmful tasks. This tendency to be active can lead to huge time and monetary costs when it is prolonged by narrow framing. Hence, employees can allocate their time and energy more efficiently if the work environment encourages them to question their tasks frequently and consider alternative options.

[18] The index is simply created as the sum of 7 relevant survey questions.

Likewise, the precision in experiments can be improved by considering active participation bias. When running an RCT, we must refrain from comparing an active behavior to the alternative of staying passive. Otherwise, it would not be possible to clearly distinguish the correct preference for that specific behavior from the general tendency to be active. In fact, the results of this paper makes many previous studies questionable. We must reevaluate many existing findings to separate the effect of active participation bias.

Furthermore, active participation bias can help us explain many puzzles in real life. For instance, it can be a reason behind the high levels of voter turnout. As a reflection of active participation bias, individuals may feel the urge to participate in a common activity even when the probability of being pivotal is minuscule[19]. Besides, active participation bias can be a motive behind the surge in stock exchange trading during COVID-19 lockdowns, both at the internal and the external margin[20]. Individuals might have been inclined towards financial trading as a way of participating in economic activities when most of the other sectors were on hold.

As the first study that examines active participation bias in detail, we mainly focus on why and how it operates. A natural next step in research would be finding some debiasing strategies. Since the active participation motive was never completely shut down in our experimental settings, we believe that providing a promising debiasing rule will be challenging.

Finally, the root source of the intrinsic preference toward activity remains an open question. Due to the relative nature of ordinal preferences, we cannot determine using our data whether activity brings positive utility or idleness evokes negative feelings. Perhaps, neural imaging can provide some answers to the question but it is beyond the scope of this paper.

Appendix 1.A Appendix

1.A.1 Theoretical Predictions

Reminder: In each group, there is a low-type (L) and a high-type (H) player. They search separately for offers and only the highest offer in the group is relevant for the payoff. At the end of the search period, each group member earns the highest offer in the group and pays individual search costs. The offers for the low-type are drawn from a uniform discrete random variable between 0 and 65 whereas the offers for the high-type are drawn from a uniform discrete random variable

[19] Both Dellavigna et al. (2017) and Rogers, Ternovski, and Yoeli (2016) reveal that social image concerns are weak and cannot explain voter turnout. On the other hand, Gerber, Green, and Larimer (2008) demonstrate that the probability to vote increases if people believe others vote, too.

[20] See Ortmann, Pelster, and Wengerek (2020) for a detailed description.

between 35 and 100. The cost of getting another offer is 1 for both players.

Equilibrium: First, we look at the individual search. Assume a player has the highest offer of x at some point in the search phase. The expected gain from drawing another offer becomes

$$V(x) = \sum_{x_i} Pr(x_i)\max\{x_i, x\} - x_i = \sum_{x_i > x} Pr(x_i)(x_i - x) \qquad (1.A.1)$$

The agent keeps drawing offers whenever the continuation value $V(x)$ is bigger than the search cost of $c = 1$. Note that, $V(x)$ is decreasing in x since both summed values $(x_i - x)$ and the summation interval $(x_i > x)$ get smaller when x gets bigger. Thus, there must be a cutoff point R such that $V(x) > c$ for $x < R$ and $V(x) < c$ for $x > R$. Hence, the optimal strategy in a sequential search game is always a threshold strategy.

In our specifications, the continuation value for a high-type at an offer of 89 is

$$V^H(89) = \sum_{x_i \in \{90,\ldots,100\}} \frac{1}{66}(x_i - 89) = \frac{1}{66}(1 + 2 + \ldots + 11) = 1 \qquad (1.A.2)$$

Since the continuation value is equal to the search cost of $c = 1$, the high-type is indifferent between drawing another offer and stopping. For simplicity, we assume an agent stops if the net gain from continuation is equal to the cost. Thus, the optimal strategy for high-types becomes a threshold strategy of $R^H = 89$. This assumption does not change any of the results.

Similarly, the continuation value for a low-type at an offer of 54 is

$$V^L(54) = \sum_{x_i \in \{55,\ldots,65\}} \frac{1}{66}(x_i - 54) = \frac{1}{66}(1 + 2 + \ldots + 11) = 1 \qquad (1.A.3)$$

Therefore, the optimal strategy for a low-type in an individual search game is a threshold strategy of $R^L = 54$.

Now, we can look at the group level. The threshold strategy of $R^H = 89$ is a dominant strategy for the high-type because the low-type cannot reach an offer above 65 and the high-type is strictly better of by searching at any offer below 89. Therefore, in equilibrium, the high-type should follow the same strategy of $R^H = 89$ regardless of the low-type's strategy.

The best response of the low-type to high-type's strategy of $R^H = 89$ is not searching, i.e., a threshold of $R^L = 0$, since it is not possible for the low-type surpass the level 89. Any draw incurs a cost without any potential gain. Therefore, any strategy other than not searching is dominated for the low-type. Hence, the unique equilibrium of the game becomes $(R^H, R^L) = (89, 0)$. In equilibrium, the high-type searches until finding an offer of 89 or above and the low-type does not search.

The equilibrium described above assumes that agents maximize their expected earnings which is based on risk neutrality. Nonetheless, using different utility functions that entail risk aversion does not change the optimal strategies greatly. Table 1.A.1 shows optimal thresholds in individual search based on different utility functions and conventional risk aversion parameters used in the literature. Constant absolute risk aversion (CARA) utility has the following functional form with risk aversion parameter $a \geq 0$:

$$u(c) = \begin{cases} (1 - e^{-ac})/a, & a \neq 0 \\ c, & a = 0 \end{cases}$$

Constant relative risk aversion (CRRA) utility with a risk aversion parameter $\rho \geq 0$ has the following functional form:

$$u(c) = \begin{cases} \frac{c^{1-\rho}-1}{1-\rho}, & \rho \neq 1 \\ ln(c), & \rho = 1 \end{cases}$$

Table 1.A.1. Optimal threshold levels in individual search game for low and high types based on different utility functions and risk aversion parameters.

	CARA			CRRA		
	$a \leq 0.04$	$a = 0.05$	$a = 0.08$	$\rho \leq 4$	$\rho = 5$	$\rho = 8$
R_{high}	89	88	88	89	88	87
R_{low}	54	53	53	54	51	51

Note that, parameters $a = \rho = 0$ indicates risk neutrality. Almost all empirical findings estimate parameters[21] $a < 0.02$ and $\rho < 2$. As seen in Table 1.A.1, even extreme risk aversion parameters do not lead to dramatically different thresholds.

Remark: In the computer treatment, we imposed a time limit of 10 minutes and a draw time of 10 seconds onto the computer algorithm. Therefore, the maximum number of draws was 60 for the computer. The probability of not achieving the threshold of 90 in 60 draws is minuscule, such that

$$Pr\{\max x_i < 90 | n = 60\} = \left(\frac{55}{66}\right)^{60} < 2.10^{-5} \tag{1.A.4}$$

Moreover, the probability of high-type not receiving an offer above 65 is even much smaller:

$$Pr\{\max x_i < 65 | n = 60\} = \left(\frac{30}{66}\right)^{60} < 2.10^{-20} \tag{1.A.5}$$

[21] See Babcock, Choi, and Feinerman (1993) and Evans (2005) as some examples.

Hence, it was almost impossible for low-types to have an offer higher than the maximum offer of the computer algorithm. Even the extreme risk aversion parameters mentioned above do not alter the optimality of not searching for the low-types. Indeed, the constraint of 60 draws was never binding in our experiments.

1.A.2 Main Treatments

Table 1.A.2. Linear regression of attention check on treatment and type. There is no statistical difference between different groups players.

	attention check successful
social, low-type	0.0606
	(0.0423)
computer, low-type	0.0146
	(0.0408)
constant (social, high-type)	0.910***
	(0.0301)
N	216

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.A.3. OLS regression of the indicator for answering the bonus question correctly on the treatment dummy.

	bonus correct
computer treatment	0.0232
	(0.0822)
constant (social treatment)	0.612***
	(0.0596)
N	141

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.A.4. Distribution of the participants according to search activity, treatment, and the bonus question.

	social		computer	
	correct	not	correct	not
searched	39	26	37	26
no search	2	0	10	1

Table 1.A.5. The effect of cognitive ability on answering the bonus question correctly. The IQ score (measured by Raven's matrices) and self-reported math ability have a positive impact.

	(1)	(2)	(3)
	bonus correct	bonus correct	bonus correct
comp. treatment	-0.0234	0.0268	0.0255
	(0.0839)	(0.0816)	(0.0819)
IQ	0.0684^{**}		
	(0.0312)		
math. ability		0.0258^{*}	
		(0.0146)	
RAT score			-0.0392
			(0.0269)
constant	0.323^{**}	0.467^{***}	0.660^{***}
	(0.144)	(0.101)	(0.0681)
N	141	141	141

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

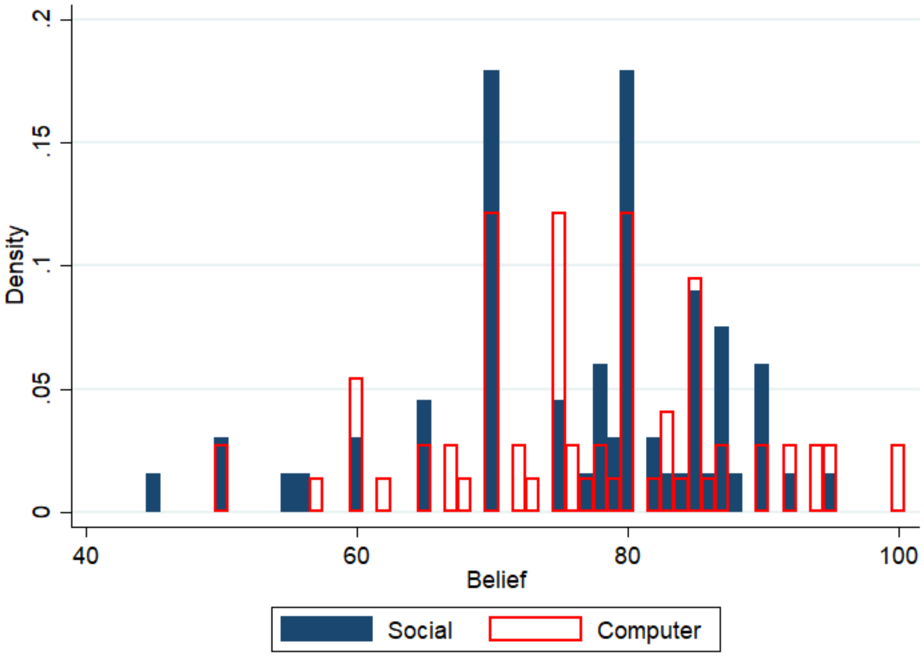

Figure 1.A.1. Histogram of the beliefs of low-types about the performance of their partners in social and computer treatments. A vast majority had beliefs above 55, particularly above 65.

Table 1.A.6. OLS regression of the dummy for zero draws on demographics and personality measures. None of the factors have a significant impact on search behavior.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
comp. treatment	0.124**	0.119**	0.119**	0.126**	0.119**	0.126**	0.119**
	(0.0485)	(0.0483)	(0.0483)	(0.0490)	(0.0483)	(0.0487)	(0.0482)
age	0.0055						
	(0.0058)						
income		-0.0116					
		(0.0397)					
education			0.0023				
			(0.0318)				
punish for me				0.0081			
				(0.00998)			
punish for others					-0.0001		
					(0.00975)		
altruism						-0.0105	
						(0.0109)	
self control							-0.0034
							(0.008)
constant	-0.107	0.0396	0.0285	-0.0003	0.0305	0.104	0.0529
	(0.148)	(0.0483)	(0.0396)	(0.0510)	(0.0600)	(0.0841)	(0.0575)
N	141	141	141	141	141	141	141

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.A.7. OLS regression of the dummy for zero draws on big five personality traits. Only neuroticism has a significant impact but it disappears in other treatments.

	(1)	(2)	(3)	(4)	(5)	(6)
	nosearch	nosearch	nosearch	nosearch	nosearch	nosearch
treatment	0.118**	0.121**	0.114**	0.132***	0.116**	0.134***
	(0.0485)	(0.0483)	(0.0495)	(0.0478)	(0.0486)	(0.0503)
extraversion	0.00333					0.00663
	(0.0278)					(0.0315)
aggreeableness		-0.0169				-0.00610
		(0.0281)				(0.0287)
conscientiousness			-0.0103			0.0146
			(0.0235)			(0.0267)
neuroticism				-0.0526**		-0.0752***
				(0.0236)		(0.0288)
openness					0.0126	0.0417
					(0.0247)	(0.0283)
constant	0.0161	0.110	0.0762	0.271**	-0.0294	0.115
	(0.120)	(0.138)	(0.112)	(0.114)	(0.121)	(0.225)
N	141	141	141	141	141	141

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.A.8. OLS regression of the indicator for recall behavior on the treatment dummy. There is no significant difference in recall frequencies.

	recall
computer treatment	0.0867
	(0.0790)
social treatment	0.231^{***}
	(0.0555)
N	128

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

1.A.3 Additional Treatments

Table 1.A.9. OLS regression of zero draws dummy on all the control variables in social, computer, and high intensity treatments. Only risk aversion, patience, income, and education have some effects.

	Not drawing any offer	
computer treatment	0.164***	(0.0499)
high treatment	0.0784	(0.0495)
Raven score	-0.0106	(0.0142)
RAT score	0.0183	(0.0137)
math ebility	0.00184	(0.00739)
belief	-0.00146	(0.00195)
trust	-0.00296	(0.00783)
risk aversion	-0.0185**	(0.00935)
patience	0.0204*	(0.0108)
age	0.00110	(0.00596)
income	-0.0727**	(0.0366)
education	0.0610*	(0.0342)
punish for me	0.0127	(0.00864)
punish for others	-0.00560	(0.00927)
altruism	-0.0127	(0.00970)
control	0.00381	(0.00761)
extraversion	0.0149	(0.0281)
aggreeableness	0.00213	(0.0258)
conscientiousness	0.0361	(0.0223)
neuroticism	-0.0404	(0.0249)
openness	0.00899	(0.0241)
constant	0.0661	(0.292)
N	214	

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.A.10. OLS regression of the dummy for drawing only one offer on all the control variables in minmax treatment. Only RAT score, risk aversion, and conscientiousness have significant impacts.

