Developing and Validating a Model-Based Method for Measuring People's Subjective Planning Costs

Dissertation
der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Valkyrie Felso
aus Raritan Township, NJ, USA

Tübingen
2024

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	03.12.2024
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter/-in:	Dr. Falk Lieder
2. Berichterstatter/-in:	Prof. Dr. Peter Dayan

Abstract

When planning, people have to trade off between the costs and benefits of additional deliberation. People who tend not to plan may have worse outcomes in various domains which impact well-being such as health, finances, and academics. Several self-report measures exist to quantify people's tendency to plan into the future (e.g., the Propensity to Plan Scale: Lynch *et al.*, 2010, the Consideration of Future Consequences Scale: Strathman *et al.*, 1994, and the Future Orientation Scale: Steinberg *et al.*, 2009). However, these measures can be seen as subjective and might be influenced by social-desirability bias or demand characteristics such as people wanting to answer in alignment with societal values to appear "good".

In this dissertation, I present a computational method for objectively quantifying individual differences in planning via subjective planning costs. Instead of asking a person to self report how much they tend to consider future outcomes, I quantify people's subjective planning costs using a combination of a planning task and a resource-rational computational model (Lieder and Griffiths, 2020). Furthermore, I present a study pairing questionnaire measures with this method. In this study, I investigate whether and how individual differences in subjective planning costs can predict symptoms of psychiatric disorders, as well as if subjective planning costs can predict people's scores on self-report measures for similar constructs.

The foundation of my method for quantifying individual differences in planning is a computational model for measuring subjective planning costs. In a situation where planning is unlikely to pay off, or where planning is particularly hard, it might make sense not to plan at all. Here I investigate how planning which looks suboptimal, may, in fact, be optimal with respect to the subjective planning costs experienced by an individual. In particular, I investigate how costs such as a planning depth cost (i.e., cost for looking into the future) might lead to suboptimal planning using a process-tracing paradigm (the Mouselab-MDP paradigm Callaway *et al.*, 2017). To do so, I extend an existing resource-rational model of planning to include subjective planning costs captured by a cost function with multiple parameters. I show that this model explains human planning better than simpler candidate models and other alternative models. This model provides a mechanistic account for why some people might engage in particular forms of seemingly suboptimal planning.

Furthermore, I introduce the application of Bayesian Inverse Reinforcement Learning to infer these cost weights for individuals. I show, in an experiment where people's planning costs are manipulated, that individual differences in a planning depth cost weight can be reliably recovered for around 70% of people. The planning depth cost weight could be useful as a measure of a person's propensity to plan into the future (e.g., far-sighted versus short-sighted planning).

I present the results from a study investigating whether individual differences in subjective planning costs might be related to symptoms of psychiatric disorders as well as other self-report measures, such as life satisfaction, regrets, and planning behaviors. I find no predictive relationship between inferred cognitive cost weights and self-report measure scores. However, I do find, in exploratory analyses, several smaller correlations between cognitive cost weights and self-report measure scores. This exploratory work expands on the existing literature on the structure of planning differences across different mental disorders.

Finally, I outline and discuss the possible limitations of the method, and future studies needed before these methods could be applied to a larger-scale, more real-world setting.

Kurzfassung

Bei der Planung müssen Menschen den Nutzen zusätzlicher Erwägungen gegen ihre Kosten abwägen. Menschen, die dazu neigen, nicht zu planen, erzielen in der Regel schlechtere Ergebnisse in verschiedenen Bereichen des Lebens, die sich auf ihr Wohlbefinden auswirken, einschließlich Gesundheit, Finanzen und Bildung. Es gibt mehrere Fragebögen, um zu messen wie sehr Menschen dazu neigen die Zukunft zu planen (z. B., die Fragebögen zur Erfassung der Planungsneigung: Lynch *et al.*, 2010, der Berücksichtigung zukünftiger Konsequenzen: Strathman *et al.*, 1994, und der Zukunftsorientierung: Steinberg *et al.*, 2009). Diese Maße können jedoch als subjektiv angesehen werden und könnten durch soziale Erwünschtheit oder Reaktivitätseffekten verfälscht werden, wie z. B. dem Wunsch der Menschen, im Einklang mit gesellschaftlichen Werten zu antworten, um "gut" zu erscheinen.

In dieser Dissertation entwickelte ich ein objektives, verhaltensbasiertes, mathematisches Verfahren zur Messung individueller Unterschiede darin, wie Menschen planen. Anstatt eine Person zu bitten, selbst zu berichten, wie sehr sie dazu neigt, zukünftige Konsequenzen in Betracht zu ziehen, schätzt meine Methode die subjektiven Planungskosten einer Person mithilfe einer Kombination aus einer Planungsaufgabe und einem rationalen kognitiven Prozessmodelll (Lieder and Griffiths, 2020). Außerdem stelle ich eine Studie vor, die Fragebogenmessungen mit dieser Methode kombiniert. In dieser Studie untersuche ich, ob und wie individuelle Unterschiede in den subjektiven Planungskosten die Symptome psychiatrischer Störungen vorhersagen können, und wie sehr mein verhaltensbasiertes Maß der subjektive Planungskosten mit dem Erleben der Person korreliert ist.

Die Grundlage meiner Methode zur Messung individueller Unterschiede in der Planung ist ein kognitives Prozessmodell das die Planungsstrategien von Personen mit unterschiedlichen Planungskosten vorhersagt . In einer Situation, in der es unwahrscheinlich ist, dass sich lohnen würde zu planen, oder in der planen besonders schwierig ist, könnte es sinnvoll sein, überhaupt nicht zu planen. Hier untersuche ich, wie Planungsstrategien, die suboptimal wirken, in Wirklichkeit optimal sein können, wenn man die subjektiven Planungskosten berücksichtigt, die ein Individuum empfindet. Insbesondere untersuche ich, wie erhöhte Kosten für vorausschauendes Planen zu suboptimaler Entscheidungen führen können, indem ich ein Prozessverfolgungsparadigma verwende (das Mouselab-

MDP-Paradigma Callaway *et al.*, 2017). Zu diesem Zweck erweitere ich ein bestehendes ressourcenrationales Planungsmodell um subjektive Planungskosten, die durch eine Kostenfunktion mit mehreren Parametern erfasst werden. Ich zeige, dass dieses Modell menschliches Planen besser erklärt als einfachere Modelle und alternative Erklärungen. Dieses Modell liefert eine mechanistische Erklärung dafür, warum manche Menschen sich auf bestimmte Formen scheinbar suboptimaler Planung verlassen.

Darüber hinaus wende ich Bayesian Inverse Reinforcement Learning an, um die Parameter dieser Planungskostenfunktion zu schätzen. In einem Experiment, in dem die Planungskosten von Personen manipuliert werden, zeige ich, dass individuelle Unterschiede in den Kosten des vorausschauenden Planens für etwa 70% der Versuchspersonen zuverlässig erfasst werden können. Dieser Parameter könnte ein nützliches Maß dafür sein, wie sehr eine Person dazu neigt in die Zukunft zu planen (z. B. langfristige versus kurzfristige Planung).

Ich präsentiere die Ergebnisse einer Studie, die untersucht, ob individuelle Unterschiede in den subjektiven Planungskosten mit den Symptomen psychiatrischer Störungen sowie mit anderen Fragebögen wie Lebenszufriedenheit, Bedauern und Planungsverhalten zusammenhängen könnten. Ich finde keine prädiktive Beziehung zwischen den geschätzten kognitiven Kosten und den Selbstauskünften der Versuchspersonen. In explorativen Analysen finde ich jedoch mehrere kleinere Korrelationen zwischen den geschätzten Planungskosten und einzelnen Fragebögen. Diese explorative Arbeit erweitert die bestehende Literatur über die Struktur individueller Unterschiede im Planen zwischen verschiedenen psychischen Störungen.

Abschließend skizziere und diskutiere ich die möglichen Grenzen der Methode und künftige Studien, die erforderlich sind, bevor diese Methode in einem größeren, realistischeren Rahmen angewendet werden kann.

Acknowledgments

First, I would like to thank Falk Lieder for his mentorship and his tireless support in helping me wrap this up. It has been a privilege to get to work with you. I would also like to thank my committee members, Peter Dayan and Martin Butz, for their feedback and guidance over the last few years. In particular, thanks to Peter Dayan for helping push me to finish my first draft of the dissertation in early 2023. Thank you also to Felix Wichmann for stepping in for the defense and being not just a member of my examination committee but an excellent chair.

Thank you to all current and past REG and RAL members and collaborators; I have been lucky to get to know and work with you! In particular, I would like to thank Jugoslav Stojcheski and Jiatong Liu for the opportunity for close cooperation on projects which are not in this work, but which taught me a lot and informed the work in this dissertation. To my fellow doctoral students, Victoria Amo Olea, Lovis Heindrich, and Ruiqi He: Thank you for the camaraderie! Thank you also to everyone who mentored me in my scientific journey up until this point, particularly Eran Eldar, Gecia Bravo-Hermsdorff, Lee Gunderson, Lee Harten, Zhenghan Qi, Yael Niv, and Yossi Yovel. Victoria Amo Olea, Lisa Eckerstorfer, and Maria Wirzberger: You have all really touched my life on multiple fronts, no matter how little time we actually shared an office. Fred Callaway, Sayan Gul, and Yash Raj Jain: Thank you for paving the way for this work through existing code. I am eternally grateful to Fred Callaway for the input given over the years, particularly for the first draft of my preprint.