	Drawing exactly one offer	
Raven score	-0.0822	(0.0609)
RAT score	0.121*	(0.0648)
math ability	-0.0257	(0.0465)
belief	0.000998	(0.00652)
trust	-0.0204	(0.0419)
risk aversion	0.0888*	(0.0497)
patience	0.0510	(0.0594)
age	0.0201	(0.0255)
income	-0.179	(0.168)
education	-0.105	(0.185)
punish for me	0.0171	(0.0409)
punish for others	0.0259	(0.0596)
altruism	-0.0427	(0.0420)
self control	0.0104	(0.0304)
extraversion	0.128	(0.123)
aggreeableness	-0.0356	(0.162)
conscientiousness	0.346**	(0.131)
neuroticism	-0.122	(0.165)
openness	0.0491	(0.0957)
constant	-1.697	(1.622)
N	42	

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.A.11. Distribution of participants by option choice and bonus question in option treatment.

	Bonus question	
Option	correct	wrong
accepted	11	1
rejected	11	16

Table 1.A.12. OLS regression of accepting the option on skill and preference measures in option treatment. None of the variables has a significant effect.

	(1) accepted	(2) accepted	(3) accepted	(4) accepted	(5) accepted
Raven's score	-0.0177				
	(0.0456)				
rat score		-0.0506			
		(0.0476)			
math			-0.0198		
			(0.0226)		
risk aversion				0.0310	
				(0.0280)	
patience					-0.00162
					(0.0356)
constant	0.394^{*}	0.383^{***}	0.405^{***}	0.156	0.322
	(0.229)	(0.0999)	(0.131)	(0.156)	(0.285)
N	42	42	42	42	42

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.A.13. OLS regression of accepting the option on demographic and personality measures in option treatment. None of the variables has a significant effect.

	Accepting the option	
age	0.0270	(0.0224)
income	0.0102	(0.132)
education	-0.192	(0.142)
punish for me	0.0497	(0.0374)
punish for others	-0.0449	(0.0452)
altruism	0.0186	(0.0429)
self control	0.0187	(0.0250)
extraversion	-0.193	(0.127)
aggreeableness	0.0143	(0.0893)
conscientiousness	-0.0963	(0.0921)
neuroticism	-0.0140	(0.111)
openness	-0.00308	(0.106)
constant	0.707	(0.812)
N	39	

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

1.A.4 Follow-up Survey

The English translation of the statements are listed below. Participants indicate how much each situation applies to them.
Active Participation Questions

(1) When a new type of investment becomes popular and has made good returns recently, I want to invest in it too.

(2) I think it is important to join public protests, even if the goals of the protests are not achieved.

(3) I feel uncomfortable when I have nothing to do.

(4) I'm more likely to read a book that's already a bestseller.

(5) I try out the latest fashion trends.

(6) Once I'm in Las Vegas, I also go to the casino and gamble.

(7) I watch old movies again to kill time.

Narrow Framing Questions

(1) I try to learn all the topics in a course, even if they won't be on the exam.

(2) If I have a medical problem, I try to get as much information about it as possible.

(3) Before I buy anything, I inform myself extensively about all alternatives.

(4) When I start something, I want to finish it, even if it's less fun than I initially thought.

(5) It bothers me not to continue watching a TV series for which more episodes are available.

(6) I finish the fries in a restaurant, even if I'm full.

(7) When I start a puzzle or crossword puzzle, I have to finish it, even if I get bored.

Internal Motivation Question: Internal motivation is more important to me than financial incentives.
We also ask two questions about the following scenario. The first question indicates active participation bias and the second question is on narrow framing.

> *Charlie is moving to a new country where cricket is a popular sport. He has seen some cricket matches in his home country, but he did not like them at all. One evening Charlie is alone at home and has nothing to do. A cricket match starts on the television.*

Question 1. Charlie should watch the match.
Question 2. He should try to better understand the playing styles of both teams so he can enjoy the game more.

Table 1.A.14. OLS regression of vignette questions on active participation bias index, narrow framing index, and internal motivation as a control variable. Starting to watch the match is associated with active participation whereas trying to learn the details of the game is associated with both indexes.

	(1)		(2)	
	Charlie start		Charlie details	
Active Participation Bias Index	0.0658***	(0.0179)	0.0399**	(0.0178)
Narrow Framing Index	0.00822	(0.0177)	0.0310*	(0.0177)
internal motivation	-0.0186	(0.0726)	0.0584	(0.0724)
constant	1.692***	(0.398)	1.589***	(0.397)
N	295		295	

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.A.15. OLS regression of the indicator for not searching on answers to the vignette question. People who report Charlie should start watching the game are more likely to show active participation in the experiment.

	(1)		(2)	
	zero search		zero search	
Charlie: start watching	-0.0399*	(0.0224)	-0.0465**	(0.0213)
Charlie: learn details	0.0532**	(0.0228)	0.0489**	(0.0218)
constant	0.0784	(0.0651)	0.0274	(0.0729)
N	190		190	
Treatment controls	No		Yes	

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.A.16. OLS regression of the indicator for not searching on active participation and narrow framing indexes together with internal motivation and treatment controls. The effect of active participation index is not significant (possibly due to small sample size) but the coefficient is negative, indicating that a higher bias measure is associated with a higher probability of active search.

	zero search	
Computer Treatment	0.119^{*}	(0.0664)
High Intensity Treatment	0.0427	(0.0623)
Minmax Treatment	-0.0291	(0.0716)
Option Treatment	0.329^{***}	(0.0739)
Active Participation Bias Index	-0.00724	(0.00569)
Narrow Framing Index	0.00492	(0.00590)
internal motivation	0.0532^{**}	(0.0229)
constant	-0.0669	(0.130)
N	190	

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

References

Abbink, Klaus, Jordi Brandts, Benedikt Herrmann, and Henrik Orzen. 2010. “Intergroup Conflict and Intra-group Punishment in an Experimental Contest Game.” *American Economic Review* 100 (1): 420–47. https://doi.org/10.1257/AER.100.1.420. [7]

Babcock, Bruce A, E Kwan Choi, and Eli Feinerman. 1993. “Risk and Probability Premiums for CARA Utility Functions.” *Journal of Agricultural and Resource Economics* 18 (1). https://doi.org/www.jstor.org/stable/40986772. [26]

Bar-Eli, Michael, Ofer H. Azar, Ilana Ritov, Yael Keidar-Levin, and Galit Schein. 2007. “Action bias among elite soccer goalkeepers: The case of penalty kicks.” *Journal of Economic Psychology* 28 (5): 606–21. https://doi.org/10.1016/J.JOEP.2006.12.001. [8]

Barberis, Nicholas, Ming Huang, and Richard H. Thaler. 2006. “Individual Preferences, Monetary Gambles, and Stock Market Participation: A Case for Narrow Framing.” *American Economic Review* 96 (4): 1069–90. https://doi.org/10.1257/AER.96.4.1069. [5, 18]

Bilodeau, Marc, Jason Childs, and Stuart Mestelman. 2004. “Volunteering a public service: an experimental investigation.” *Journal of Public Economics* 88 (12): 2839–55. https://doi.org/10.1016/J.JPUBECO.2004.02.001. [7]

Carpenter, Jeffrey, Allison Liati, and Brian Vickery. 2010. “They come to play: Supply effects in an economic experiment.” *Rationality and Society* 22 (1). https://doi.org/10.1177/1043463109358486. [7]

Chen, Daniel L., Martin Schonger, and Chris Wickens. 2016. “oTree—An open-source platform for laboratory, online, and field experiments.” *Journal of Behavioral and Experimental Finance* 9: 88–97. https://doi.org/10.1016/J.JBEF.2015.12.001. [10]

Dechenaux, Emmanuel, Dan Kovenock, and Roman M Sheremeta. 2015. “A survey of experimental research on contests, all-pay auctions and tournaments.” *Experimental Economics* 18 (4): 609–69. https://doi.org/10.1007/s10683-014-9421-0. [6]

Dellavigna, Stefano, John A. List, Ulrike Malmendier, and Gautam Rao. 2017. “Voting to tell others.” *Review of Economic Studies* 84 (1). https://doi.org/10.1093/restud/rdw056. [24]

Evans, David J. 2005. “The Elasticity of Marginal Utility of Consumption: Estimates for 20 OECD Countries.” *Fiscal Studies* 26 (2): 197–224. https://doi.org/10.1111/J.1475-5890.2005.00010.X. [26]

Forsythe, Robert, Joel L. Horowitz, N. E. Savin, and Martin Sefton. 1994. “Fairness in Simple Bargaining Experiments.” *Games and Economic Behavior* 6 (3): 347–69. https://doi.org/10.1006/GAME.1994.1021. [9]

Gerber, Alan S., Donald P. Green, and Christopher W. Larimer. 2008. “Social pressure and voter turnout: Evidence from a large-scale field experiment.” *American Political Science Review* 102 (1). https://doi.org/10.1017/S000305540808009X. [24]

Goerg, Sebastian J., Sebastian Kube, and Jonas Radbruch. 2019. “The Effectiveness of Incentive Schemes in the Presence of Implicit Effort Costs.” *Management Science* 65 (9): 4063–78. https://doi.org/10.1287/MNSC.2018.3160. [7]

Hey, John D. 1982. “Search for rules for search.” *Journal of Economic Behavior and Organization* 3 (1). https://doi.org/10.1016/0167-2681(82)90004-X. [11]

Hörisch, Hannah, and Oliver Kirchkamp. 2010. “Less fighting than expected.” *Public Choice* 144 (1-2). https://doi.org/10.1007/s11127-009-9523-y. [7]

Killingsworth, Matthew A., and Daniel T. Gilbert. 2010. “A wandering mind is an unhappy mind.” *Science* 330 (6006). https://doi.org/10.1126/science.1192439. [8]

Kuhn, Peter, Peter Kooreman, Adriaan Soetevent, and Arie Kapteyn. 2011. "The Effects of Lottery Prizes on Winners and Their Neighbors: Evidence from the Dutch Postcode Lottery." *American Economic Review* 101 (5): 2226–47. https://doi.org/10.1257/AER.101.5.2226. [8]

Lei, Vivian, Charles N. Noussair, and Charles R. Plott. 2001. "Nonspeculative Bubbles in Experimental Asset Markets: Lack of Common Knowledge of Rationality vs. Actual Irrationality." *Econometrica* 69 (4): 831–59. https://doi.org/10.1111/1468-0262.00222. [5, 7]

Li, Jiawei, Stephen Leider, Damian Beil, and Izak Duenyas. 2021. "Running online experiments using web-conferencing software." *Journal of the Economic Science Association* 7 (2). https://doi.org/10.1007/s40881-021-00112-w. [10]

Lugovskyy, Volodymyr, Daniela Puzzello, and Steven Tucker. 2010. "An experimental investigation of overdissipation in the all pay auction." *European Economic Review* 54 (8): 974–97. https://doi.org/10.1016/J.EUROECOREV.2010.02.006. [7]

Ortmann, Regina, Matthias Pelster, and Sascha Tobias Wengerek. 2020. "COVID-19 and investor behavior." *Finance Research Letters* 37. https://doi.org/10.1016/J.FRL.2020.101717. [24]

Rogers, Todd, John Ternovski, and Erez Yoeli. 2016. "Potential follow-up increases private contributions to public goods." *Proceedings of the National Academy of Sciences of the United States of America* 113 (19): 5218–20. https://doi.org/10.1073/pnas.1524899113. [24]

Schotter, Andrew, and Yale M. Braunstein. 1981. "Economic Search: An Experimental Study." *Economic Inquiry* 19 (1): 1–25. https://doi.org/10.1111/J.1465-7295.1981.TB00600.X. [11]

Sheremeta, Roman M. 2010. "Experimental comparison of multi-stage and one-stage contests." *Games and Economic Behavior* 68 (2): 731–47. https://doi.org/10.1016/J.GEB.2009.08.001. [7]

Sheremeta, Roman M. 2011. "Perfect-Substitutes, Best-Shot, and Weakest-Link Contests Between Groups." *SSRN Electronic Journal,* https://doi.org/10.2139/ssrn.1516019. [7]

Wilson, Timothy D., David A. Reinhard, Erin C. Westgate, Daniel T. Gilbert, Nicole Ellerbeck, Cheryl Hahn, Casey L. Brown, and Adi Shaked. 2014. "Just think: The challenges of the disengaged mind." *Science* 345 (6192): 75–77. https://doi.org/10.1126/SCIENCE.1250830. [8]

Zeelenberg, Marcel, and Rik Pieters. 2004. "Consequences of regret aversion in real life: The case of the Dutch postcode lottery." *Organizational Behavior and Human Decision Processes* 93 (2): 155–68. https://doi.org/10.1016/J.OBHDP.2003.10.001. [8]

Chapter 2

Voting Under Salience Bias and Strategic Extremism[⋆]

Joint with Günnur Ege Bilgin

2.1 Introduction

The political economy literature classically presumes that office-oriented candidates observe the electorate and take positions that match the majority voter preferences. However, there are political candidates worldwide who take extreme positions on some issues and propose radical policies. Yet, some of these politicians get elected and implement their pledged policies.[1]

Besides, we observe another phenomenon. The policy positions of candidates influence the preferences of voters. Consequently, political positioning can be employed as a strategic tool to shape preferences into a more favorable distribution. In this study, we present a model along with experimental evidence and demonstrate that politicians can optimally choose extreme policies even when none of the voters have extreme preferences. We examine how policy proposals affect preferences and lead to extreme policies, thereby explaining the extreme policy choices of candidates.