I am grateful to the RAL & REG lab managers: Zahra Tahmasebi, Nisha Tyagi, Sierra Kaiser, and Bastian Nowozimski. Without you, administrative hurdles would've been insurmountable! Thank you also to the MPI-IS Administration, particularly Kornelia Sigloch, for all the help in the last years with contract changes and mobile working contracts. Thank you to Heidi Schuster and Ulrich Braun at the Max Planck Society for their input and efforts in helping me improve my ethics applications. A huge thank you to the MPI-IS IT team, particularly Telintor Ntounis and Sebastian Stark, for their help in setting up the infrastructure to run the final experiment. Thanks to the MPI-IS Scientific Computing ZWE for their patient assistance. Thank you to Ruiqi He and Diego Baptista Theuerkauf for graciously donating me cluster money to finish the final simulations.

Without you, my cluster jobs would still be slowly running. Thank you to Jean-Claude Passy, my current supervisor, for his support and flexibility.

I would also like to acknowledge the support I've received from the International Max Planck Research School for Intelligent Systems. Leila Masri's support and advocacy over the last few years have been crucial in helping me finish. Thank you! Thanks to Sara Sorce and the volunteer scholars for their work in organizing S4 Workshops. Without the dissertation writing workshop, this dissertation would be much worse organized (and may not even exist)!

I would also like to thank everyone who has formed or participated in a coworking session or writing group with me over the last few years. Thank you to Sabine and Markus Wuermeling for providing a place to work in peace in January 2023. I am also incredibly grateful to Katy Peplin for lending an ear in some of the most difficult times in the process.

Thank you to all the participants from my studies. I feel privileged to be able to work with and analyze your data.

A part of this dissertation was previously submitted to Behavior Research Methods. I want to thank the anonymous reviewers for their extensive feedback.

I would like to thank my family for their support and patience over the last few years. Finally, I cannot begin to express my gratitude to Johannes. Without your support, finishing this would have been impossible.

Declaration of Collaborative Work

My advisor, Dr. Falk Lieder, supervised this research, secured funding, and contributed significantly to the conceptualization of the research. Some work in the dissertation, specifically Chapters 2, 3 & 4 stem from a preprint (Felso and Lieder, 2023) that I wrote with Dr. Lieder. Dr. Lieder contributed to the original draft of the background and discussion sections in the manuscript as well as to the review and editing of the entire manuscript.

I developed the research goals and aims with Dr. Lieder's supervision. I collected the datasets reported in the dissertation (except for two publicly available datasets, clearly marked in Chapter 3), curated the data, ran all statistical analyses and prepared all visualizations. Finally, I wrote the original draft of the preprint, with Dr. Lieder's help as stated above.

Contents

1 Introduction **1**

1.1 Research objectives . . . 2
1.2 Outline . . . 3

2 Foundations **5**

2.1 Measuring planning propensity . . . 5
2.1.1 Survey-based measures . . . 5
2.1.2 Task-based measures . . . 6
2.1.3 Tower of Hanoi and Tower of London . . . 6
2.1.4 Plan-a-Day Task . . . 6
2.1.5 Two-step Task . . . 7
2.1.6 Mouselab-MDP . . . 8
2.2 Sequential decision-making . . . 10
2.2.1 Markov Decision Processes . . . 10
2.2.2 Reinforcement Learning . . . 10
2.2.3 Inverse Reinforcement Learning . . . 11
2.2.4 Meta-level Markov Decision Processes . . . 11
2.3 Resource Rationality . . . 12

3 Modeling the cost of planning **13**

3.1 Model development . . . 15
3.1.1 Meta-level MDP Model of the Mouselab-MDP Task . . . 15
3.1.2 An extended model of the cost of planning. . . . 16
3.1.3 Meta-Level Action Selection . . . 18
3.1.4 Considered Models . . . 19
3.1.5 Model Fitting Details . . . 19
3.2 Experiment 1: Which model describes the cost of human planning best? . . . 20
3.2.1 Introduction . . . 20
3.2.2 Methods . . . 20
3.2.3 Results . . . 21
3.2.4 Conclusion . . . 26

3.3 Does the environment structure matter for model fit? 26
3.3.1 Methods 27
3.3.2 Results 28
3.3.3 Conclusion 31
3.4 Discussion 32

4 Measuring individual differences in planning cost weights 35
4.1 Method 36
4.1.1 Applying Bayesian inverse reinforcement learning to the Mouselab-MDP paradigm 36
4.1.2 Prior 36
4.1.3 Inference 37
4.1.4 Quantifying the posterior distributions of individual variables . . 37
4.2 The fixed effects method yields similar results to random effects methods. 38
4.2.1 Possible method outputs 38
4.2.2 Simulation data 38
4.2.3 Results 38
4.3 The fixed effects method yields sufficient parameter recovery. 40
4.3.1 Methods 40
4.3.2 Results 41
4.4 Can we differentiate between simulated data generated from different models? 43
4.4.1 Methods 43
4.4.2 Results 44
4.4.3 Conclusion 44
4.5 Experiment 2: Validation Experiment 44
4.5.1 Methods 47
4.5.2 Results 48
4.5.3 Conclusion 53
4.6 Discussion 54

5 Relation between inferred planning cost weights and self-report measures 57
5.1 Methods 59
5.1.1 Participants 59
5.1.2 Materials 59
5.1.3 Procedures 65

5.2 Results 66
5.2.1 Participant data was best fit by the cost model with all five cost weights. 66
5.2.2 The inferred parameter values were similar to those in Experiment 1. 66
5.2.3 Inferred model parameters could not explain the variance in the self-report measures. 68
5.2.4 Simpler behavioral measures can explain the variance in the Cognitive Reflection Test. 69
5.2.5 Inferred strategy could not explain variance in the self-report measures. 70
5.2.6 Planning cost weights still correlate with questionnaire measures. 71
5.2.7 Forward search is associated with increased age. 74
5.3 Discussion 75

6 Conclusions and Future Directions **79**
6.1 Summary of contributions 79
6.2 Limitations & future directions 80
6.2.1 Inverse problems 80
6.2.2 Investigation of larger scale environments 81
6.2.3 Continuous optimization 81
6.2.4 Are cost weights trait-like? 82
6.2.5 Do cost weights vary by context? 82
6.2.6 Executing plans 82
6.2.7 Individualized planning interventions 83
6.3 Conclusions 85

A Appendix **87**
A.1 Chapter 3 87
A.1.1 Participant strategies for Experiment 1 87
A.1.2 Simulated data from the best model matches human behavior just as well as data generated from alternative models 87
A.1.3 Does the addition of a discount rate and a concave utility function explain the data better than the depth parameter? 88
A.2 Tables for Chapter 3 90
A.3 Figures for Chapter 3 90
A.4 Chapter 4 95
A.4.1 Both blocks in Experiment 2 are fit just as well as the test block in Experiment 1 99

A.4.2 The order of the baseline block has little effect on behavior or inferred parameter values 99
A.5 Tables for Chapter 4 101
A.6 Chapter 5 104
A.6.1 Hypotheses 104

Symbols **117**

Abbreviations **119**

Bibliography **127**

Chapter 1

Introduction

Planning can be daunting. For example, when planning out a research project, there are many unknown factors to consider: what are the milestones, when to run each experiment, what to do if things do not work out as planned, etc. Planning is also ubiquitous, not just occurring in the context of work but also in the contexts of finances, family, recreation and more, as well as on group or institutional levels. Failing to plan, or making insufficient plans can be costly, with those around the (non-)planner feeling the consequences.

Several questionnaire measures exist which could be used to measure planning or future thinking: the Propensity to Plan Scale (Lynch *et al.*, 2010), the Consideration of Future Consequences Scale (Strathman *et al.*, 1994), and the Future Orientation Scale (Steinberg *et al.*, 2009). However, self-reports of planning may not always be accurate: a participant may not be comfortable answering truthfully, either because they want to appear better to the experimenter ("social-desirability bias ") or because their own self-perception is distorted. Such inaccuracies could be further exacerbated by stressful situations (like job interviews), mental disease and/or personality traits (e.g., "dark" personality traits). For example, someone with a high positive self-image due to hypomania or narcissism may report that they are better at planning than they really are. At the same time, someone with higher social anxiety may under- or over-report due to the social setting of the experiment. Additionally, reporting may be hard for those who do not have good insight into their own planning due to lack of introspection. Therefore, having a diagnostic task-based measure such as a short game, could be invaluable for the study of individual differences in human planning.

Such a task-based measure may help not just in understanding individual differences but also in improving people's planning. There are many tools and resources which offer to help people when making personal plans (project planning frameworks, coaching, forecasting tools, etc.). However, it is not always clear which tool might help a specific person or group. Being able to extract exactly how a person's planning is suboptimal could be helpful in pairing people with planning tools. Understanding such individual differences could also be useful for the development of planning interventions or decision-support systems (such as the work done in persuasive technologies: Alslaity *et al.*, 2023, educational technology: Klašnja-Milićević *et al.*, 2011 and cognitive training for older adults: Peretz *et al.*, 2011.)

One central assumption of my work is that people differ in planning ability because they find different aspects of planning more or less costly. For example, someone with a higher cost for prospection (thinking into the future) may only consider near-future consequences, while someone with a higher general cost for planning may not plan at all. While both people may exhibit impaired planning, their planning will be impaired in different ways.

In this dissertation, I present work developing and validating an objective measure of individual's planning costs. I develop a reinforcement learning model of individual differences in human planning. I then combine the model with a planning task and statistical inference to yield a method for uncovering such planning costs for individuals. Finally, I investigate whether differences in planning costs as extracted by the method correlate with personal characteristics, especially psychiatric symptoms, and real-world planning.

1.1 Research objectives

The central aim of this dissertation is to develop an objective method for measuring individual variation in planning.

The first objective of this research is to develop and validate a reinforcement learning model of human planning. This model incorporates subjective cognitive planning costs. This model will serve as foundation for all further work in this dissertation, and will be a cornerstone in the eventual method for inferring individual variation in planning.

The second objective is to develop and validate a method for extracting individual planning costs using this model. This method will build on the model of human planning that incorporates subjective cognitive planning costs.

Finally, the third objective is to assess the relationship between the subjective planning costs outputted by the method and self-reported personality and psychiatric symptomatology. Finding a predictive relationship (or the absence thereof) between the subjective planning costs and the planning and future thinking measures would highlight the value of using this method to study real-world planning ability. At the same time, any correlation between psychiatry symptomatology and planning costs may aid in understanding the relation between mental health and individual planning differences.

1.2 Outline

This remainder of this dissertation comprises a general foundations chapter, three main chapters presenting the development and validation of the method to extract individual subjective planning costs, and a final chapter outlining conclusions and future directions.

In Chapter 2, I first introduce several questionnaires and tasks used for studying human planning, motivating the choice of the task that will be used throughout the dissertation. Next, I introduce the concepts from reinforcement learning that will be used throughout the dissertation, particularly in model and method development. Finally, I describe the concept of resource rationality, which will be a cornerstone of the model presented in Chapter 3.

In Chapter 3, I present a new model of planning, which supposes that people act rationally according to individual cognitive costs. My model differs from the model presented in Callaway *et al.* (2022), by including several additional sources of planning cost. A person with high cognitive costs to plan may not plan at all, as a rational adaption to this cost. I present results for a study, showing that the model with all defined cognitive costs explains participant data the best, taking into account both model simplicity and accuracy. I then show, using data published by other researchers, that even in environments with different reward variability settings, this model explains participant data best.

In Chapter 4, I present and validate a method to extract individual cognitive costs using the model introduced in Chapter 3 and Bayesian Inverse Reinforcement Learning (Ramachandran and Amir, 2007). Using simulated datasets, I demonstrate that a method using this model has sufficient parameter recovery. I show that the model's uncertainty in parameter estimates can be reliably quantified with highest posterior density intervals (similar to credible intervals). I also present a validation experiment where cognitive costs were assigned to participants, and show that my method can reasonably infer imposed planning depth costs.

In Chapter 5, I present a large-scale online study, aimed at investigating the relation between cognitive costs and personal characteristics and real-world planning behavior. I present results of preregistered statistical analyses relating to the relation between psychiatric symptoms and subjective planning costs. In addition, I explore whether simpler behavioral measures or a separate measure of planning strategy in the task relate to personal characteristics and real-world planning behavior. I also explore whether any correlations between self-report measures and planning cost weights exist.

Finally, in Chapter 6, I discuss limitations and future studies needed before my findings could be put into practice. I conclude the dissertation by outlining how the work could inform the development of personalized planning interventions.

Chapter 2

Foundations

This chapter provides the necessary, overarching background for later chapters in the dissertation. First, I highlight some current methods for measuring human planning and how they vary from one another. Next, I introduce the topic of sequential decision-making and provide an overview of ways such problems can be modeled using reinforcement learning. This background is the foundation needed to understand the model and method development in Chapter 3 and Chapter 4. Finally, I provide an overview of resource rational analysis – a type of analysis which takes into account the limits in human cognition. The model presented in Chapter 3 will be based on resource-rational assumptions.

2.1 Measuring planning propensity

Planning is a complicated process, even on the individual level. Here I briefly introduce survey-based and task-based measures which measure how much people plan.

2.1.1 Survey-based measures

Several self-report measures exist to measure how people plan and consider future consequences: the Future Orientation Scale (Steinberg *et al.*, 2009), the Consideration of Future Consequences Scale (Strathman *et al.*, 1994) and the Propensity to Plan Scale (Lynch *et al.*, 2010). Please see Chapter 5 for more information on all three scales.

Scores in these self-report measures correlate with health, financial wellbeing (Joireman *et al.*, 2005; Lynch *et al.*, 2010), and educational outcomes (Joireman, 1999; Peters *et al.*, 2005). For example, in the health domain, higher future orientation has been shown to mitigate onset of depressive symptoms in youth after emotional victimization (Hamilton *et al.*, 2015). Van Beek *et al.* (2013) were also able to predict people's food and

Some of the content of this chapter is from the background section in my preprint available here: https://psyarxiv.com/xmf3y/.

exercising behaviors using specifically-adapted versions of the Consideration of Future Consequences Scale.

Although these self-report measures are generally connected to real-world planning and well-being, there may be some cases where participants may under- or over-report planning. As discussed in the previous chapter, this could arise from factors such as environment (e.g., high stress settings), mental disorders or even personality types. This is my motivation for developing a task-based measure of individual differences in planning.

2.1.2 Task-based measures

Task-based measures of planning vary in their similarity to real-world planning (from abstract games to planning in a simulated setting.) Planning in a simulated setting has an advantage as it is closer to planning in the real world, with all its complexities. On the other hand, abstract games allow researchers to measure planning by itself, removed from outside factors such as domain-specific knowledge or preferences. Here I will introduce four particular tasks which vary in complexity.

2.1.3 Tower of Hanoi and Tower of London

The Tower of Hanoi and Tower of London tasks are particularly influential for measuring human planning and problem-solving (Simon, 1975; Shallice *et al.*, 1982). In both tasks, a participant finds themselves faced with three rods and discs or balls which can be placed on the rods. The participant is tasked with moving the discs or balls along the rods into a pre-specified arrangement. In the Tower of Hanoi task, the participant must move the discs so that they are sorted by size, with wider discs on the bottom. In the Tower of London task, participants are given a target arrangement of balls as well as a target number of moves.

These tasks, while abstract, have been the standard way of measuring planning ability in the lab (Unterrainer *et al.*, 2004). Performance does seem to be related to real-world planning, for example, chess players have been shown to perform better in the Tower of London task (Unterrainer *et al.*, 2006). Unfortunately the task is quite abstract and does not appear to be ecologically valid (Kafer and Hunter, 1997; Campbell *et al.*, 2009).

2.1.4 Plan-a-Day Task

On the other hand, tasks such as the Plan-a-Day task mirror real planning to a larger degree. In the computerized Plan-a-Day task (Funke and Krüger, 1993), participants are given a map, and a set of assignments they must complete in a day. For example, they may have to go to the post office at noon and to an important meeting at 2 PM, with a 15-minute walking distance between the two buildings. A participant must select their schedule to fulfill as many daily assignments as possible.

Interestingly, performance in the Plan-a-Day task seems to remain the same as people age, whereas performance in the Tower of London task goes down (Phillips *et al.*, 2006). This may have to do with the fact that older adults can compensate for decline in cognitive abilities with their ability to focus on task-relevant features, possibly due to domain knowledge (Kliegel *et al.*, 2007). Another successful application of the Plan-a-Day task is in measuring planning ability of patients diagnosed with schizophrenia. Holt *et al.* (2011) found that the Plan-a-Day task appears to have high ecological validity for measuring executive functioning, making it a good candidate for assessment: not too abstract and not as involved as real-world observation.

2.1.5 Two-step Task

The two-step task is a task developed by Daw *et al.* (2011) and consists of interdependent binary choices. In the task, participants must choose between two fractals in two steps (or stages) before receiving a reward or loss. Crucially, the fractal chosen in the first step leads to one set of fractals in the second step with high probability ("common" transition), and another set of fractals in the second step with low probability ("rare transition"). Participants who exploit the underlying structure of the environment are said to plan in a "model-based" way, whereas participants who seem to ignore the structure are said to plan in a "model-free" way.

For example, say you decide on two cafeterias for lunch, A and B. There are also two chefs in this world, A and B. The chefs are usually at their respective cafeteria. However, sometimes the chefs switch between the cafeterias, but you do not know their exact schedules. If you have a bad meal at Cafeteria A when Chef B is visiting, as a model-based learner you should update your valuation of Chef B's cooking. However, if you were a model-free learner, you would now update your valuation of Cafeteria A based on this bad experience.

Most relevant to this dissertation, this task has been used to identify differences in model-based versus model-free planning in mental disorders (Gillan *et al.*, 2016; Seow *et al.*, 2021). Specifically, people higher in OCD traits appear to behave in a more model-free way. However, due to the two-step nature of the task, the task lacks the complexity often seen in real-world decision-making.