We show that when voters exhibit salience bias (i.e., overemphasize the importance of salient issues), candidates can manipulate this bias by adopting radical stances in a policy where they have an advantage. This way they draw attention

[⋆] I am thankful to Florian Zimmermann and Thomas Dohmen for many discussions, and to Lucas Coffman, Frank Schilbach, and Mehmet Ekmekci for helpful comments. Funding by the Reinhard Selten Institute (RSI) through Selten Grant is gratefully acknowledged. This study is preregistered under Open Science Framework: https://osf.io/cfgdy . The ethical approval by German Association for Experimental Economic Research e.V.: No. J2S1c2TN

[1] For instance Donald Trump builds a multi-billion dollar wall (BBC News (2017)). See Carothers and O'Donohue (2019) for an overview of different countries.

to that issue and create demand for it by shifting the preferences of voters. When we extend the same argument to two candidates, the electoral competition may become an arms-race scenario where each candidate aims to take an extreme position on a different issue and tries to persuade voters that their issue is the most relevant one.

We implement the probabilistic voting model by Persson and Tabellini (2002) with two politicians and a two-dimensional policy platform. As a special case of their model, our voters are not divided into groups and are quite similar to each other in taste, apart from some noise factors. The voters consider the utility they would get from each candidate, which would be driven by the policy choices of candidates. In addition, the utility of voters is affected by salience. A policy dimension becomes more salient as candidates become more diverse, and more salient issues are overemphasized by the voters, similar to Bordalo, Gennaioli, and Shleifer (2012). The voter then votes for the candidate whose policy choices would bring higher utility.

Politicians are aware of the salience bias and by choosing and committing to the two-dimensional policy proposal, they maximize their probability of winning, i.e. their vote share. They are constrained by the same governmental budget. However, they differ in their marginal costs to implement each policy, which reflects both pecuniary and non-pecuniary costs. Pecuniary costs reflect the sources the candidate has such as tools, factories, and manpower, which enables the candidate to provide policies at lower costs. Non-pecuniary costs, on the other hand, reflect the connections of the candidate. For instance, if the candidate's main supporting lobby is in favor of a policy, then shifting resources to the other is more costly for her.

In any equilibrium of the model with two candidates, always the same issue is salient for both candidates. Both candidates invest more in the salient issue, and extremism is enhanced with the salience bias. The candidate who is relatively more advantageous in the salient issue can increase her advantage by choosing even higher levels. Which issue will be salient in the equilibrium will be dictated by the parameters of the model, namely the relative cost advantages.

There is a new but sizable literature on extremism in politics. Unlike our approach, most of the studies assume that there are existing divides in each society and politicians use that polarization to gain power. However, we show that a radical vote base is not necessary for extreme policies. Politicians can promote an issue as the most crucial aspect of the election by taking an extreme position on that. This way, they can manufacture radicalization. We also show that extremism is exacerbated if an issue is already a hot topic. Hence, existing polarization in a society would have a multiplicative effect on extremism.

Our model also explains increased mobilization through extremism. When candidates take disparate positions, the welfare difference between candidates gets larger. Thus, voters have a greater incentive to vote. Additionally, our model can be used to analyze run-off elections and the effect of existing polarization. Furthermore, the tractable form of the model can be used in most of the more complex models to investigate various phenomena.

Additionally, we test the predictions of the model through an experiment with a representative sample of Turkey. We ask subjects to vote on a hypothetical election where hypothetical candidates differ in their positions on climate and defense policy proposals. The experimental findings support the model and confirm that politicians can increase their vote shares by promising extreme policies. We also show that the salience of an issue is the main driver of the voting decision as assumed in the model.

The paper is organized as follows: In Section 2.2, we provide an overview of the related literature. In Section 2.3, the model is described and the equilibrium analysis is provided in Section 2.4. We analyze comparative statics about the optimal choices in Section 2.5. The possible implications of the model regarding mobilization and second-term elections are explained in Section 2.6. Section 2.7 provides the experimental design and the main results. Section 2.8 concludes.

2.2 Literature

This paper lies at the intersection of two strands of literature: extremism and salience. In the extremism literature, most studies try to explain radical politicians as a response to radical voters. This bottom-up argument mainly states that the political preferences of (at least some) people in society shift toward extreme attitudes and politicians take extreme stances to match the demands of their voters.

For instance, Matějka and Tabellini (2021) argues that small groups with stark preferences can alter the political outcomes in their favor. They advocate that electoral candidates give those groups a disproportionately large weight in their policy choices since they are more responsive compared to moderate voters. Similarly, Jones, Sirianni, and Fu (2022) argue that if voters with moderate preferences are less likely to vote, politicians take extreme positions to attract more eager radical voters.

Furthermore, there are studies analyzing extremism as a result of identity politics (Kuziemko and Washington (2018), Grossman and Helpman (2021)), communalism (Enke (2020), Enke, Rodríguez-Padilla, and Zimmermann (2022)), globalization (Rodrik (2021)), and polarization (Nunnari and Zápal (2017), Bursztyn, Egorov, and Fiorin (2020), Enke, Polborn, and Wu (2022)).

On the other hand, many studies show that the salience of an issue is a critical factor in voters' decisions. Colussi, Isphording, and Pestel (2021) clearly show that anti-Muslim parties gain votes if the elections are held right after Ramadan. They also demonstrate the effect of the salience of Muslim minorities as the main mechanism. Likewise, Aragonès and Ponsatí (2022) depict a similar phenomenon using the data from the UK and Catalonia. They show that political parties adjust their positions when an exogenous shock makes an issue more salient.

On top of that, there are studies that show the effect of salience is exacerbated when combined with existing stereotypes. Bordalo, Gennaioli, and Shleifer (2020) and Bonomi, Gennaioli, and Tabellini (2021) show that a salient divide in society creates radical preferences via negative stereotypes. Furthermore, Spirig (2023) shows the strength of salience using Swiss data. When immigration becomes more salient, not only the voter preferences but also the decisions of judges become less favorable for minorities.

Furthermore, some studies show that politicians strategically manipulate the salience of some issues to gain an advantage. For instance, Lewandowsky, Jetter, and Ecker (2020) provide evidence for Donald Trump using Twitter to manipulate the salience of some issues. Similarly, Glaeser (2005) show that politicians can supply hate stories to shape the preferences of individuals. Balart, Casas, and Troumpounis (2022) also show that politicians can exploit social media platforms to push radical opinions.

However, none of the papers in the literature examines the positioning of candidates as a potential manipulation of the salience of different issues. Yet, the idea of politicians positioning themselves in different attributes is very similar to firms choosing different price, quality, and/or quantity levels to compete with other firms. Although there are differences between firms and politicians, the closest resemblance to our model can be found in IO literature. Several papers show that firms design their menus such that they influence the salience of some aspects of products. The canonical paper by Bordalo, Gennaioli, and Shleifer (2016) (together with Bordalo, Gennaioli, and Shleifer (2013)) provides a model that explains the product choices of firms to exploit the salience bias of consumers.[2]

2.3 Model

There are two purely office-oriented candidates running for the election, $i = \{A, B\}$. Both candidates announce and commit to two policy choices $q = (x_i, y_i) \in \mathbb{R}^2_+$, which represent the government spending they will allocate to the two subjects.

There is a continuum of voters. Voters do not have the option to abstain. Following the Probabilistic Voting Model by Persson and Tabellini (2002), they simply vote for the candidate whose policy proposal is more favorable. Observing the policy choices, a single voter's utility from candidates is as follows:

$$v(i) = \ln x_i + m \ln y_i$$

[2] See the book chapter by Herweg, Müller, and Weinschenk (2018) for an analysis of these models and their implications.

However, we assume that the agents have bounded rationality and their attention is limited a la BGS. To be more specific, as the policies in one spectrum are wider spread from each other, this drives the voters' attention to that aspect, resulting in an increase of the relative utility weight that issue in their utility function. In particular, the policy choices of the politicians affect voter preferences such that for $\delta > 1$:

$$v(i) = \begin{cases} \delta \ln x_i + m \ln y_i & \text{if } \frac{|x_i - \bar{x}|}{\bar{x}} > \frac{|y_i - \bar{y}|}{\bar{y}} \\ \ln x_i + m \ln y_i & \text{if } \frac{|x_i - \bar{x}|}{\bar{x}} = \frac{|y_i - \bar{y}|}{\bar{y}} \\ \ln x_i + \delta m \ln y_i & \text{if } \frac{|x_i - \bar{x}|}{\bar{x}} < \frac{|y_i - \bar{y}|}{\bar{y}} \end{cases}$$

BGS uses a more general salience function. However, in this version of the paper we are restricting our attention to a more specific one, which indicates that a policy attribute is more salient for a candidate whenever he deviates from the average spending more, relative to the other policy. Other than the partiality due to salience, the utility function is the sum of two logarithmic utility functions, with a slight adjustment by m that represents the relative importance of issue y for the voters. Voters receive strictly positive utility from both policies, therefore $m > 0$. If $m < 1$, voters care more about policy x without the interference of the salience bias.

Policy choices are not the only factors that affect voter preferences. Additionally, ideological bias towards candidate B denoted by $\beta \sim U\left[\frac{-1}{2\phi}, \frac{1}{2\phi}\right]$ and relative popularity of B denoted $\epsilon \sim U\left[\frac{-1}{2\varphi}, \frac{1}{2\varphi}\right]$ represent the noise in the elections. Once the candidates select their positions, salience reveals and voters calculate the utility they would get from each candidate. Furthermore, the noise factors β and ϵ realize and a voter votes for A if $v(A) > v(B) + \beta + \epsilon$[3].

Both politicians are trying to maximize their probability of winning, which, with the logic explained above, is equal to $[v(i) - v(j)]\,\varphi + \frac{1}{2}$ for candidate i. Furthermore, they are bounded by a budget constraint $c_x^i x_i + c_y^i y_i = G$. This budget constraint represents the pecuniary and non-pecuniary costs of each policy for both candidates. For example, if a candidate possesses tools that would ease implementing a policy, he has a lower marginal cost. These tools might be material such as factories, skilled teams, and other apparatus. However, they could also represent other structures such as networks and lobbies. If the main lobby that supports a candidate is in favor of policy x, then implementing policy y would be more costly for him. Since voters get positive utility from both policies, for a non-trivial analysis of equilibrium policy choices, we impose $c_x^A < c_x^B$ and $c_y^B < c_y^A$.

[3] β realizes for each individual, whereas ϵ realizes as a common variable for the whole electorate

Simple intuition would hint at the fact that both candidates would want to highlight the dimension in which they have a comparative advantage. At this point, a bridging fact that is shown by BGS simplifies our analysis a lot:

Lemma 2.1. *x is salient by $A \iff x$ is salient by B. (BGS 2012)*

2.4 Equilibrium Analysis

As a result of the features discussed above, a voter with $\tilde{\beta} = v(A) - v(B) - \epsilon$ indifferent between the two candidates and the vote share of A can be calculated as $\Pi_A = \mathbb{P}(\beta < \tilde{\beta}) = \left[\tilde{\beta} + \frac{1}{2\phi}\right]\phi$ and the probability of candidate A winning the election is $\mathbb{P}(\Pi_A > \frac{1}{2}) = \mathbb{P}\left(v(A) - v(B) - \epsilon + \frac{1}{2\phi} > \frac{1}{2\phi}\right) = [v(A) - v(B)]\,\varphi + \frac{1}{2}$

Furthermore, as discussed in the previous section, candidates try to maximize their probability of winning. They only have control over their own policy choices and take other candidate's positioning as given. Therefore, candidate A's problem is:

$$\max_{\{x_A, y_A\}} [v(A) - v(B)]\,\varphi + \frac{1}{2} \tag{2.1}$$

$$\text{s.t. } c_x^A x_A + c_y^A y_A = G \tag{2.2}$$

A key analysis requires embranchment after this point. This is due to the fact that both $v(A)$ and $v(B)$ depend on the salient issue in the election. From the lemma, we know that the same issue will be salient for both candidates, therefore we can call it the salience issue of the election. As the first branch, suppose there exists an x—salient equilibrium. Then, the maximization problem of candidate A is quite straightforward:

$$\max_{\{x_A, y_A\}} [\delta \ln x_A + m \ln y_A - \delta \ln x_B - m \ln y_B]\,\varphi + \frac{1}{2} \tag{2.3}$$

$$\text{s.t. } c_x^A x_A + c_y^A y_A = G \tag{2.4}$$

Since the candidates can only affect their own positions, the problem resembles a basic utility maximization problem of a consumer with a budget constraint. As usual, optimality of the interior solution requires:

$$\frac{\delta y_A}{m x_A} = \frac{c_x^A}{c_y^A}$$

Proposition 2.2. *In an x-salient equilibrium, the optimally chosen policy profiles of both candidates are as in the following table, and the equilibrium indeed is x-salient iff $\frac{c_x^B}{c_x^A} > \frac{c_y^A}{c_y^B}$.*

$x_A^* = \frac{G\delta}{(\delta+m)c_x^A}$	$x_B^* = \frac{G\delta}{(\delta+m)c_x^B}$
$y_A^* = \frac{G\delta}{(\delta+m)c_y^A}$	$y_B^* = \frac{G\delta}{(\delta+m)c_y^B}$

Observe that in such an equilibrium $x_A^* > x_B^*$ and $y_B^* > y_A^*$. Furthermore, this equilibrium can be sustained if and only if $c_x^B/c_x^A > c_y^A/c_y^B$, meaning that the relative cost advantage of candidate A in policy x should be higher than the relative cost advantage of candidate B in policy y. Furthermore, candidate A wins if and only if $\delta \ln \frac{c_x^B}{c_x^A} - m \ln \frac{c_y^A}{c_y^B} > \epsilon$. The equilibrium policy choices and the necessary condition of a y-salient equilibrium can be found in the appendix.

2.5 Comparative Statics

In this section, we provide comparative statics of the equilibrium and provides explanations. First of all, in both x-salient and y-salient equilibria, $x_A^* > x_B^*$ and $y_B^* > y_A^*$. This is not related to salience but is solely due to the different cost functions of the candidates. Each candidate prefers higher amounts in the policy that is less costly for him.