2.1.6 Mouselab-MDP

The Mouselab-MDP paradigm is a process-tracing paradigm for planning, inspired by the Mouselab paradigm (Johnson *et al.*, 1989). Process-tracing paradigms allow researchers to observe people's decision-making process through information-gathering behavior alone (Schulte-Mecklenbeck *et al.*, 2011). A well-developed process-tracing task is cheaper and easier to use than alternative paradigms such as eye-tracking, click-tracing or brain image approaches (such as fMRI with categories easily distinguishable in the brain.)

In the Mouselab paradigm, people are faced with a choice between several gambles. In order to make the decision, they are given a grid of potential outcomes and gambles, with the outcomes hidden behind a box. Participants are told that they can click on the boxes in order to temporarily reveal the potential outcomes. This clicking on the box allows the researcher to "watch" a participant's decision-making process. This paradigm allows researchers to uncover how participants' planning heuristics adapt to changes in the environment (Payne *et al.*, 1988). However, the paradigm is only geared for one-step decision-making.

The Mouselab-MDP paradigm (Callaway *et al.*, 2017) is an extension of the Mouselab paradigm which allows researchers to observe a participant's planning process in a sequential decision-making task. Much in the same way that one could piece together how someone planned a trip by looking at their web browsing history (which destination, airline, etc. did the user consider?), the researcher can infer what is going through a person's head by looking at what information they gather.

In the task, participants are shown a map comprising multiple locations, called *nodes*, and paths that connect them (see Figure 2.1). Each node contains a certain monetary gain or loss, which participants experience if they visit that node. In the displayed version of the task, participants are told that they have to lead a money-loving spider through a web of cash. The spider has to traverse the web in the direction of the arrows until it reaches one of the outermost nodes (backwards movements are not allowed). The participant's task is to gather as much money as possible. Participants must first click on the starting node before the other nodes appear.

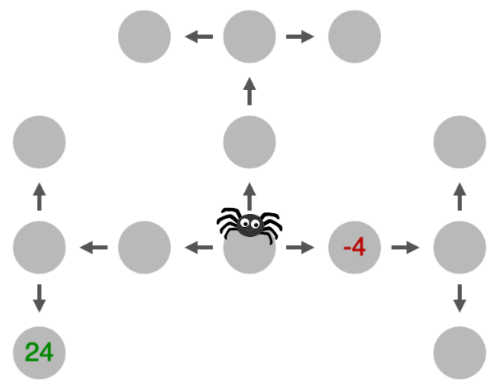


Figure 2.1: Example Mouselab-MDP trial.

A key element of this process-tracing paradigm is that the values of the nodes are hidden at the start of a trial. Participants are told they can click on nodes to reveal a node's value for a cost of \$1. Participants can then use this information to decide whether to later traverse the node and collect its reward/loss or not. After participants feel they have gathered enough information, they can move the spider out of the web with the arrow keys. Once participants have started moving the spider through the web, they are not allowed to inspect further nodes.

Of particular note is a configuration of the spider web that has previously been used for research on the "present bias" (Lieder *et al.*, 2019b). The values of nodes further out into the future (i.e., the nodes at the edge of the spider web) are drawn from a higher variance distribution than those in the middle, followed by those at the beginning of the web. The possible values at the first (inner) level are $\{-4,-2,2,4\}$; for the middle level, they are $\{-8,-4,4,8\}$; and for the final (outer) level, they are $\{-48,-24,24,48\}$. Given knowledge of this structure, the optimal planning strategy would click on the outer nodes first because they carry the most information about a path's total value.

Jain *et al.* (2022) developed a computational method for inferring which planning strategies participants use from the information gathering actions they perform in the Mouselab-MDP paradigm. They observe two types of suboptimal planning: a bias towards inspecting nodes close to the start even when they are less informative than the nodes further from the start, and the tendency to not plan (click) at all. However, it remains unclear why some people engage in these suboptimal forms of planning but not others. This method lies at the descriptive stage of computational modeling, and does not provide a mechanistic account for suboptimal planning.

In this dissertation, I will study planning primarily with the Mouselab-MDP task. The

Mouselab-MDP task lies in a midpoint between abstract and real-world, and the process-tracing aspect allows us to easily model people's planning processes.

2.2 Sequential decision-making

We make many decisions every day, spanning different domains and levels of consequence (trivial to life-altering): what to eat, whether it's safe to cross the street, how to plan the day. Many decisions that are encountered in everyday life are *sequential* in nature – these decisions are not one-off. In such situations, future decisions can be improved by incorporating knowledge gained from previous decisions. For example, when deciding what to eat in an unfamiliar setting (say, visiting a locale with a cuisine you are unfamiliar with), one might initially randomly order dishes. As time goes on, the decision on what to eat can now be informed by the result of the previous meals (what food tastes good or bad, discovered allergies, etc.)

2.2.1 Markov Decision Processes

Sequential decision-making problems are often modeled as **Markov decision processes** (Sutton and Barto, 2018). A Markov decision process is defined by a set of states $\mathcal{S}$, a set of actions $\mathcal{A}$, a transition function $\mathcal{T}(s,a,s') = \Pr(s' \mid s,a)$ and a reward function $\mathcal{R}(s,a) = r$.

This mathematical framework can be used when the Markov property is satisfied (or at the very least, assumed). This means all future decisions should be independent of past decisions. In practice, this means the present state should encode all information about the past that is needed for future decisions. In the example of choosing a meal, even though the success of a decision may depend on the last chosen meal (for example, you may need a break from your favorite meal every once in a while), the history of chosen meals can be encoded into the present state (through memory).

2.2.2 Reinforcement Learning

Reinforcement learning focuses on developing algorithms that allow agents behaving in a particular environment to optimize their rewards (or minimize their costs). This sub-branch of machine learning borrows ideas from the study of animal behavior such as operant conditioning, where an animal learns to perform a certain behavior through the administration of rewards or punishments.

In the typical reinforcement learning framework, one is given an agent's state space $\mathcal{S}$, an action space $\mathcal{A}$, a transition function $\mathcal{T} = Pr(s' \mid s,a)$, and a reward function $\mathcal{R} = Pr(r \mid s,a)$ (Sutton and Barto, 2018). The goal is to derive a policy $\pi : \mathcal{S} \times \mathcal{A} \to [0,1]$. For smaller problems, exact solutions via dynamic programming exist. For large problems, solutions can only be found via approximation methods such as function-approximation.

2.2.3 Inverse Reinforcement Learning

The inverse reinforcement learning paradigm puts this on its head: we observe an agent's behavior, a trajectory τ of state-action tuples (s_t, a_t) for time steps $t = 1, 2, 3, \ldots, T$. Knowing the agent's state space $\mathcal{S}$, action space $\mathcal{A}$, and the transition dynamics of the environment, $\mathcal{T}(s, a, s') = Pr(s' \mid s, a)$ the most common goal is to uncover the reward function $\mathcal{R}(s, a) = r$. Inverse reinforcement learning can also be used for apprenticeship or imitation learning (Kubala *et al.*, 2019; Arora and Doshi, 2021).

In Bayesian inverse reinforcement learning (Ramachandran and Amir, 2007; Rothkopf and Dimitrakakis, 2011), a posterior distribution over potential reward functions is inferred:

$$P(\mathcal{R} \mid \tau) \propto P(\tau \mid \mathcal{R}) P(\mathcal{R})$$

The likelihood $P(\tau \mid \mathcal{R})$ is generally assumed to arise from a softmax action-selection policy $P(a \mid s)$. Kubala *et al.* (2019) have extended this algorithm to learning agents through the addition of a behavior rule that outputs the policy given the history of past outcomes for actions in states, a discount factor, and reward parameter values.

2.2.4 Meta-level Markov Decision Processes

A meta-level MDP (Hay *et al.*, 2014) focuses on the meta-level decision process. Rather than focus on the object-level decision, "which action to take", a meta-level decision focuses on the level above, "which information to gather, in order to make the object-level decision". For example, when planning transportation to a conference, an object-level decision might be "what transportation do I take?". A meta-level decision might be "how long do I work on researching transportation options?".

In the Mouselab-MDP task, on the object level, a participant is choosing a path to move the spider. On the meta-level, the participant is choosing what information to gather to aid in this decision and when to stop gathering information. Please see Chapter 3 for more specifics on how planning behavior in the Mouselab-MDP task is modelled as an MDP.

2.3 Resource Rationality

Planning has traditionally been modeled as searching for a good path through a tree of state-action sequences (Newell and Simon, 1972; Huys *et al.*, 2015; Van Opheusden *et al.*, 2017; Callaway *et al.*, 2022). A crucial challenge for these models is to explain which state-action sequences people consider and which they ignore (Huys *et al.*, 2015). The principle of resource-rationality suggests that this decision should follow a rational cost-benefit analysis (Lieder and Griffiths, 2020; Callaway *et al.*, 2022). That is, people should consider a potential course of action if and only if the improvement in decision quality it can be expected to achieve is larger than the cost of the time and cognitive resources that this additional planning would require.

Resource-rational analysis is an extension of Anderson's rational analysis (Anderson, 2013), in which cognitive constraints are incorporated into rational models of cognition. It thereby constrains the space of potential cognitive models by requiring the model to accomplish the task with limited computational resources (Lieder and Griffiths, 2020). Seemingly suboptimal or irrational behavior, such as heuristics or biases, can be seen as a rational adaptation to people's limited cognitive resources or to environmental constraints (Todd and Gigerenzer, 2012). Resource-rational analysis has already been applied to study planning in the Mouselab-MDP paradigm (cf. Callaway *et al.*, 2018, 2022). However, so far, there are no resource-rational models of individual differences in planning.

Chapter 3

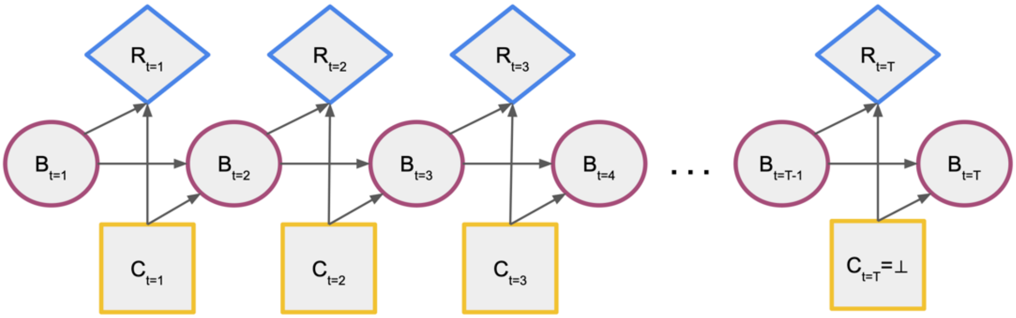
Modeling the cost of planning

Imagine planning how to travel to a nearby conference. Should you drive, take a train, or even a plane? There are many factors that go into this decision: which mode of transport will be most comfortable, quicker, or allow you to do more work at the time, budgets or other financial constraints, and even further-out environmental considerations. Perhaps you may find yourself too busy to plan, or find planning too difficult cognitively, leading to suboptimal plans, such as waiting until the week before the conference to book your travel. Some of these factors that influence planning may be individual, or differ between people: some people may have a passion for planning travel whereas it instills dread in other people.

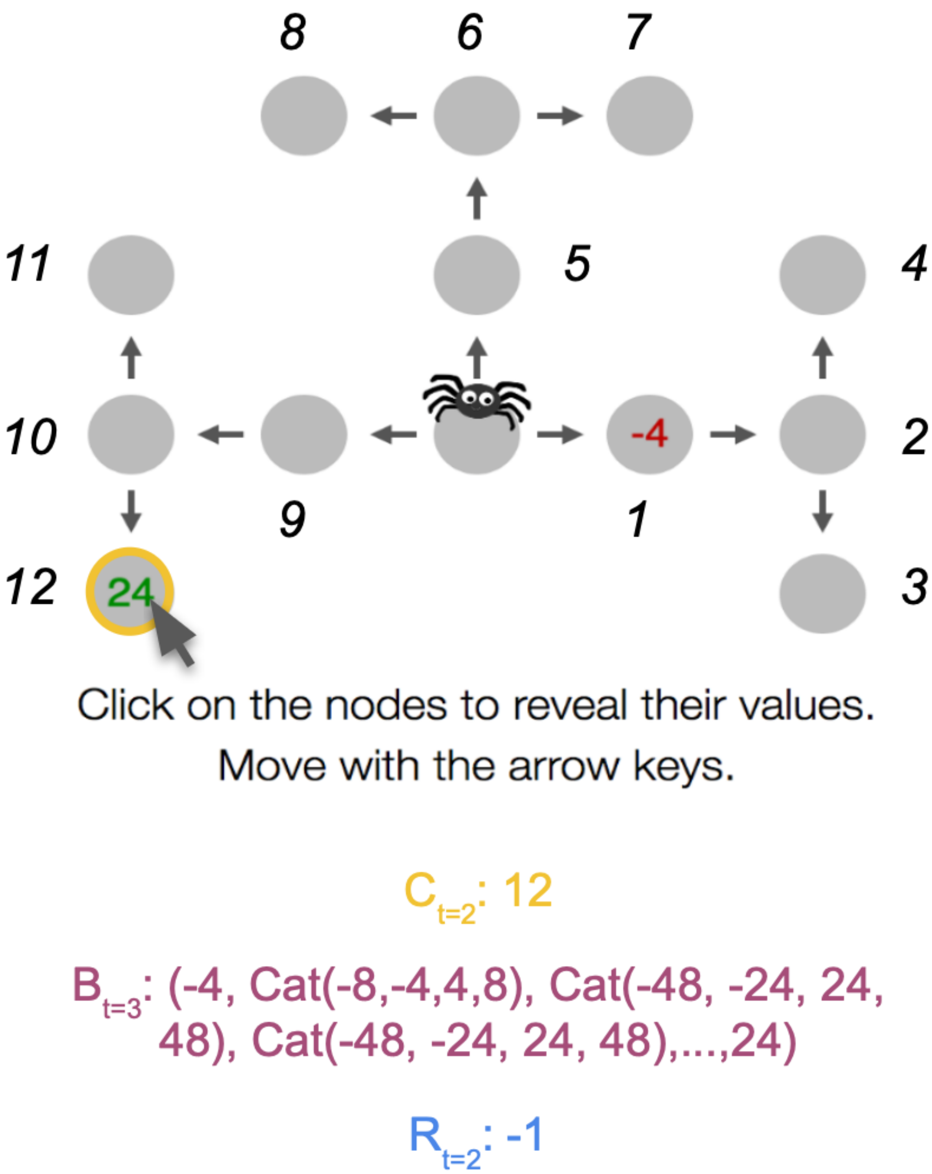
In the Mouselab-MDP task, we often see participants plan in many seemingly suboptimal ways. A participant may only click on nodes which do not provide much information about the path value, or they may click on no nodes at all, foregoing planning (Lieder *et al.*, 2019a; Callaway *et al.*, 2022; Jain *et al.*, 2022). Using the lens of rational analysis (Anderson, 1989), such behavior can be modelled as arising from differences in task goal or in yet-to-be modelled environmental constraints.

In this chapter, I will present a general model of planning in the Mouselab-MDP task which augments the current state-of-the-art planning model (Callaway *et al.*, 2022) by including a newly parametrized planning cost model. This work may help to explain why we see people make plans which seem suboptimal, both in the Mouselab-MDP task and in real-world planning situations. In Chapter 4, I will go on to validate a method using this model which infers individual differences in the cost of planning.

Much of the content of this chapter is from the section "Modeling the cost of planning" in my preprint available here: https://psyarxiv.com/xmf3y/.



(a)



(b)

Figure 3.1: (a) A diagram of the meta-level MDP. (b) In this example time step, the participant just clicked on the node highlighted in yellow ($C_{t=2}$) with a cost of -1 ($R_{t=2}$) and ends up in a new belief state ($B_{t=3}$). The numbers outside the nodes denote each node's index in the belief state.

3.1 Model development

To make it possible to infer individual differences in planning costs from behavior, I developed a generative model that predicts planning behavior in the Mouselab-MDP paradigm based on individual planning costs. I approached modelling people's planning in the Mouselab-MDP task from a resource-rational point of view, assuming that people plan near-optimally given their limited cognitive resources. The model thus assumes that people act optimally with respect to subjective internal costs, even if behavior may seem suboptimal to an outside observer. Please note that I do not strive to fully explain all aspects of human planning, but merely to develop a model that is useful for measuring individual differences in the cost of planning.

The model is, in part, an extension of the resource-rational model of human planning put forward by Callaway *et al.* (2022). I therefore first introduce the standard meta-level MDP model of the Mouselab-MDP Task in Subsection 3.1.1. Afterward, I introduce two new additions to the model: how I model the cost of planning (with a new cost parametrization) and the selection of planning operations (meta-level action selection).

3.1.1 Meta-level MDP Model of the Mouselab-MDP Task

We model human planning as an approximately optimal solution to a meta-level Markov Decision Process (MDP; Hay *et al.*, 2014; Griffiths *et al.*, 2019; Callaway *et al.*, 2018). A meta-level MDP $(\mathcal{B}, \mathcal{C}, T_{\text{meta}}, R_{\text{meta}})$ is an MDP where the states $\mathcal{B}$ are the beliefs the agent can arrive at through reasoning and the actions are the computations $\mathcal{C}$ the agent can perform to improve the accuracy of its beliefs. In the case of the Mouselab-MDP paradigm, a belief $b \in \mathcal{B}$ is a probability distribution over the possible values of rewards at the nodes of the environment shown in Figure 2.1. The available computations $\mathcal{C} = \{c_1, c_2, c_3, \cdots \perp\}$ comprise the planning operations $(c_1, c_2, \cdots)$ and the decision to terminate planning $(\perp)$. The planning operation c_i evaluates and processes the reward of the i^{th} node. The decision to end planning selects the path out of the spider web that is best according to the information processed by the preceding planning operations. The transition matrix $T_{\text{meta}}(b, c, b')$ specifies the probability of arriving in belief state b' after performing the planning operation $c \in \mathcal{C}$ in belief state b. For performing a planning operation in the environment described in the previous subsection, $T_{\text{meta}}(b, c, b')$ is uniformly $1/4$ since there are four equally likely values that may be revealed upon clicking a node. Performing the termination operation $\perp$ deterministically leads to a terminal state.