Moreover, in x-salient equilibrium, x_i^* increases with δ and in y-salient equilibrium, y_i^* increases with δ. This explains that politicians respond to salience in the sense that they provide more on the salient issues. Thus, the salience has an overshooting effect such that voters' utility from the salient issue increases even more.

The probability of candidate A winning the election in an x-salient equilibrium increases with the salience of x and the cost advantage of A in policy x and decreases with the relative importance of issue y and the cost advantage of B in policy y as expected.

Observe that A prefers an x-salient equilibrium since he has the absolute advantage and will provide more than B in any case. However, which equilibrium is to be sustained will be determined by exogenous variables and the candidates have no means of choosing the equilibrium. With two candidates, they respond to salience only by choosing their own policies, not by the salience structure of the equilibrium.

However, even with this simple strategic behavior, in x-salient equilibrium, $x_A^* - x_B^*$ increases with δ and $y_B^* - y_A^*$ decreases with δ. This sustains the salience bias in policy x.

In the following section, we consider an extension to the model where another candidate is introduced into the environment.

2.6 Extensions and Implications

2.6.1 Introduction of a Decoy Candidate

Similar to the industrial organization literature, an interesting implication of this model occurs when a decoy candidate appears on the election platform. In marketing, the *decoy effect* is the phenomenon whereby consumers tend to have a specific change in preference between two options when also presented with a third option that is dominated. In social choice, it is known as *indepence of irrelevant alternatives* (Cane and Luce (1960)) and in matching theory, the notion corresponds to *irrelevance of rejected contracts* (Aygün and Sönmez (2013)). In any of the fields, the flavor is similar: An alternative that is not going to be chosen by the decision-makers should not affect the choice process at all.

In this paper, a candidate is a *decoy* if he is unlikely to be chosen but affects the election outcomes by interfering with salience. We show that, for a given policy choice, an initially disadvantageous candidate might benefit from the existence of a decoy candidate.

Consider an initial setup where candidates A and B choose relatively moderate locations in policy y, whereas their policy choices are wider spread in policy x, such that policy x is the salient issue for both candidates. Additionally, suppose B chooses a higher level of x and for non-triviality, and A chooses a higher level in y. In such a scenario, candidate B has a relatively upper hand by choosing more in the salient issue.

Now we introduce a third candidate C in the election. Candidate C is a far-extremist in policy y and will not allocate any budget to policy x. This simple assumption ensures that candidate C will not be chosen in any kind of equilibrium due to the utility function of the voters. The following proposition shows that, even though C will not be voted for by any voter, his existence can affect the outcome of the election by interfering with salience, and salience only.

Proposition 2.3. *Suppose the alignment of the candidates is as in the table below, and* $h > \varepsilon > 0$, $\frac{h}{\bar{x}} > \frac{\varepsilon}{\bar{y}}$ *and* $\bar{x} > h > \frac{\bar{x}}{3}$.

	A	B	C
x	$\bar{x}-h$	$\bar{x}+h$	0
y	$\bar{y}+\varepsilon$	$\bar{y}-\varepsilon$	ω

Then, introduction of an extremist candidate C where ω is large enough ($\omega > \frac{4\bar{x}\bar{y}+6\bar{y}h+6\bar{x}\varepsilon}{3h-\bar{x}}$ *and* $\omega > \frac{2\bar{y}h-2\bar{x}\varepsilon}{\bar{x}-h}$) *increases the vote share of candidate A if* $m\ln(\bar{y}+\varepsilon) > \ln(\bar{x}-h)$.

First of all, observe that candidate C's choice of 0 in policy x indeed ensures him not being elected. In the initial positions, candidate A would prefer making y

salient. With the far extremist C, policy x is still salient for candidate B. However, with three candidates, it is now possible that different issues are salient for different candidates. If C is extremist enough policy y becomes salient for candidate A. If the utility A creates with policy y exceeds the utility A creates with x, policy y becoming salient for A increases the probability of him winning the election.

The proposition shows that, if voters' rationality is bounded by salience bias, introducing a third candidate can interfere with the election outcome, even though the third candidate is *irrelevant*, in the sense that he does not attract any votes. This candidate only serves as an agenda setter and attracts voters' attention to the policy, in which the initially disadvantageous candidate has a comparative advantage.

2.6.2 Polarization in the Electorate

For this extension, suppose there is an already existing polarization in the electorate. Namely, apart from their ideological bias towards candidate B, the voters also differ in the importance they attribute to policy y. Recall that in the benchmark model, m reflected the relative importance of policy y from the voters' perspective. Now, a voter either belongs to the group that intrinsically cares less about policy y with m_L (with probability p) or more with m_H (with probability $1-p$), where $m_L < m < m_H$.

Solving the model for such parameters shows that the optimal policy choices of the candidates depend only on the average relative importance of policy y in the society, namely $pm_L + (1-p)m_H$. How the optimal policy choices and winning probabilities change is then the same question as the comparative statics with respect to m. Interestingly, the candidates' positions are not affected as long as the weighted average of relative importance remains the same in the electorate.

2.6.3 Mobilization

In line with the probabilistic voting model, our agents simply vote for the candidate they like better. However, we could also consider a scenario where voters do not simply go to the ballot box. Instead, similar to Coate, Conlin, and Moro (2008), they might require the election to be sufficiently important. The importance of the election can be reflected in the utility difference between the two candidates. The next proposition suggests that as the salience bias gets stronger, no abstention is ensured and all voters indeed vote.

Proposition 2.4. *Suppose voting is costly, and voters vote if and only if the utility difference they get from both candidates exceeds the cost of voting. If the cost of voting is bounded from above, i.e.* $c_v < \infty$, $\exists \underline{\delta} < \infty$ *such that for all* $\delta > \underline{\delta}$ *everybody in the electorate votes.*

The above proposition suggests that, apart from affecting candidate positioning, salience bias can also be a factor that incentivizes people to vote. Therefore, increasing the salience of an issue can be used as a tool to increase voter turnout.

2.7 Experiment

As our theoretical framework suggests plausible dynamics, we also conduct a supplementary experiment to test whether the implications are applicable in real life. Namely, in the experiment, we test whether the prediction of the model about the positive effect of extremism on the vote share hold.

The main goal of the experiment is to investigate two conjectures of the model. First, we check if a candidate can gain more votes by choosing an extreme policy. Secondly, we assess attention as the main driver of policy preferences and voting decisions.

2.7.1 Experimental Design

The experiment is in a survey format. Each participant answers simple questions using the online platform. Our main goal is to test the predictions of the model in a stylized context. Specifically, participants are presented with a hypothetical election scenario and asked to vote for one of the two candidates. The positions of the hypothetical candidates regarding climate and defense policies are either extreme or moderate (2x2 design). The experiment is in a between-subject design, hence subjects are only aware of a single scenario. The timeline of the experiment is as follows:

(1) Demographics: In this part, we ask simple demographic questions about age, gender, education, employment, city of residence, and per-person income in the household.

(2) Political Engagement: We use agreement with four statements to measure general interest in politics. The statements are about following the news, attachment to an ideology, being influenced by the election polls, and regular voting. We also ask participants whether they have ever voted and if they are registered members of any political party.

(3) Issue Ranking: We ask them to rank political issues such as health services, economic stability, and freedom of speech according to subjective importance. We mainly focus on the ranking of climate and defense policies.

(4) Voting: We present hypothetical candidates (A and B) and ask participants to vote for one. They see the information about the verbal proposals of candidates on climate and defense policies, in addition to their age, gender, education, and family status. For both candidates, climate and defense policy can either be extreme or moderate. Treatment manipulation is implemented here.

(5) Key Factors: We ask participants to state the factors that were crucial for their voting choice in the previous question. We use this question to detect the salient issues.

(6) Donation: Participants are asked to divide 10.000 Turkish Liras among two charities. One participant is going to be randomly selected and her choice of donations is implemented. One charity (TEMA) is one of the biggest associations in Turkey that focus on the environment, whereas the other charity supports the war veterans and families of martyrs. The donations would reflect the importance of climate and defense policies, respectively.

Participants will be randomly allocated to one of the four (2×2) treatments differing only in the voting question:

- Moderate-Moderate (MM) Treatment: There are 2 candidates and they have moderate proposals on both climate and defense policies.
- Extreme-Moderate (EM) Treatment: There are two candidates and they have extreme and opposing views on climate policies such that one promises urgent solutions to the climate crisis and the other does not find it necessary to take any action. Defense proposals are moderate.
- Extreme-Moderate (ME) Treatment: There are two candidates and they have extreme and opposing views on defense policies such that one considers border security as a top priority issue and the other does not attach much importance to it. Climate proposals are moderate.
- Extreme-Extreme (EE) Treatment: There are two candidates and they have extreme and opposing views on climate policies. Defense proposals are moderate.

2.7.2 Experimental Results

The experiment is conducted with 604 participants in September 2022 in Turkey with a representative sample of the country's adult population in terms of geographical region, age, gender, and socio-economical status. The data is collected by a third-party company to reach a representative subject pool. We conduct the experiment in Turkey because the political conjuncture is similar to our model environment where presidential elections are run with two opposing candidates. The experiment takes around 10 minutes and the participation fee is 4 Euros.

The main result of the experiment is in line with the model prediction such that the vote share of a candidate increases as she takes more extreme positions in her strong policy. As you can see in Table 2.1, participants are more likely to vote for the climate-oriented candidate (Candidate B) when climate policy proposals are extremely different, and vice-versa.

The second result of the experiment is about the underlying channel of this effect. As shown in Table 2.2, people who report that they considered climate

Table 2.1. OLS regression of voting for candidate B on treatment variations. The baseline is the MM treatment in the first two regressions.

	Vote for climate-oriented candidate			
	votes B	votes B	votes B	votes B
EM	0.185***	0.165***		
	(0.0544)	(0.0531)		
ME	-0.119**	-0.128**		
	(0.0544)	(0.0531)		
EE	0.0199	-0.00344		
	(0.0544)	(0.0530)		
extreme climate			0.162***	0.145***
			(0.0384)	(0.0374)
extreme defense			-0.142***	-0.149***
			(0.0384)	(0.0376)
constant	0.351***	0.029***	0.363***	0.033***
	(0.0385)	(0.241)	(0.0333)	(0.241)
N	604	604	604	604
Control vars.		✓		✓

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

proposals while voting have a higher probability of voting for the climate-oriented candidate, and the opposite is true for defense proposals. Crucially, the coefficients are similar when we control for the importance of those policies before the voting decision. Hence, paying more attention to a policy increases the likelihood of voting for the stronger candidate in that policy.

Table 2.2. OLS regression of voting for candidate B on indicators of (self-reported) considered policies and donation for the environmental charity.

	Vote for climate-oriented candidate			
	votes B	votes B	votes B	votes B
considered climate	0.210***	0.185***		
	(0.0398)	(0.0401)		
considered defense	-0.268***	-0.235***		
	(0.0388)	(0.0399)		
donation for climate			0.0331***	0.0277***
			(0.0107)	(0.0106)
constant	1.471***	1.077***	1.213***	0.881***
	(0.0351)	(0.233)	(0.0553)	(0.246)
N	604	604	604	604
Control vars.		✓		✓

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Both findings support the implications of our model. Politicians can attract voters via choosing extreme positions in a policy and they achieve that by drawing the attention of the voters to that specific policy. These results suggest that extremist policies can arise as a way to stand out in the competition and catch the attention of voters.

2.8 Conclusion

We provided a model that explains the mechanism behind the extreme policy proposals by electoral candidates. We assume voters involuntarily pay more attention to issues where candidates take extreme positions and overstate the importance of those salient issues. As a result, intrinsically differentiated politicians exploit this bias by strategically positioning themselves in extreme positions and trying to attract attention to their strong issues.

This model shows the top-to-bottom process of extremism and polarization. Unlike the existing studies, an already polarized vote base is not a necessary condition, and all the results hold for homogeneous voters. We also show that the supply-driven extremism that we propose gets exacerbated if there is an already existing polarization in the society. Hence, the results of this paper can also be seen as a multiplier of previous findings on extremism.

Additionally, our model clearly shows the effect of extremism on the mobilization of voters. When the candidates take extreme positions to exploit salience bias, the utility difference between them for the voter gets larger. This creates an extra incentive for individuals to vote which leads to higher turnout.

The model can also be used to analyze the second-round elections. If a moderate candidate is chosen by the opposition party, extremist politicians in the opposition can help her to gain votes by manipulating the salient issues. For instance, in the 2020 US presidential elections, more radical politicians such as Bernie Sanders and Elizabeth Warren may have had a positive impact on Joe Biden by attracting attention to some issues different than Donald Trump's campaign.

We also conducted an experiment with a representative sample to test the predictions of the theory. The results of the experiment provide supportive evidence for our model. The vote share of candidates increases when they take extreme positions and the salience of their strong issues is the main channel of this increase.

The natural next step in this line of research would be investigating the ways to combat this supply-driven extremism. Raising awareness about those strategies and more informative media consumption are likely promising channels but their analysis is beyond the scope of this paper.

Appendix 2.A Appendix

2.A.1 Proofs

Proof of Lemma 2.1

Proof.

$$\frac{|x_A - \bar{x}|}{\bar{x}} > \frac{|y_A - \bar{y}|}{\bar{y}} \iff \frac{|x_A - \frac{x_A+x_B}{2}|}{\bar{x}} > \frac{|y_A - \frac{y_A+y_B}{2}|}{\bar{y}} \tag{2.A.1}$$

$$\iff \frac{|\frac{x_A-x_B}{2}|}{\bar{x}} > \frac{|\frac{y_A-y_B}{2}|}{\bar{y}} \iff \frac{|\frac{x_B-x_A}{2}|}{\bar{x}} > \frac{|\frac{y_B-y_A}{2}|}{\bar{y}} \tag{2.A.2}$$

$$\iff \frac{|x_B - \frac{x_A+x_B}{2}|}{\bar{x}} > \frac{|y_B - \frac{y_A+y_B}{2}|}{\bar{y}} \iff \frac{|x_B - \bar{x}|}{\bar{x}} > \frac{|y_B - \bar{y}|}{\bar{y}} \tag{2.A.3}$$

□

Values in y-salient equilibrium: In a y-salient equilibrium, the optimally chosen policy profiles of both candidates are as in the following table, and the equilibrium indeed is x-salient iff $\frac{c_y^A}{c_y^B} > \frac{c_x^B}{c_x^A}$.