The cost of inspecting a node which is not terminal, that is $R_{\text{meta}}(b, c)$ where $c \in \mathcal{C} - \perp$, is \$1. Participants are told they will have to pay this fee for inspecting a node before beginning the task. The cost of completing a trial and moving the spider through the web, $R_{\text{meta}}(b,\perp)$, is defined as the maximum expected return of any available path according to the current belief state b. Please see Figure 3.1 for a diagram of the meta-level MDP as well as an example time step using the reward distribution described in the previous subsection. The improved model of planning will extend the definition of this meta-level MDP.

3.1.2 An extended model of the cost of planning.

In the meta-level Markov Decision Process described above in Subsection 3.1.1, the cost of performing planning operation c in belief-state b (i.e., $R_{\text{meta}}(b,c)$) is typically assumed to be identical for all non-terminal planning operations. It is typically set to the monetary cost participants actually pay to inspect nodes in a given experiment. However, people may experience additional costs that go beyond the financial cost of clicking on nodes, namely cognitive costs, motor costs, or time costs.

To address this, in this subsection, I introduce five possible cost components which may yield a more realistic and more flexible model of the cost of planning. It is beyond the scope of the present work to develop a comprehensive cost model. Rather, I seek to parametrize the cost function in a way that could be useful for understanding participant behavior and providing a mechanistic account of planning variation.

Note also that I will use the letter c (for "cognitive operations", although "clicks" also works for non-terminal actions) to refer exclusively to non-terminal planning operations (i.e., $c \in \mathcal{C} - \perp$). The lowercase letter $r(b,c)$ will refer to the cost/reward component[1] values in belief state b for cognitive operation c. The full reward function $R_{meta}(b,c)$ will consist of multiple reward components and individual weights, i.e., $R_{meta}(b,c) = \sum_{i=1}^{n} w_i r_i(b,c) = w_1 r_1(b,c) + w_2 r_2(b,c) + \cdots + w_n r_n(b,c)$.

3.1.2.1 Planning Depth

I measure how effortful it is for a person to plan out to further planning depths via the planning depth cost r_{depth}. Let d_n be the depth of node n. For the first level of the web (closest to the start), d_n is 1, while for the leaf nodes (furthest from the start) d_n is 3. Therefore, the depth cost of the planning operation c_n that inspects the node n in belief state b is $r_{depth}(b,c_n) = -d_n$.

This cost is then multiplied by a cost weight w_{depth}. A person with higher w_{depth} might find planning further into the future more difficult and will make short-sighted decisions because they prefer to inspect earlier nodes that represent more proximal outcomes

[1] "Reward" and "cost" here since reward and cost can be used interchangeably in the reinforcement learning setting.

rather than later nodes that represent more distant outcomes. A person with $w_{depth} = 0$ will not be influenced by the depth of a node when planning.

3.1.2.2 Effort Cost

I measure a person's sensitivity to planning in general with the effort cost weight $w_{eff.}$. This cost applies to all non-terminal planning operations, i.e., $r_{eff.}(b, c_n) = -1$. A person with higher $w_{eff.}$ may plan less in general or not perform any planning operations at all, if the cost of planning exceeds the value of planning.

3.1.2.3 Distance Cost

This distance parameter formalizes the alternate explanation that people are not experiencing a cognitive planning depth cost, but instead experience a motor cost from moving the cursor all the way to the outermost nodes. I denote the distance cost weight as $w_{dist.}$. Individuals with a higher distance parameter weight might be more likely to consider options closer in physical space.

The distance itself depends on the distance between the locations evaluated by the current planning operation, c_t, and the previous planning operation, c_{t-1}. I define $l(n_i, n_j)$ as the Euclidean distance between the nodes n_i and n_j inspected by the planning operations c_i and c_j, respectively. If the position of node n_i is (x_i, y_i) then $r_{dist.}(b, c_{n_i}, c_{n_j}) = -l(n_i, n_j) = -(x_i - x_j)^2 - (y_i - y_j)^2$. The first planning operation is assumed to be the spider's starting node n_0. To make certain of this, in all experiments I ran, participants must click on the starting node to begin each trial.

3.1.2.4 Non-Forward Search Cost & Non-Backward Search Cost

The non-forward search cost and non-backward search cost both depend on the relation between a given node and the already-revealed nodes. Recently, there has been evidence that participants may employ forwards search or backwards search strategies in the Mouselab-MDP task, depending on which strategy is more well-suited for the environment they find themselves in (Callaway *et al.*, 2022).

§1 Non-Forward Search Cost Best-first-search has also been hypothesized to be a common strategy people employ when planning (Newell and Simon, 1972). To capture this, the non-forward search cost increases the cost to click on nodes whose parent nodes are not already revealed. This cost captures the idea that there may be individual variation in whether people find it easier to continue planning action sequences they have already thought about than to start making an entirely new plan.

I define $\text{Parents}(n_i)$ as the direct predecessor nodes of n_i, and I define N_{revealed} as the set of already revealed nodes. Please note that as mentioned in 3.1.2.3, the node the spider starts on must be clicked by the participant at the beginning. This node is counted as an uncovered parent node of the first level nodes, although it isn't technically "revealed".

Using this notation, the cost can be written as

$$r_{forw.}(b, c_n) = \begin{cases} 0 & \text{if } \text{Parents}(n) \cap N_{\text{revealed}} \neq \emptyset \\ -1 & \text{otherwise} \end{cases},$$

where n is the node inspected by the planning operation c_n. A person with a higher forward cost weight $w_{forw.}$ should engage in more forward planning.

§2 Non-Backward Search Cost Likewise, the non-backward search cost increases the cost for clicking on nodes whose children (or successor) nodes are not already revealed. The furthest out nodes also do not have any added cost, as someone following a backwards search strategy would click on these first.

Using the notation introduced in §1 and further defining the outermost nodes as $N_{d_n=3}$ the cost is defined as

$$r_{back.}(b, c_n) = \begin{cases} 0 & \text{if } \text{Children}(n) \cap N_{\text{revealed}} \neq \emptyset \lor n \in N_{d_n=3} \\ -1 & \text{otherwise} \end{cases},$$

where n is the node inspected by the planning operation c_n. A person with a higher non-backward search cost weight $w_{back.}$ should engage in more backward planning.

3.1.3 Meta-Level Action Selection

I assumed that people select planning operations according to a softmax policy over an approximation of the state-action values of the meta-level MDP. More specifically, that the probability of observing a planning operation c_t in belief state b_t under reward parameters θ is

$$P(c_t \mid b_t, \theta) \propto e^{\hat{Q}_\theta(b_t, c_t)/\beta}$$

where β is the decision-making temperature and $\hat{Q}_\theta$ is an approximation of the optimal meta-level Q-function for the problem where the parameters of the cost function are given by θ. Participants who are fit better with very high β values (higher temperature) are more likely to select actions almost uniformly at random, whilst participants with lower β are more likely to stick to selecting the best action, according to their subjective state-action values.

I approximate the optimal meta-level Q-function by the myopic value of computation (Russell and Wefald, 1991). This approximation calculates the value of a computation (click) assuming that the optimal policy will terminate planning in the next step:

$$\hat{Q}(b_n, c_n) = \begin{cases} \mathrm{VOI}(b_n, c_n) + R_{\mathrm{meta}}(b_n, c_n) + Q(b_{n+1}, \perp) & \mathrm{c} \in \mathcal{C} - \{\perp\} \\ R_{\mathrm{meta}}(b_n, \perp) & \mathrm{c} = \perp \end{cases},$$

where the myopic VOI is defined as $\mathrm{VOI}(b_n, c_n) = \sum_{b_{n+1}} [T(b_n, c_n, b_{n+1}) \cdot Q(b_{n+1}, \perp)] - Q(b_n, \perp)$. The myopic value of computation approximates the resource-rational planning strategy which corresponds to the Directed Cognition model by Gabaix and Laibson (2005). This model has been found to explain human behavior almost as well as the fully resource-rational model in the original Mouselab paradigm (Gul *et al.*, 2018).

3.1.4 Considered Models

I considered one model for each possible subset of the five cost parameters. This set of models includes a model where the cost is fixed to the imposed financial costs ("Null (Given Costs)"). I also include a model which assumes completely random planning operation selection, naive to the optimal state-action values ("Null (Random)"). This is the only model which does not assume softmax meta-level action selection. To compute the meta-level state-action values, I assumed that the participant know the distribution of possible payoffs at each step of the planning task. Since participants are not informed of the payoff structure, I only fit the data of participants' later trials, when participants have had enough experience to learn the possible payoffs.

3.1.5 Model Fitting Details

When fitting participant data to these models, I performed a grid search on a range of possible model parameter values. The possible values on the grid were chosen based on a range of possible behavior seen in simulations. I individually fit participants, finding an approximation of the maximum-likelihood estimate (MLE) for each participant. Please see Section 4.1 in the following chapter for more details.

3.2 Experiment 1: Which model describes the cost of human planning best?

3.2.1 Introduction

In my first experiment, I aimed to see which cost model best explains real people's planning data. I will use the best model as a key component for my method to measure individual differences in the costs of planning. Therefore, model selection is a crucial step in developing such a method.

3.2.2 Methods

3.2.2.1 Participants

I recruited 164 participants on Amazon Mechanical Turk, using CloudResearch (Litman *et al.*, 2017). 137 participants completed the experiment; the other participants quit during the instructions phase. Of these 137 participants, 15 participants were unable to pass the instructions quiz. So, I have complete data for 122 participants (52 females, 69 males, 1 non-binary; median age 42, age range 20-75).

3.2.2.2 Materials

Each participant completed 40 trials of the Mouselab-MDP task as described in the Section "Mouselab-MDP Paradigm" (see also Figure 2.1). Concretely, I used the exact same version of this task that has previously been used to study the "present bias" (Lieder *et al.*, 2019b). That is, the click cost was 1 game dollar and the possible rewards were $\{-4,-2,2,4\}$ at the first step, $\{-8,-4,4,8\}$ at the second step, and $\{-48,-24,24,48\}$ for the third and final step. At the beginning of this task all participants were given a 50 game dollar endowment. Moreover, participants were incentivized to perform well in this task by a performance-dependent bonus of 1 cent for every \$5 they earned in this game. Participants completed the task in a median time of 13.42 minutes and earned a median \$3.24 bonus in addition to the \$0.50 base pay.

3.2.2.3 Procedures

All participants gave informed consent. Participants were then given instructions and a comprehension test before beginning the task. The comprehension test checked participants' understanding of the task and bonus payment. Participants were only allowed to move on to the task if they passed the pre-task comprehension test within four attempts.

Participants then completed 40 trials of the Mouselab-MDP Task. After that, participants were asked four multiple-choice questions testing their knowledge of the task. The first three questions related to the range of node values for each of the three levels of nodes in the task. The fourth question checked their knowledge of the cost for inspecting a node before moving.

3.2.3 Results

3.2.3.1 Participant behavior was relatively stable for the last half of trials.

Because the model assumes that participant behavior is stable, I checked that participant behavior did not vary too much in the last 20 of the 40 trials. To do so, I used Jain et al. 's Computational Microscope to infer the most likely strategies the participant used in each trial (Jain *et al.*, 2022). I found that, on average, participants used less than two different strategies in the last half of trials (M = 1.951, SD = 1.170). Moreover, around half of the participants never changed their strategy during the last half of the trials and $\approx 86\%$ of participants used at most three different strategies during this period (see Subsection A.1.1 in the appendix for further details on the Computational Microscope results). I can therefore apply my method to the last 20 trials of the experiment.

3.2.3.2 The full cost model explains participants' data best.

To determine which cost function my method should assume, I compared all the models using both the Bayesian Information Criterion (BIC; Schwarz, 1978) and Bayesian Model Selection (BMS; Stephan *et al.*, 2009; Rigoux *et al.*, 2014). To compute Bayes factors (Kass and Raftery, 1995), I used the Bayesian Information Criterion as an approximation of the negative log-model evidence (i.e., $-\log(P(\text{Model} \mid \text{Data}))$; Claeskens and Hjort, 2008). The model comparison based on the Bayesian Information Criterion was a fixed-effects analysis, looking for the best model for all participants, whereas the Bayesian Model Selection method performed a random effects analysis, which takes into account that different participant's data may be better explained by different models.

My fixed-effects analysis provided decisive evidence that the full cost model explains the data better than any of the alternative models (all log-Bayes factors ≥ 156; see Figure 3.2a). The second-best model was the model with all cost parameters except the effort cost weight.

I next conducted Bayesian model selection, which is a mixed-effect analysis. This means that I calculate the fit per model per participant, rather than for all the participants combined. Bayesian model selection (see Table 3.1) favored simpler models with one or two cost parameter weights rather than the full model, with models containing only the distance weight parameter explaining the most (19%) participants' planning behavior. The second-best model, the model containing only the effort cost weight, best explained the planning behavior of approximately 16% of the participants. The full cost model had a model probability of only 1%. The model which consists of only the non-forward search cost weight and effort cost weight which mimics the state-of-the-art model proposed by Callaway *et al.* (2022) had a similar model probability of 1%. In fact, several of the newly considered models explained human planning better than this model.

Taken together, these results suggest that the full cost model may be more appropriate when applying the same model to all subjects. With this result, I will continue with the full cost model in my efforts to extract individual differences in cost weights.

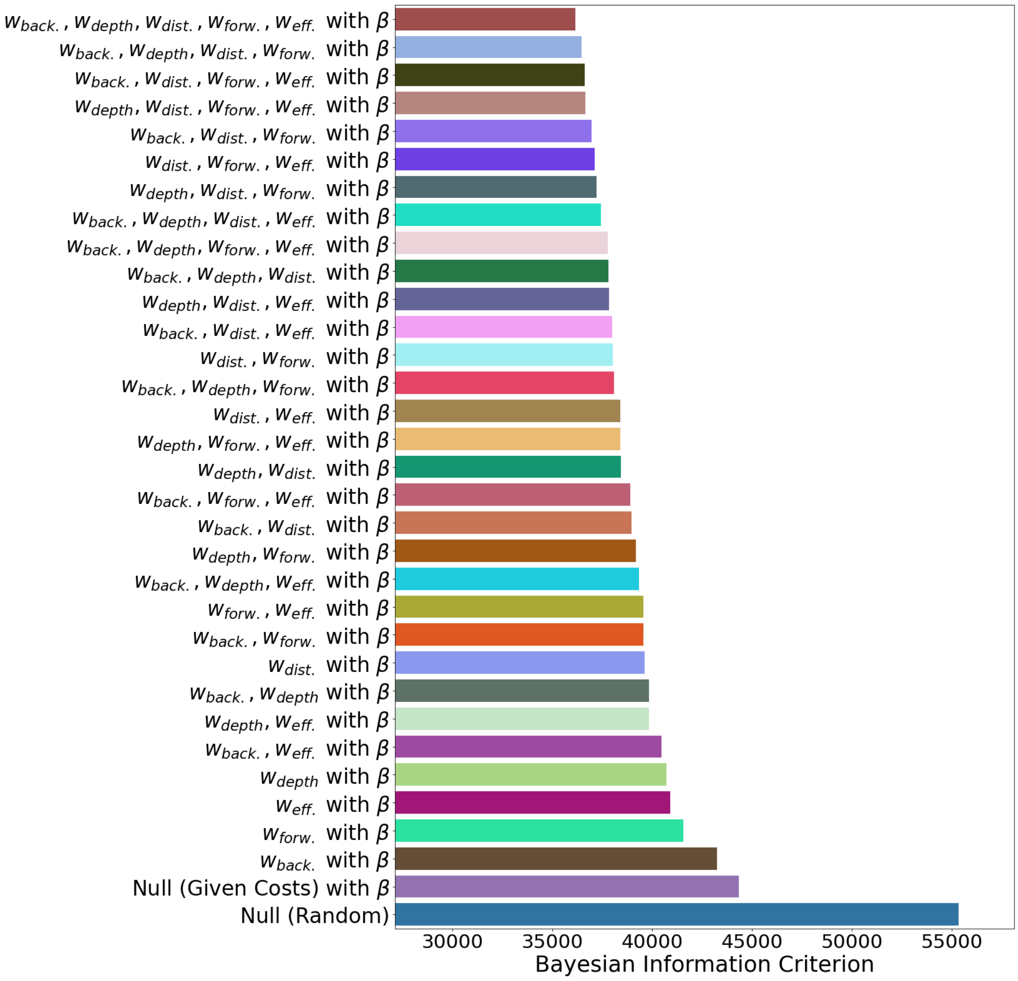
3.2.3.3 The full cost model is best when considering all trials.

Recall that I specifically inferred costs for the last 20 trials of each participant, when participants should be less engaged in structure-learning. There was no correlation between the average planning likelihood and time for the full cost model over the last 20 trials (Spearman's $\rho(2440) = 0.03$, $p = 0.089$, 95% C. I. $[-0.01, 0.07]$). Additionally, there was also only a slight correlation between the average planning likelihood and time for the full cost model (Spearman's $\rho(4880) = 0.06$, $p < 0.001$, 95% C. I. $[0.04, 0.09]$; see Figure 3.2b). This suggests that the earlier trials, at least absolutely, are just as well fit as later trials.