$x_A^* = \frac{G}{(1+\delta m)c_x^A}$	$x_B^* = \frac{G}{(1+\delta m)c_x^B}$
$y_A^* = \frac{G\delta m}{(1+\delta m)c_y^A}$	$y_B^* = \frac{G\delta m}{(1+\delta m)c_y^B}$

Proof of Proposition 2.3:
At the initial positioning without candidate C, policy x is salient for both candidates. However, with the introduction of candidate C, different policies may become salient for both candidates. The assumptions $h > \varepsilon > 0$ and $\frac{h}{\bar{x}} > \frac{\varepsilon}{\bar{y}}$ ensure that both policies are positive values initially. Furthermore, $\bar{x} > h > \frac{\bar{x}}{3}$ ensures x is salient for candidate B even after C comes on stage.

For ω is large enough ($\omega > \frac{4\bar{x}\bar{y}+6\bar{y}h+6\bar{x}\varepsilon}{3h-\bar{x}}$ and $\omega > \frac{2\bar{y}h-2\bar{x}\varepsilon}{\bar{x}-h}$), policy y becomes salient for candidate A, in which A proposes a higher budget than B. Since candidate C offers 0 in policy x, this candidate does not attract any votes. Then, candidate A benefits from the introduction of C if the utility it creates with policy y is larger than the utility created by the proposal for x.

Polarization in the Electorate:
Suppose that a voter either has m_L with probability p or m_H with probability $(1-p)$. Note that the salience is not affected by m values. Therefore, the valuation for both types is as follows:

$$v_L(i) = \begin{cases} \delta \ln x_i + m_L \ln y_i & \text{if } \frac{|x_i-\bar{x}|}{\bar{x}} > \frac{|y_i-\bar{y}|}{\bar{y}} \\ \ln x_i + m_L \ln y_i & \text{if } \frac{|x_i-\bar{x}|}{\bar{x}} = \frac{|y_i-\bar{y}|}{\bar{y}} \\ \ln x_i + \delta m_L \ln y_i & \text{if } \frac{|x_i-\bar{x}|}{\bar{x}} < \frac{|y_i-\bar{y}|}{\bar{y}} \end{cases}$$

$$v_H(i) = \begin{cases} \delta \ln x_i + m_H \ln y_i & \text{if } \frac{|x_i - \bar{x}|}{\bar{x}} > \frac{|y_i - \bar{y}|}{\bar{y}} \\ \ln x_i + m_H \ln y_i & \text{if } \frac{|x_i - \bar{x}|}{\bar{x}} = \frac{|y_i - \bar{y}|}{\bar{y}} \\ \ln x_i + \delta m_H \ln y_i & \text{if } \frac{|x_i - \bar{x}|}{\bar{x}} < \frac{|y_i - \bar{y}|}{\bar{y}} \end{cases}$$

Among the voters with m_L, voters with $\tilde{\beta_L} = v_L(A) - v_L(B) + \beta + \epsilon$ vote for A and among the voters with m_H, voters with $\tilde{\beta_H} = v_H(A) - v_H(B) + \beta + \epsilon$ vote for A.

Hence, vote share of A boils down to $\phi[p\tilde{\beta_L} + (1-p)\tilde{\beta_H}] + \frac{1}{2}$, which turns A's winning probability into:

$$[p(v_L(A) - v_L(B)) + (1-p)(v_H(A) - v_H(B))]\,\varphi + \frac{1}{2}$$

Therefore, A's problem becomes a weighted average:

$$\max_{\{x_A, y_A\}} [pv_L(A) + (1-p)v_H(A)] \tag{2.A.4}$$

$$\text{s.t. } c_x^A x_A + c_y^A y_A = G \tag{2.A.5}$$

In return, this leads to a replacement of m in the original problem by $pm_L + (1-p)m_H$ in the optimality conditions. Nothing else changes.

Proof of Proposition 2.4:

Suppose we are in an x-salient equilibrium. The utility difference that a voter gets from both candidates is formulated as follows:

$$|\delta lnx_A + mlny_A - \delta lnx_B - mlny_B|$$

Plugging in the equilibrium policy choices of both candidates yield

$$|\delta ln\frac{c_x^B}{c_x^A} + m\frac{c_y^B}{c_y^A}|$$

We know that $\delta \geq 1$ and $m > 0$. Because $c_x^B > c_x^A$ and $c_y^B < c_y^A$, the first term is positive and the latter is negative. If $\delta ln\frac{c_x^B}{>}c_x^A > mln\frac{c_y^A}{<}c_y^B$, the whole term in absolute value is positive and therefore increases with δ.

References

Aragonès, Enriqueta, and Clara Ponsatí. 2022. “Shocks to issue salience and electoral competition.” *Economics of Governance* 23 (1). https://doi.org/10.1007/s10101-022-00267-0. [42]

Aygün, Orhan, and Tayfun Sönmez. 2013. “Matching with Contracts: Comment.” *American Economic Review* 103 (5): 2050–51. https://doi.org/10.1257/AER.103.5.2050. [46]

Balart, Pau, Agustin Casas, and Orestis Troumpounis. 2022. “Technological change, campaign spending and polarization.” *Journal of Public Economics* 211: 104666. https://doi.org/10.1016/J.JPUBECO.2022.104666. [42]

BBC News. 2017. “Trump orders wall to be built on Mexico border.” *BBC News* (26, 2017). Accessed January 30, 2022. https://www.bbc.com/news/world-us-canada-38740717. [39]

Bonomi, Giampaolo, Nicola Gennaioli, and Guido Tabellini. 2021. “Identity, Beliefs, and Political Conflict.” *Quarterly Journal of Economics* 136 (4). https://doi.org/10.1093/qje/qjab034. [42]

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2012. “Salience theory of choice under risk.” *Quarterly Journal of Economics* 127 (3). https://doi.org/10.1093/qje/qjs018. [40]

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2013. “Salience and consumer choice.” *Journal of Political Economy* 121 (5). https://doi.org/10.1086/673885. [42]

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2016. “Competition for Attention.” *Review of Economic Studies* 83 (2). https://doi.org/10.1093/restud/rdv048. [42]

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2020. “Memory, Attention, and Choice.” *Quarterly Journal of Economics* 135 (3). https://doi.org/10.1093/qje/qjaa007. [42]

Bursztyn, Leonardo, Georgy Egorov, and Stefano Fiorin. 2020. “From Extreme to Mainstream: The Erosion of Social Norms.” *American Economic Review* 110 (11): 3522–48. https://doi.org/10.1257/AER.20171175. [41]

Cane, Violet, and R. Duncan Luce. 1960. “Individual Choice Behavior: A Theoretical Analysis.” *Journal of the Royal Statistical Society. Series A (General)* 123 (4). https://doi.org/10.2307/2343282. [46]

Carothers, Thomas, and Andrew O’Donohue. 2019. *Democracies Divided: The Global Challenge of Political Polarization.* Brookings Institution Press. [39]

Coate, Stephen, Michael Conlin, and Andrea Moro. 2008. “The performance of pivotal-voter models in small-scale elections: Evidence from Texas liquor referenda.” *Journal of Public Economics* 92 (3-4). https://doi.org/10.1016/j.jpubeco.2007.08.007. [47]

Colussi, Tommaso, Ingo E. Isphording, and Nico Pestel. 2021. “Minority Salience and Political Extremism.” *American Economic Journal: Applied Economics* 13 (3). https://doi.org/10.1257/app.20190703. [41]

Enke, Benjamin. 2020. “Moral values and voting.” *Journal of Political Economy* 128 (10). https://doi.org/10.1086/708857. [41]

Enke, Benjamin, Mattias Polborn, and Alex Wu. 2022. “Values as Luxury Goods and Political Polarization.” *SSRN Electronic Journal,* https://doi.org/10.2139/ssrn.4098323. [41]

Enke, Benjamin, Ricardo Rodríguez-Padilla, and Florian Zimmermann. 2022. “Moral Universalism and the Structure of Ideology.” *Review of Economic Studies,* https://doi.org/10.1093/restud/rdac066. [41]

Glaeser, Edward L. 2005. “The political economy of hatred.” *Quarterly Journal of Economics* 120 (1). https://doi.org/10.1162/0033553053327434. [42]

Grossman, Gene M., and Elhanan Helpman. 2021. "Identity Politics and Trade Policy." *Review of Economic Studies* 88 (3): 1101–26. https://doi.org/10.1093/RESTUD/RDAA031. [41]

Herweg, Fabian, Daniel Müller, and Philipp Weinschenk. 2018. "Salience in markets." *Handbook of Behavioral Industrial Organization,* 75–113. https://doi.org/10.4337/9781784718985.00010. [42]

Jones, Matthew I., Antonio D. Sirianni, and Feng Fu. 2022. "Polarization, abstention, and the median voter theorem." *Humanities and Social Sciences Communications* 9 (1). https://doi.org/10.1057/s41599-022-01056-0. [41]

Kuziemko, Ilyana, and Ebonya Washington. 2018. "Why did the democrats lose the south? Bringing new data to an old debate." *American Economic Review* 108 (10). https://doi.org/10.1257/aer.20161413. [41]

Lewandowsky, Stephan, Michael Jetter, and Ullrich K.H. Ecker. 2020. "Using the president's tweets to understand political diversion in the age of social media." *Nature Communications* 11 (1). https://doi.org/10.1038/s41467-020-19644-6. [42]

Matějka, Filip, and Guido Tabellini. 2021. "Electoral Competition with Rationally Inattentive Voters." *Journal of the European Economic Association* 19 (3). https://doi.org/10.1093/jeea/jvaa042. [41]

Nunnari, Salvatore, and Jan Zápal. 2017. "Dynamic elections and ideological polarization." *Political Analysis* 25 (4). https://doi.org/10.1017/pan.2017.24. [41]

Persson, Torsten, and Guido Tabellini. 2002. *Political Economics: Explaining Economic Policy.* MIT Press. [40, 42]

Rodrik, Dani. 2021. "Why Does Globalization Fuel Populism? Economics, Culture, and the Rise of Right-Wing Populism." *Annual Review of Economics* 13. https://doi.org/10.1146/annurev-economics-070220-032416. [41]

Spirig, Judith. 2023. "When Issue Salience Affects Adjudication: Evidence from Swiss Asylum Appeal Decisions." *American Journal of Political Science* 67 (1). https://doi.org/10.1111/ajps.12612. [42]

Chapter 3

Learning with State-Dependent Preferences[*]

3.1 Introduction

State-dependent utilities are prevalent in almost all economic decisions. From consumption to labor or finance, the benefits of different options depend on exogenous factors such as the health of the consumer, other people's actions, or macroeconomic situations. The extensive literature on learning with state-dependent preferences repeatedly shows that individuals systematically underestimate the effect of the decision environment on their utility and attribute the impact of the state to the good or service they choose[1]. This causes them to make wasteful purchases or investments that are affected by random factors at the time of the decision. Yet, the dynamics of these errors and learning pattern has not been studied in detail before. Most studies show a bias in a single experience and extend it to multiple experience cases.

Nevertheless, understanding the mechanism of updating with state-dependent preferences and the long-term learning efficacy is crucial for designing the right policies to protect consumers, workers, and investors. For instance, in many countries, customers have a right of withdrawal from any transactions within a certain period of time. This interval allows them to understand the quality of the goods or services they purchased and return them if they are not satisfied. However, the policy would be ineffective if people make inaccurate inferences about the quality. Even with customers eventually estimating the quality correctly, the duration for the right of withdrawal may depend on the pace of learning. With better insight

[*] I am thankful to Florian Zimmermann and Thomas Dohmen for helpful comments. Funding by briq Institute on Behavior and Inequality and by Bonn Graduate School of Economics is gratefully acknowledged.

[1] These types of systematic errors are called projection bias or attribution bias. The descriptions and differences of these two biases are explained in Section 3.2

into the bias, we can determine the duration more optimally which would protect the customer with a minimum burden on sellers.

In this study, I conduct a lab experiment where participants try a good twice, one week apart such that the consumption states are altered as a treatment variation. By measuring their valuations for the good repeatedly in an incentive-compatible way, I analyze the dynamics of updating and learning at the limit. The experimental results show that while valuations are affected by the consumption state initially, individuals eventually learn the impact of the state that is manipulated in the experiment. Additionally, I investigate the underlying mechanism of the bias and show the importance of attention and salience. Although participants can distinguish the effect of the treatment manipulation of a state, the bias on other less salient states remains even after multiple experiences. Therefore, awareness about the relevant states for consumption is the key factor for debiasing and correct updating.

Recently, some studies show that consumption states cause economic agents to form biased valuations in various contexts. For instance, Haggag et al. (2019) conduct an experiment and show that the evaluation of a drink is heavily affected by the thirst levels of consumers at the time of trial. They also show that bad weather during a visit to a famous theme park drastically reduces the probability of visiting again. Similarly, Chang, Huang, and Wang (2018) show that daily fluctuations in air pollution are positively correlated with health insurance purchases, and Simonsohn (2010) shows the causal effect of cloudy weather during campus visits on a higher likelihood of selecting academic-oriented colleges. Likewise, Busse et al. (2015) demonstrate that the sales of convertible cars increase during sunny days whereas SUV sales increase on rainy days even though the irrelevance of the weather on demo day on future utility. Furthermore, Haggag et al. (2021) show that small details such as the schedule of the introductory course in the first year of college can influence large stake decisions such as major choice. Therefore, understanding the dynamics of learning with state-dependent preferences is essential for preventing substantial mistakes in many critical decisions. However, most of the empirical evidence mentioned above is based on one-shot decisions. To be able to understand the underlying mechanism of the bias, we would need to analyze the evaluations repeatedly in controlled environments.

Lab experiments provide an ideal setup to investigate the effect of experience state and update patterns of individuals since all the relevant factors can be supervised and regulated. Hence, I conducted a lab experiment similar to Haggag et al. (2019) and modified it for a multiple experience framework. Likewise, I look at the effect of thirst level on the evaluation of a drink. As a first experience, participants come to the lab and taste some fruit juice. Before the trial, one group is asked to drink a low amount of water (100 ml) and the other group drinks a high amount of water (500 ml). One week later, they come to the lab and try the same juice again. This time, all the participants are asked to drink 500 ml of water beforehand. The participants' willingness to pay (WTP) for a box of juice (to be delivered at the end of the second week) is measured before and after each trial. By comparing the WTP measures of the low & high water group (LH) to the high & high water group (HH), we can see the effect of changing states on valuations.

The first main result of the experiment shows the effectiveness of the treatment manipulation and replicates the previous findings about the impact of the current state on the valuation of the good. The mean of the first WTP measure is 28.9 Euro cents for the high water group whereas it is 37.1 Euro cents for the low water group. Drinking 500 ml water instead of only 100 ml creates a 28% difference in initial valuations since participants reflect the current lack of appetite onto their future consumption.

The second result is about the long-run learning pattern. The mean WTP measures at the end of the second week are 25.4 Euro cents for the LH group and 27.7 Euro cents for the HH group. The huge gap in initial valuations disappears and even slightly reverses after multiple experiences. Since there is no statistical difference between the WTP measures of the two groups, we can deduce that individuals eventually learn the impact of water consumption on their utility and adjust for their satiation.

The last result sheds light on the mechanism of the bias. Together with a short questionnaire about the current physiological conditions, participants are also asked about their thirst levels at the beginning of each session. As noted above, the effect of treatment manipulation, i.e. drinking less or more water, diminishes after multiple experiences. Nonetheless, the effect of original thirst levels on valuations remains strong and stable throughout all measures. Participants that indicate an above the median thirst level at the beginning of the second session are willing to pay 33.1 cents while other participants are willing to pay only 15.7 cents. This more than double difference in valuation stems from the fact that individuals cannot pinpoint the impact of the initial satiation level since it is not as salient as drinking 500 ml water. Thus, the bias from a consumption state vanishes after multiple experiences only if agents pay attention to it.

On top of these main findings, I compare different models of learning with state-dependent preferences and test their predictions. Two main models at focus are the model suggested in Haggag et al. (2019) (HP) and the model with reference dependent utilities by Gagnon-Bartsch and Bushong (2022) (GBB). The key difference between these models is the opposite effects of past experiences. In HP, current evaluations are biased toward past experiences. Hence, a good memory affects the impression of quality positively. On the other hand, agents in GBB compare their experiences with their expectations. Thus, a good first experience can lead to lower evaluations in the future by creating an over-optimistic reference point.

The experimental data contradicts the main predictions of the HP whereas it is in line with GBB. Contrary to HP, better initial valuations do not lead to a higher willingness to pay after the second experience and agents update their evaluations even under constant states. I also analyze the standard benchmark models of purely naive or fully sophisticated agents and show that they cannot explain the experimental findings.

I describe the theory and summarize two main models in Section 3.2. Experimental design is explained in Section 3.3 and results are provided in Section 3.4. I compare different models in Section 3.5 and conclude in Section 3.6.

3.2 Theory

In many economic models, it is assumed that agents are aware of their preferences in each state. In other words, the utility of good q in state s, $u(x|s)$ is known by the individual herself. This is broadly based on two presumptions. First, the value of good q should be known. Second, the effect of the state on the enjoyment of the good should be known. However, the separate effects of these two components are usually not observed and the individuals only recognize the composite utility of a good at a state. In other words, observing $u(q|s)$ may not be enough to identify q, s, and their interaction. As a result, it becomes hard to disentangle the effect of the state and the value of the good itself. Furthermore, this becomes especially difficult when individuals do not have experience in several different states.

Starting with Loewenstein, O'Donoghue, and Rabin (2003), many studies have documented that people make systematic mistakes in assessing the impact of the state and incorrectly forecast the utility in another state. This systematic error is called *projection bias* and it is shown to be effective in many daily decisions like car purchases, education, and even medical choices. More recently, another bias called *attribution bias* has been suggested by Haggag et al. (2019) to explain a similar phenomenon of underestimating the effect of the state and attributing some of its impact to the good itself[2]. Despite the close connection between them, it is useful to differentiate the distinct features of projection bias and attribution bias.

The key aspect of projection bias is being future-oriented. It focuses on the errors in estimating the utility in another state. On the other hand, misattribution is related mostly to past events and it occurs when evaluating the quality of a good after an experience. Of course, forecasting future utility depends on the evaluation of past experiences, and in that sense, investigating attribution bias can be seen as a natural step forward from the projection bias. In this paper, I focus on attribution bias but almost all findings can be applied to projection bias.

[2] In psychology, there is a huge literature on attribution theory. However, it mostly focuses on interpersonal relations and is not very relevant in our context. For a recent review with a chronological summary see Weiner (2010). Also, Gawronski (2004) discusses the findings of attribution literature.

To clarify the concept, it is worth mentioning what does the value of a good or its *true quality* mean as any consumption experience has to happen in some state. Any consumption experience has to happen in some state. As an example, whenever we drink some wine, it is always in some ambiance either in a restaurant or at home. However, we would expect a dominance in quality conditional on each state. We can think that a more delicious wine is always preferable to a bad one wherever the consumption happens. Nevertheless, it would be difficult to choose between the two wines if we tried them in different states. Since states are usually numerous (or even continuous variables such as heat, hunger, thirst, etc.), it is almost impossible to try two goods or services under exactly the same conditions. Thus, understanding the effect of the state is crucial in forming preferences.

Alternatively, one can think of the value of a good as the expected utility one gets from it. When buying a product (especially a durable good), people are often uncertain about in which state they are going to consume it. Therefore, they form an expectation of utility in various states. Hence, this expected enjoyment could be considered as the value of a good or the utility in a *neutral state*. Moreover, when people share experiences among themselves, it is not always possible to describe utility in various different states. This is mainly because of too many possible states or the fact that a medium does not always allow for a detailed explanation. For instance, the rating systems in restaurant review platforms or online markets are generally constructed as one-dimensional such as a 5-star scale. Hence, customers are expected to summarize their experiences as a single number which is believed to indicate the *true quality* of the good or service.

The aspect of attribution bias that this study focuses on is the learning mechanism. One may argue that misattribution is only relevant in cases with little experience. According to this "learning argument", people would eventually figure out the effect of the state on utility and the correct form of their utility function as they are exposed to more observations. However, recent studies revealed that this may not be the case. Haggag et al. (2019) show that visitors of a theme park mistakenly attribute weather conditions to the quality of the park even if they are locals or they have visited the park at least six times. Gagnon-Bartsch and Bushong (2022) argued that having experiences in opposite states can make the signals so volatile that beliefs would be inconsistent at the limit. Therefore, investigating attribution bias helps us understand not only rare decisions with little experience but also long-run behaviors with repetitive choices. For that purpose, identifying the underlying mechanism of the misattribution process would clarify the learning pattern, if it exists at all.

In the following subsections, I will briefly analyze two recent misattribution models by Haggag et al. (2019) and Gagnon-Bartsch and Bushong (2022) to test their predictions with experimental data.

3.2.1 Model I: Misattribution with Imperfect Adjustment

With the same essence of *simple projection bias model* in Loewenstein, O'Donoghue, and Rabin (2003), Haggag et al. (2019) propose a model of misattribution where agents have state dependent preferences $u(c, s_t)$ in which c denotes consumption good and s_t denotes the state the consumption takes place. In the model, the state is left intentionally broad to capture any internal factors such as hunger, thirst, or mood and also external factors like weather, or noise. To be more precise, state s_t can be thought of as a vector of all the relevant aspects of the consumption experience other than the good itself. [3]

After observing the utility from the past experience $u(c, s_{t-1})$ in state s_{t-1}, the consumer tries to predict (current or future) utility at state s_t. However, she cannot fully adjust the effect of the state and her prediction falls between the utility in actual and past states. Formally, there exists $\gamma \in [0, 1]$ such that for all c, s_{t-1} and s_t; $u(c, s_t | s_{t-1}) = \gamma u(c, s_{t-1}) + (1-\gamma) u(c, s_t)$. Higher γ indicates a higher bias since the prediction is closer to past experience rather than the correct estimation in the new state. Hence, for an agent with full bias, γ equals 1, and similarly for an agent who is unaffected by the bias γ is zero.

They also suggest an extension of the model with multiple prior experiences. After experiencing consumption utility $t-1$ times in the past, the prediction for the current utility becomes

$$u(c, s_t | s_1, s_2, ..., s_{t-1}) = (1-\gamma) u(c, s_t) + \gamma \frac{1}{t-1} \sum_{\tau=1}^{t-1} \delta_\tau u(c, s_{t-\tau})$$

where $\delta_\tau \in [0, 1]$ is a discount rate that is possibly period-specific and might be context specific. Some natural discount functions they mention are equal discounting with $\delta_\tau = 1$ for any τ, exponential discounting where $\delta_\tau = \delta^\tau$ or a discount function with a specific emphasis on the first and/or the last experience such as $\delta_1 > 0, \delta_{t-1} > 0$ and $\delta_2 = ... = \delta_{t-2} = 0$.[4]

3.2.2 Model II: Misattribution with Reference Dependence

The misattribution model of Gagnon-Bartsch and Bushong (2022) is based on the reference dependent preferences in Koszegi and Rabin (2007). Agent's consumption utility x_t follows a stochastic process depending on the parameter θ such

[3] In the simple projection bias of Loewenstein, O'Donoghue, and Rabin (2003), agents try to predict utility in the future and often in uncertain state s' based on their past or current state s. Whereas in attribution bias models, agents are certain about the state s' and still cannot predict the utility correctly. With this point, the attribution bias model of Haggag et al. (2019) clearly separates itself from the effects of uncertainty about the state and hence any belief structure for the future.

[4] The dominant effect of the first experience is called the primacy effect and similarly, the dominance of the last experience is called the recency effect.

that $x_t = \theta + \epsilon_t$ where ϵ_t is assumed to have a normal distribution $N(0, \sigma^2)$ with known variance.[5] Hence, the parameter θ denotes the average consumption utility of the good or its quality, and ϵ_t is the effect of the state. However, the agent does not know the parameter θ and has a prior belief $\theta \sim N(\theta_0, \rho^2)$.

In the model, the experienced utility of an agent is assumed to have two additively separable parts. The first part is consumption utility (denoted as x above) and the second part is gain-loss utility stemming from the comparison of consumption utility to a reference. Specifically, the reference is assumed to be the expected level of x, denoted as $\hat{\theta}$. Therefore, total experience utility is $u(x|\hat{\theta}) = x + \eta n(x|\hat{\theta})$ where the exact functional form of gain-loss utility $n(\cdot|\cdot)$ is as follows

$$n(x|\hat{\theta}) = \begin{cases} x - \hat{\theta} & \textit{if} \quad x \geq \hat{\theta} \\ \lambda(x - \hat{\theta}) & \textit{if} \quad x < \hat{\theta} \end{cases}$$

where $\eta > 0$ is the relative strength of gain-loss utility compared to consumption utility and $\lambda \geq 1$ allows for loss aversion i.e., a loss gives bigger disutility than the utility of same level gain.

After correctly observing total utility u_t, the agent updates her belief about the parameter θ using Bayes' rule. However, the agent underestimates the effect of gain-loss utility such that she thinks parameter as $\hat{\eta} < \eta$. Therefore, she infers $u_t = \hat{u}(\hat{x}_t|\hat{\theta}_{t-1}) = \hat{x}_t + \hat{\eta} n(\hat{x}_t|\hat{\theta}_{t-1})$ instead of calculating with the correct parameter η. Therefore, incorrectly encoded outcome $\hat{x}_t$ can be calculated as

$$\hat{x}_t = \begin{cases} x_t + \kappa^G(x_t - \hat{\theta}_{t-1}) & \textit{if} \quad x_t \geq \hat{\theta}_{t-1} \\ x_t + \kappa^L(x_t - \hat{\theta}_{t-1}) & \textit{if} \quad x_t < \hat{\theta}_{t-1} \end{cases}$$

with $\kappa^G = \left(\frac{\eta - \hat{\eta}}{1+\hat{\eta}}\right)$ and $\kappa^L = \lambda\left(\frac{\eta - \hat{\eta}}{1+\hat{\eta}\lambda}\right)$. Here, κ^G and κ^L show the relative size of the error in gain and loss domain respectively. Hence, the agent overestimates the consumption utility if the outcome exceeds the expectation and underestimates it for outcomes below expectation. Note that for $\lambda > 1$, there is an asymmetry in gain and loss utilities since $\kappa^L > \kappa^G$. Therefore, the error in interpreting signal is higher for a surprisingly low experience $x_t < \hat{\theta}_{t-1}$ compared to the error when $x_t \geq \hat{\theta}_{t-1}$. This asymmetry implies a negatively biased estimate for x and consequently for θ when agents observed many experiences within gain and loss domains.

For an unbiased and fully rational agent, extending belief updating to multiple prior experiences gives a recursive formula for $\hat{\theta}_t$ as the estimate after t-th update such that $\hat{\theta}_t = \alpha_t \hat{x}_t + (1 - \alpha_t)\hat{\theta}^r_{t-1}$ where $\alpha_t = 1/(t + \frac{\sigma^2}{\rho^2})$ measures the informativeness of the signal and $\hat{\theta}^r_{t-1}$ is the prior rational belief before the last experience. On the other hand, a biased agent would use misencoded outcome $\hat{x}_t$

[5] This setup can be seen as a special case of the state-dependent preferences where the state is a random variable and its effect on utility is linear.

and biased prior belief $\hat{\theta}_{t-1}$ for the update. Thus, the biased updated belief of the misattributor becomes

$$\hat{\theta}_t = \alpha_t(1+\kappa_t)x_t + [1-\alpha_t(1+\kappa_t)]\hat{\theta}^r_{t-1}$$

where κ_t is either κ^G or κ^L depending on x_t being a gain or loss respectively.

3.2.3 Comparison of Two Models

Before going into the details of the comparison, we need some refinement in the model by Haggag et al. (2019). Authors claim that with several multiple experiences in a particular state, individuals learn the true value of a good in that state. They also present some evidence for it in their amusement park experiment. On the other hand, they did not specify a discounting pattern in their model. However, to be able to achieve convergence to the true valuation of a good in a state, we need to rule out some discounting parameter combinations.

Assumption 1: The sum of discount rates is equal to their quantity, i.e. for any $k \geq 1$, $\sum_{\tau=1}^{k} \delta_\tau = k$.

Without this assumption, their model can lead to unintentional results. Consider the trivial case of only one possible state. In their model, they assume that individuals observe their utilities in a certain state after they experience it. Hence, according to their model, an agent's evaluation for that good after trying it several times must be unbiased. Nevertheless, this is true only under Assumption 1. Under a fixed state $\bar{s}$, the predicted utility after k trials becomes

$$u(c,\bar{s}|s_1,s_2,...,s_k) = (1-\gamma)u(c,\bar{s}) + \gamma\frac{1}{k}\sum_{\tau=1}^{k}\delta_\tau u(c,\bar{s})$$

which cannot be equal to $u(c,\bar{s})$ unless Assumption 1 is satisfied. The main intuition for this assumption is considering discount parameters as weights. Without the sum of the discount rates being equal to the number of past periods, we would observe a general upward or downward shift in the evaluation of past experiences. However, this is not the intended feature of their model. Therefore, we will continue our comparison of two models under this assumption to keep the essence of the model in a useful frame.

To be able to see the differences between the two models we need a decision with at least two prior experiences because with only a single experience both models make the same prediction. According to both models, an individual who tried the good in a pleasant state (or equivalently who had a positive shock) overestimates the quality of it and believes that she will get a higher utility compared to the rational belief in all other states. Also, she will underestimate the good after having a single bad experience (or a negative shock).

An interesting situation happens when we allow multiple prior experiences. To compare the models in detail, we can look at the predictions with two prior experiences. Assume an individual consumes good c at state s twice and let s' be any other state. According to HP model, the agent estimates the utility of the good after trying it once in state s so that for any other state s', $u(c,s'|s) = (1-\gamma)u(c,s') + \gamma u(c,s)$. After trying it once more in the same state s, the agent's evaluation becomes

$$u(c,s'|s,s) = (1-\gamma)u(c,s') + \gamma\frac{1}{2}(\delta_1 u(c,s) + \delta_2 u(c,s))$$

which is equal to $u(c,s'|s)$ since $\delta_1 + \delta_2 = 2$ under Assumption 1. Hence, HP predicts no change between the first and the second trials in the same state s because there is no new information to help in learning.

On the other hand, according to GBB, the agent estimates the utility of the good after trying it once in state s. For any reference $\hat{\theta}_0$, the assessment after the first trial is $\hat{x}_1 = x_s + \kappa_1(x_s - \hat{\theta}_0)$ where x_s is the consumption utility in state s. Also, κ_1 is equal κ^L if the consumption utility is below the initial expectation and equal to κ^G if the consumption utility exceeds the expectation. Since this assessment after the first trial is generically not equal to the initial expectation[6], reference is adjusted so that $\hat{\theta}_1 = \hat{\theta}_0 - \alpha_1(1+\kappa_1)(\hat{\theta}_0 - x_s)$. The new reference point $\hat{\theta}_1$ is either above or below the consumption utility in state s, that is x_s. Now, trying it again in the state s gives the second signal $\hat{x}_2 = x_s + \kappa_2(x_s - \hat{\theta}_1)$ where $\kappa_2 = \kappa^L$ if the expectation is above the low state level and $\kappa_2 = \kappa^G$ otherwise. Therefore, the reference is adjusted again with the Bayesian update rule. Hence, GBB predicts a change in the valuation of the good between the first and the second trials.

As a result, to identify the underlying model, we can compare the valuations of the individuals after the first and second trials in a fixed state. A result with no change between trials would support HP whereas any change would be a piece of evidence for GBB.

3.3 Experimental Design

The experimental procedure is similar to Haggag et al. (2019) and extended to a multiple-experience framework. Participants try the same drink twice, one week apart and their thirst level is manipulated as the treatment variance. This way, the effect of an underlying consumption state and the dynamics of updating can be observed. In each session, willingness to pay measures are collected for the juice they tried. Each session lasted approximately 15 minutes and the average earning was €14.8. For the experiment, zTree software by Fischbacher (2007) is used.

Session 1: At the first session, participants are informed that their willingness to pay (WTP) will be measured four times (two in each session), and only one

[6] 6. Theoretically, it is possible to have exactly $\hat{\theta}_0 = x_s$ but it is a zero probability event if the state is continuous. Therefore, this case is omitted.

of them will be effective with equal probability. First, each participant is asked to fill out a demographic survey and a detailed questionnaire about their current state at the beginning of the session before drinking anything in the lab. The questionnaire contains many queries such as the current thirst, hunger, mood, health, sleep levels, and their general affinity with fruit juices.

Later, they start the juice trial phase. At the beginning of this phase, they drink either 100 ml (low amount) or 500 ml (high amount) of water depending on the random group allocation. The number of participants in each group is roughly equal and the participants are not informed about the existence of another group. After this treatment variation which exogenously manipulates their thirst levels, their WTP for a 200 ml box of the juice is measured before they taste the drink. The multiple price list method is used for all WTP measures throughout the experiment with €0.1 increments from €0.1 to €1.[7] This first WTP is used as the initial expectation in a specific state later in the analysis. The drink used in the experiment is mango juice imported from Egypt, chosen purposefully as an unfamiliar drink to avoid the spillover effects from previous experiences.

After the initial WTP measure, participants are asked to drink half of the 160 ml glass of fruit juice and write a few sentences about the taste. This description task in the middle of the drinking is also used in Haggag et al. (2019) to make the experience more memorable[8]. Then, they finish the rest of the juice and their WTP is measured again.

Finally, they are asked to indicate “How much did you like the juice?” on an unmarked slider with “not at all” and “very much” expressions on each end of the slider. This measure is designed as a continuous measure of liking, independent of any monetary concerns. After this step, the first session is over.

Session 2: The second session is similar to the first session. After a short reminder of the experimental procedure, participants again start with a questionnaire about the same consumption states such as thirst, hunger, and sleep. Additionally, there is an additional query about their juice consumption between the two sessions. By this question, we can detect if subjects are influenced by any interaction with another juice. Later, they are asked to indicate “How much did you like the juice from last week?” on the same unmarked slider as in the first session. This question is designed to capture the memory of their previous experience.

Later, they are asked to drink water as in the first session. However, they all drink the same amount of 500 ml water this time. Hence, our two treatment groups are those who drink a high amount in both sessions (HH group) and those who drink a low amount of water in the first and a high amount in the second session (LH group). Following the water consumption, their WTP is measured for

[7] The market price for the 200 ml juice is €0.5.

[8] Apparently, the goal of choosing an unfamiliar drink is successful so that almost none of the subjects are able to guess that it was mango juice in this description task.

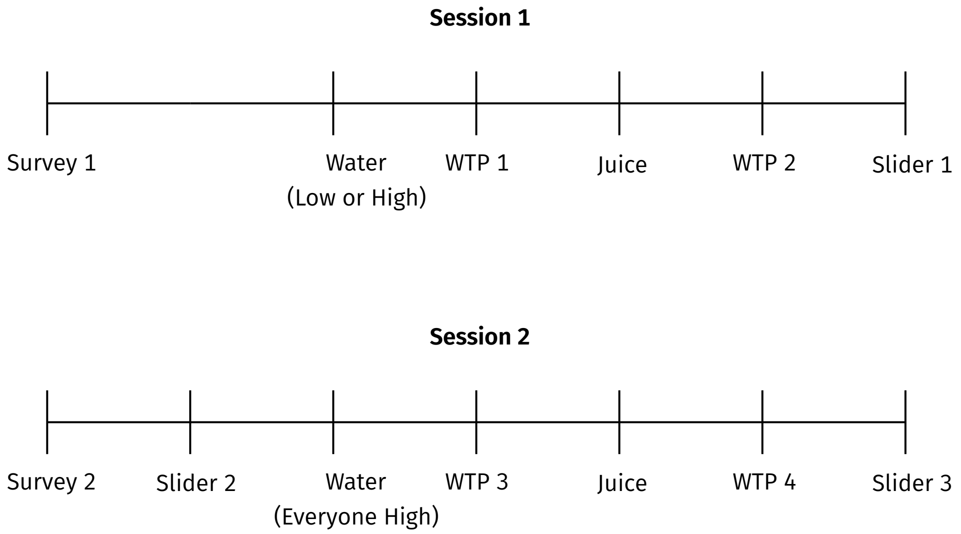

Figure 3.1. Timeline of the experiment

the third time using the same multiple price list method. Afterward, they drink half of the juice and write a short description of their experience again. Then, they finish the juice and their WTP is measured for the fourth and the last time. Then, they indicate their liking again on an unmarked slider.

Finally, they are asked a hypothetical question similar to Haggag et al. (2019). They answer whether they would like to drink another glass of juice at that moment if it was provided to them. This question enables us to understand their current enjoyment at that specific state. Then, the experiment is over. You can see the timeline of the experiment in Figure 3.1.

3.4 Results

The experiment was conducted with 168 participants in BonnEconLab in two waves in February and November 2020.[9] 129 participants attended the second session and attrition is balanced in treatment groups. The formal tests are provided in Appendix.

The results of the experiment show that the effect of the thirst level manipulation vanishes after multiple experiences although it is effective initially. In other words, participants can mostly disentangle the effect of water consumption on their utility. I will first describe the main outcomes of the experiment. Then, the findings about the underlying mechanism of the bias will follow.

[9] The outbreak of the COVID-19 pandemic caused the delay between the two rounds of data collection.

3.4.1 Main Outcomes

The first finding is that the treatment manipulation is effective. On average, participants who drink 100 ml of water are willing to pay 28% more in their first willingness to pay measures compared to those who drink 500 ml. As expected, their expected utility from consumption is shifted by exogenously changing their thirst level before the trial.

Result 1: WTP1 is higher in the low water group compared to the high water group.

Moreover, the effect size of the treatment manipulation is decreasing with weight and height. This decline is also expected since the same amount of water would shift the satiation less with increasing body sizes. You can see the regression results in Table 3.1.

Table 3.1. OLS Regression of the first WTP measure on treatment and personal characteristics.

	WTP1	WTP1	WTP1	WTP1
LH	8.214*	26.07**	127.2**	7.596*
	(4.340)	(11.07)	(56.61)	(4.441)
LH×weight (in kg)		-0.245*		
		(0.140)		
LH×height (in cm)			-0.686**	
			(0.325)	
Control vars.				✓
N	168	168	168	168

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In fact, the impact of the satiation manipulation is similar to the effect of the variation in the original thirst levels of the participants at the beginning of the session. The difference in WTP measures between low and high water consumption groups is comparable to the difference between participants with above and below median thirst levels initially. Hence, participants are willing to pay more for a box of fruit juice delivery next week as their current thirst level rises. This indicates that they cannot fully disentangle the effect of the current state from their evaluations. The summary statistics of the WTP1 measures are provided in Table 3.2.

Table 3.2. Mean willingness to pay measures in Euro cents by treatment and original thirst level.

	thirst≤median	thirst>median	Total
High water (500ml)	24.24	31.96	28.93
Low water (100ml)	32.81	39.81	37.14
Total	28.46	35.92	33.04

The second finding is about the evaluations after the second trial. WTP measures are not statistically different between the LH and HH groups. The effect of the thirst manipulation in the first WTP measure completely vanished and even slightly reversed after the second trial. Similarly, the outcome of the slider for liking after the second trial is not statistically different between the two groups. The formal tests are in Table 3.3 and the full list of mean WTP measures is in Table 3.4.

Result 2: There is no significant difference between the evaluations of the two groups after the second trial.

Table 3.3. OLS Regression of the WTP measure and the slider for liking after the second trial on treatment variation.

	WTP4	WTP4	Slider3	Slider3
Low water	-2.369	-2.919	0.728	2.034
	(5.104)	(5.065)	(5.367)	(5.502)
Control vars.		✓		✓
N	129	129	129	129

Standard errors in parentheses.

Table 3.4. Mean willingness to pay measures in Euro cents by treatment.

	WTP1	WTP2	WTP3	WTP4
Low & High (LH)	37.1	30	24.0	25.4
High & High (HH)	28.9	29.6	24.0	27.7

The results show that the manipulation of the thirst levels has no effect after the second trial. Thus, we can deduce that the participants can identify the effect of drinking water before trying the juice and form their evaluations accordingly.

3.4.2 Mechanism: The Effect of Salience

An interesting finding is about where individuals pay attention and how they update their valuations accordingly. In the experiment, there are two factors that affect the satiation level of participants. First, they arrive at the lab with different thirst levels and we ask about their original thirst levels at the beginning of each session. Secondly, the treatment variation (low vs. high water consumption) alters their initial satiation.

As mentioned in Section 3.4.1, the effect of treatment manipulation diminishes after the second trial. However, the original thirst levels stay effective throughout all willingness to pay measures. As shown in Table 3.5, a higher level of thirst at the beginning of a session leads to a higher willingness to pay for the fruit juice[10]. Similarly, a higher value on the slider for liking at the second session is associated with a higher thirst level at the beginning of that session.

Table 3.5. OLS Regression of the first and the last evaluations on treatment and the initial thirst levels at the beginning of the sessions.

	WTP1	WTP4	Slider3
LH	7.925*	-3.396	-0.135
	(4.297)	(4.872)	(5.223)
thirst 1	3.468**		
	(1.637)		
thirst 2		6.762***	5.680***
		(1.817)	(1.948)
N	168	129	129

Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Combining this result with Result 2 from Section 3.4.1, we can deduce that individuals can distinguish the effect of the state and form the right evaluations only if they pay attention to the relevant states[11]. In this experiment, they can identify the impact of drinking water before fruit juice consumption as a very salient factor. On the other hand, the original thirst level at the beginning of a session is a much less salient factor and participants do not adjust their expectations considering it. Thus, it causes biased evaluations even at the end of the second week. The participants who indicate an above the median thirst level at the beginning of the second session are willing to pay almost three times the participants with below the median thirst levels. You can see all mean willingness to pay measures by treatment and original thirst levels by the median split[12] in Table 3.6.

Table 3.6. Mean willingness to pay measures in Euro cents by treatment and thirst levels. t1 and t2 represent thirst levels at the beginning of sessions 1 and 2, respectively.

	WTP1		WTP2		WTP3		WTP4	
	t1≤4	t1>4	t1≤4	t1>4	t2≤4	t2>4	t2≤4	t2>4
LH	24.24	31.96	22.42	34.31	14.17	30.26	18.75	33.42
HH	32.81	39.81	22.5	34.62	13.6	30.24	12.8	32.86

[10] Since participants did not know about the water consumption in the experiment, they cannot strategically prepare for it. Hence, we can take the original thirst level (especially in the first session) as exogenous.

[11] The experimental data show a similar bias in willingness to drink an extra cup of fruit juice. Similarly, WTP measures are biased as a result of the mood of the participants. Individuals who indicate a better overall feeling report a higher willingness to pay in the second session (see Appendix for details.)

[12] The thirst level is measured on a scale from 1 to 7. The median is 4 in both sessions.

3.5 Model Comparison

In this section, I check the predictions of the models by Haggag et al. (2019) and Gagnon-Bartsch and Bushong (2022) to see whether the data is in line with them. I will also evaluate the results of this study in comparison to some benchmark models of fully sophisticated and fully naive agents. The results clearly show that the misattribution model by Haggag et al. (2019) fails to explain the experimental data in a multiple experience framework and the model by Gagnon-Bartsch and Bushong (2022) is in line with the data.

3.5.1 Model by Haggag et al. (2019)

The main insight of the model is that current evaluations are directly proportionate to past experiences. Hence, better states in the past lead to better evaluations in the future, and there will be no update if the state is constant. However, the data contradict these predictions and I present some counter evidence below.

Counter evidence 1. A better initial experience does not lead to a higher valuation: $WTP4_L H \ngtr WTP4_H H$.

We would expect a higher willingness to pay by participants who were in a more favorable state in their initial experience if evaluations are biased towards the past states. Nonetheless, the fourth measure of willingness to pay is even higher in the group with high water consumption in the first session. This shows that, contrary to common perception, a good first impression is not necessarily advantageous in further interactions.

Counter evidence 2. Agents with the unchanged states still update after the second experience.

As indicated in section 3.2.3, HP model predicts no change when an experience is repeated in the same state since the combination of equal utilities would be the value of the first utility itself. Hence, that model does not allow for updating in the same state. However, we can see an increase between WTP3 and WTP4 of over 15% after the second trial in the group that consumed high amounts of water in both sessions.

Counter evidence 3. Different initial states lead to the same valuation: $WTP2_L H = WTP2_H H$ and $WTP3_L H = WTP3_H H$.

According to the HP model, agents are biased toward their past experiences. Hence, the only possibility for having the same valuation is to have the same utility from past experience. However, we can see that participants try the juice in two distinctive states by looking at WTP1.

3.5.2 Model by Gagnon-Bartsch and Bushong (2022)

Now, we can ask whether the experimental results can be explained by the GBB model. Remember that according to GBB, expectations are a crucial factor in evaluating new experiences. Agents underestimate the comparison utility and attribute some of the gain-loss utility to the quality of the good. Therefore, having a good prior experience (so a high expectation) might lead to a negative comparison and hence a downward bias in future trials.

As you can see in Table 3.4 in Section 3.4.1, participants in the low water group start with particularly high expectations and decrease their valuations in WTP2 after the trial. Similarly, participants in the high water group have a larger valuation after the second trial (WTP4) compared to the low water group possibly because of their familiarity with the high water consumption. Consequently, the GBB model would be in accordance with the data under the right parameters and hence we cannot falsify it.

3.5.3 Full Naivety

We can think that individuals are totally ignorant about the effect of the state and just take the average of past experiences (i.e. use Bayesian updating without states) to estimate the quality of the good. This case would have almost the same predictions as the HP model since experiences would have a linear effect on valuations. Thus, an experience in a more favorable state must lead to a better valuation. Moreover, repeated experiences in the same state would indicate having averages of equal utilities, so no updating would occur. However, as we have seen in the previous subsection, data speak against these predictions. We can reject the premise of fully naive agents using the same three pieces of counter evidence in section 3.5.1.

3.5.4 Full Sophistication

On the other hand, we can assume that agents are fully aware of the impacts of the states and they can pinpoint the true quality of a good by disentangling the state effects. This assumption would mean that any exogenous variation in environmental factors would not affect the valuations of the participants. On the contrary, results show that initial willingness to pay measures are affected by the treatment variation and original thirst levels are significant determinants of the willingness to pay measures although they are not relevant for future consumption. These results show that the assumption of fully sophisticated agents is not supported by the data. Conversely, people are susceptible to the effects of the consumption states.

3.5.5 Selective Memory and Attention:

The dichotomy between the disappearing impact of treatment manipulation and the robust effect of the original thirst levels can show the importance of salience and attention (see Section 3.4.2). As suggested by Bordalo, Gennaioli, and Shleifer

(2020), a person might retrieve different memories and form different norms depending on the salient cues. Then, any deviation from this norm attracts more attention and disproportionately influences choices.

Because of the high salience of the treatment manipulation by seeing a huge glass of water and trying to drink all of it (versus drinking a small amount of water), participants retrieve relevant memories and were not surprised by the experience with the juice. On the other hand, original thirst levels are not so salient. Hence, thirsty and not-so-thirsty subjects might form the same average norm and satiated subjects might be disappointed with the juice experience whereas thirsty subjects are positively surprised. This mechanism of excess influence of unexpected experiences can perfectly explain the experimental results.

Additionally, at the beginning of the second session, participants are asked to indicate how well they can remember their experience in the first session. In line with the predictions of Bordalo, Gennaioli, and Shleifer (2020), participants who had an extraordinary experience of drinking 500 ml water before the juice report a more vivid memory compared to the 100 ml group[13]. This provides suggestive evidence about the role of memory in learning with repeated experiences.

3.6 Conclusion

In this study, the updating dynamics with state-dependent preferences and the long-run learning pattern are investigated using a lab experiment. In the experiment, participants try an unfamiliar fruit juice twice, one week apart while their thirst levels are manipulated as the treatment variation.

The results of the experiment show that individuals cannot fully understand the effect of the decision environments initially and make biased choices. Participants who drink a high amount of water indicate a much less willingness to pay for a drink to be delivered in the future compared to the low water group. They reflect their current lack of appetite on future consumption. Hence, they are biased toward the current state while deciding about the future states.

Yet, participants are able to understand the effect of the state after multiple experiences and adjust their valuations accordingly. After the second trial, the initial effect of the treatment manipulation disappears and the willingness to pay measures of groups with different past experiences converge. Thus, we can deduce that participants can learn the impact of water consumption after multiple trials.

However, learning only occurs about the salient factors. Apart from drinking water as the treatment variation, the initial thirst levels of participants determine their ultimate satiation levels. While participants can adjust for the effect of water consumption, the initial thirst levels lead to biased evaluations in all WTP mea-

[13] See Appendix for the formal test.

sures, even after the second week. Since initial thirst levels are much less salient than drinking a huge glass of water, participants do not consider that and update their beliefs accordingly. Thus, as suggested by Bordalo, Gennaioli, and Shleifer (2020), attention and salience are key components of the bias and updating mechanism.

Lastly, I compare two main models of learning with state-dependent preferences and test their predictions. The model by Haggag et al. (2019) suggest that better past experiences induce better evaluations after multiple experiences and agents do not update under constant states. However, both of these predictions contradict the experimental data. On the other hand, the model in Gagnon-Bartsch and Bushong (2022) suggests that individuals compare their experiences to the reference based on their past experiences. Thus, a better past experience might generate a worse evaluation after multiple experiences by producing an optimistic expectation. The experimental data confirm this prediction and the results are in line with the model.

These findings help us understand the learning patterns with state-dependent preferences and the mechanism of the attribution bias. As a consequence, we can design better policies to protect individuals from biased decisions. In many countries, economic agents have a right of withdrawal from any transaction within a certain period of time. Consumers, workers, or employers can recover their erroneous actions if they recognize them. However, those policies are based on the (untested) assumption that individuals can form unbiased evaluations about a good or service with a few trials. This study shows that agents can disentangle the effect of the decision environment after multiple trials, but only if they pay attention to a specific factor. Hence, the ideal policies must have two key features: enough time to allow for repeated experiences, and awareness about the relevant consumption states. This way, inefficient transactions can be avoided without creating unnecessary cancellations.

Appendix 3.A Appendix

Table 3.A.1. Randomization check for attrition.

	absent	
LH	-0.0736	(0.0685)
thirst1	0.0273	(0.0270)
WTP1	0.00231	(0.00161)
WTP2	-0.00113	(0.00155)
age	-0.00320	(0.00394)
female	0.0542	(0.0706)
income	-0.00998	(0.0736)
likes juice	-0.0298	(0.0277)
freq. drinks juice	0.0707	(0.0492)
mood1	-0.0353	(0.0421)
hunger1	-0.000225	(0.0206)
health1	-0.0256	(0.0397)
day so far	0.000397	(0.0368)
N	168	

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.A.2. Regression of willingness to drink an extra cup of fruit juice on treatment and thirst measure.

	Would you like to drink one more cup of juice?				
	OLS	OLS	oprobit	probit	logit
LH	0.173	0.177	0.271	-0.223	-0.341
	(0.17)	(0.16)	(0.23)	(0.29)	(0.47)
thirst2		-0.225***	-0.333***	0.412***	0.684***
		(0.06)	(0.09)	(0.13)	(0.22)
N	108	108	108	86	86

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.A.3. OLS regression of WTP and liking measures on the indicated mood at the beginning of the session.

	WTP1	WTP2	Slider1	Slider2	WTP3	WTP4	Slider3
LH	8.186*	0.371	-0.825	-1.700	1.616	-0.813	2.593
	(4.350)	(4.559)	(4.676)	(6.082)	(4.782)	(5.067)	(5.295)
mood at week 1	-1.191	0.589	1.545				
	(2.203)	(2.309)	(2.368)				
mood at week 2				4.246*	4.405**	4.235**	5.077***
				(2.264)	(1.749)	(1.853)	(1.936)
constant	35.53***	26.38**	50.84***	38.48***	0.585	5.199	33.85***
	(12.60)	(13.21)	(13.55)	(12.84)	(9.916)	(10.51)	(10.98)
N	168	168	168	108	129	129	129

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.A.4. OLS regression of the memory question on treatment. The dependent variable is the answer to a 7-point Likert-scale question of "How well can you remember the session from the last week?".

	memory
LH	-0.357*
	(0.184)
constant	6.059***
	(0.134)
N	108

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

References

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2020. “Memory, Attention, and Choice.” *Quarterly Journal of Economics* 135 (3). https://doi.org/10.1093/qje/qjaa007. [72–74]

Busse, Meghan R., Devin G. Pope, Jaren C. Pope, and Jorge Silva-Risso. 2015. “The psychological effect of weather on car purchases.” *Quarterly Journal of Economics* 130 (1). https://doi.org/10.1093/qje/qju033. [58]

Chang, Tom Y., Wei Huang, and Yongxiang Wang. 2018. “Something in the air: Pollution and the demand for health insurance.” *Review of Economic Studies* 85 (3). https://doi.org/10.1093/restud/rdy016. [58]

Fischbacher, Urs. 2007. “Z-Tree: Zurich toolbox for ready-made economic experiments.” *Experimental Economics* 10 (2). https://doi.org/10.1007/s10683-006-9159-4. [65]

Gagnon-Bartsch, Tristan, and Benjamin Bushong. 2022. “Learning with misattribution of reference dependence.” *Journal of Economic Theory* 203. https://doi.org/10.1016/j.jet.2022.105473. [59, 61, 62, 71, 74]

Gawronski, Bertram. 2004. “Theory-based bias correction in dispositional inference: The fundamental attribution error is dead, long live the correspondence bias.” *European Review of Social Psychology* 15 (1). https://doi.org/10.1080/10463280440000026. [60]

Haggag, Kareem, Richard W. Patterson, Nolan G. Pope, and Aaron Feudo. 2021. “Attribution Bias in Major Decisions: Evidence from the United States Military Academy.” *Journal of Public Economics* 200. https://doi.org/10.1016/j.jpubeco.2021.104445. [58]

Haggag, Kareem, Devin G. Pope, Kinsey B. Bryant-Lees, and Maarten W. Bos. 2019. “Attribution Bias in Consumer Choice.” *Review of Economic Studies* 86 (5). https://doi.org/10.1093/restud/rdy054. [58–62, 64–67, 71, 74]

Koszegi, Botond, and Matthew Rabin. 2007. “Reference-Dependent Risk Attitudes.” *American Economic Review* 97 (4): 1047–73. https://doi.org/10.1257/aer.97.4.1047. [62]

Loewenstein, George, Ted O’Donoghue, and Matthew Rabin. 2003. “Projection bias in predicting future utility.” *Quarterly Journal of Economics* 118 (4). https://doi.org/10.1162/003355303322552784. [60, 62]

Simonsohn, Uri. 2010. “Weather To Go To College.” *Economic Journal* 120 (543): 270–80. https://doi.org/10.1111/J.1468-0297.2009.02296.X. [58]

Weiner, Bernard. 2010. “The development of an attribution-based theory of motivation: A history of ideas.” *Educational Psychologist* 45 (1). https://doi.org/10.1080/00461520903433596. [60]