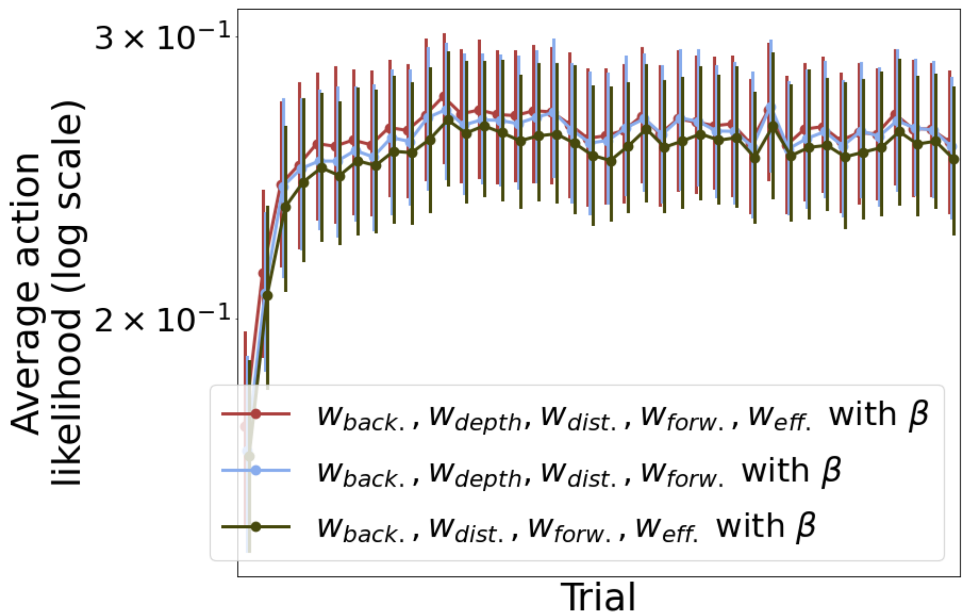
I looked further at the average action likelihood for participants in Experiment 1 using the MLE parameter estimates based on the last twenty trials ($\mathrm{M} = 0.37, \mathrm{SD} = 0.17$) versus all forty trials ($\mathrm{M} = 0.26, \mathrm{SD} = 0.14$). Using a two-sided Wilcoxon signed-rank test, I found that the average action likelihood was significantly higher with the parameters inferred from the last half of trials than with the parameters inferred from all trials ($W = 211292.00$, $\mathrm{RBC} = 0.83$, $p < 0.001$, two-sided). However, I also found that the full cost weight model still fits the data best, compared to the alternative models tested in Experiment 1, when using all 40 trials (see Figure 3.3). All in all, this suggests that while the best model would be the same regardless of the trials considered, behavior is more constant in the last 20 out of 40 trials.

3.2.3.4 The behavior of simulated participants matches participant behavior.

To check which aspects of people's planning behavior the full cost model can capture, I looked at the average number of clicks participants used to inspect immediate, intermediate, and final outcomes (clicks per level) for participants versus the fitted model as a form of posterior predictive check (Roecker, 1991; Gelman *et al.*, 1996; see Wilson and



(a)



(b)

Figure 3.2: (a) The Bayesian Information Criteria for each model. The lower the value, the better the fit. The vertical line denotes the Bayesian Information Criterion of simulated participants with costs matched to those of the winning model. (b) The average trial log likelihood for each trial, for the top three models.

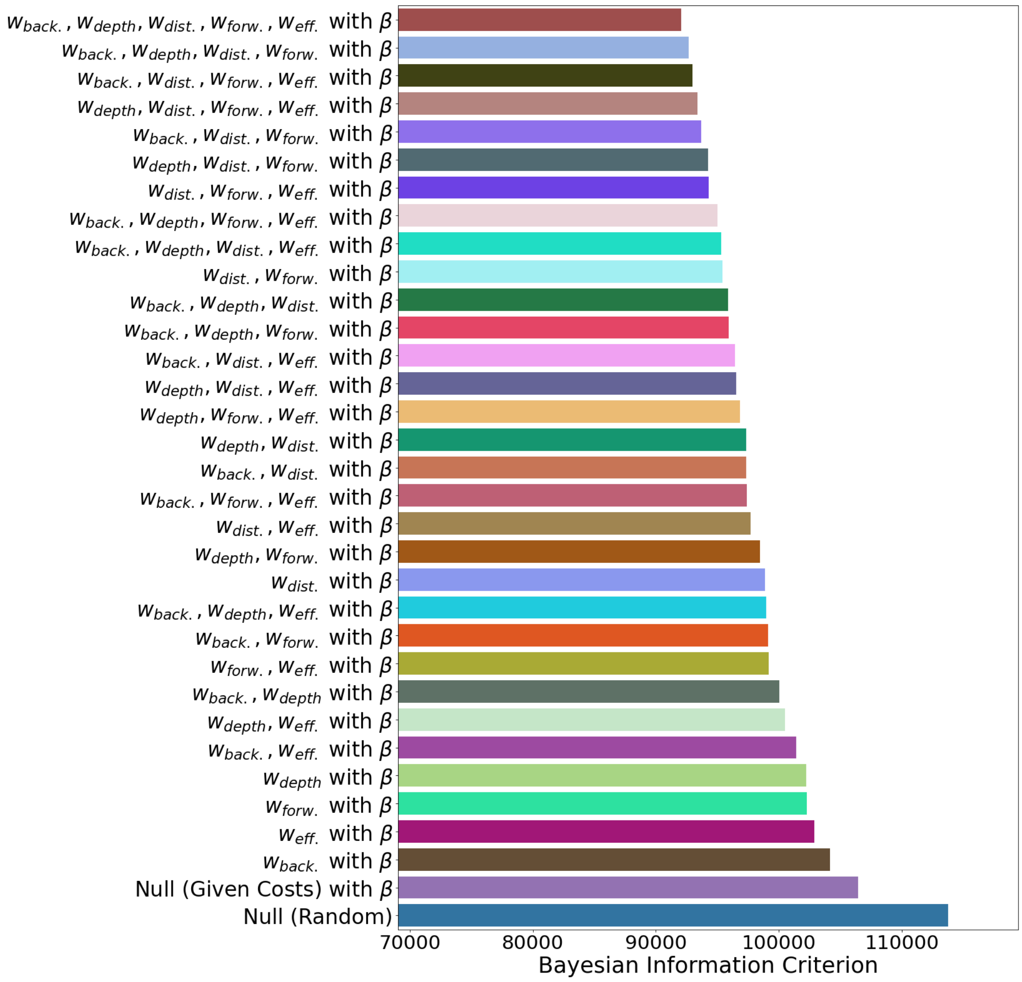


Figure 3.3: Bayesian information criterion when inferring MLE parameter estimates from all 40 trials.

Collins, 2019). Specifically, for each possible cost weight combination, I simulated 20 trials for optimal participants and computed the number of clicks per level, per trial for each simulated participant. I then computed the average node-level click rate for each cost weight combination for these simulated participants. To compare simulated and real data, I computed the correlation between click behavior for each participant and the average simulated click behavior of the model with cost weight values matching each participants' MLE parameters.

For the full cost model, I found that there was a high Pearson correlation for all metrics of planning behaviors I considered: clicking on first-level nodes (Spearman's $\rho(122) = 0.76$, $p < 0.001$, 95% C. I. $[0.67, 0.82]$), clicking on middle nodes (Spearman's $\rho(122) = 0.65$, $p < 0.001$, 95% C. I. $[0.54, 0.74]$), clicking on last-level nodes (Spearman's $\rho(122) = 0.59$, $p < 0.001$, 95% C. I. $[0.46, 0.69]$), and the total number of nodes clicked (Spearman's $\rho(122) = 0.58$, $p < 0.001$, 95% C. I. $[0.44, 0.68]$). This suggests that participant data and simulated data from the full-parameter model exhibit similar node-level clicking behavior. This strengthens the argument for using the full-parameter model to infer individual differences in planning behavior. Additionally, I found similar correlations for the two next best models according to BIC. See Subsection A.1.2 in the Appendix for further details.

3.2.3.5 Participants who made fewer clicks still understood the environment's structure reasonably well.

These models assume that people know the structure of the environment. Here, that means that I assumed people already know the possible values behind each node (and their respective probabilities) by the time they start the last 20 trials. This raises the question of whether the planning behavior of people with higher cost weight values, who may not click enough to learn the environmental structure, might simply be less well explained by these models. To look into this, I inspected the average log likelihoods of participants' planning operations as a function of their score on the four post-task questions.

I found that the average log likelihood of a planning operation did *not* differ significantly between participants who answered all four post-task questions correctly versus those who answered at least one post-task question incorrectly (Kruskal-Wallis $H(3) = 1.11$, $p = 0.775$). Figure A.2 in the Appendix shows the post-task quiz performance versus the average planning operation likelihood. This suggests that the model explains participants who lacked knowledge of the environment just as well as those who were able to answer the post-task questionnaire completely correctly. However, it could also be, that the short post-task questionnaire was not comprehensive enough to capture real differences in environment/task understanding.

Participants' scores on the post-task quiz also had no correlation with the MLE parameter estimates for any cost weight or the temperature parameter ($w_{forw.}$: Kruskal-Wallis $H(3) = 1.53$, $p = 0.675$; $w_{back.}$: Kruskal-Wallis $H(3) = 3.66$, $p = 0.301$; $w_{eff.}$: Kruskal-Wallis $H(3) = 0.81$, $p = 0.848$; $w_{dist.}$: Kruskal-Wallis $H(3) = 5.42$, $p = 0.143$; w_{depth}: Kruskal-Wallis $H(3) = 1.36$, $p = 0.716$; β: Kruskal-Wallis $H(3) = 2.05$, $p = 0.563$.) This suggests there is no pattern between cost parameters and poor understanding of the task environment.

As a final check, I investigated whether post-task scores correlated with average number of clicks per trial. I did not find a significant difference between any of the score groups (Kruskal-Wallis $H(3) = 1.30$, $p = 0.729$). This again could be due to the structure of the post-task questionnaire. Since the post-task questionnaire contained multiple choice answers, it could be that participants who almost understood the structure were able to rule out the other answers to the reward questions.

3.2.4 Conclusion

In this experiment, I established that the full cost model fits participant data best according to BIC, compared to my other candidate models. In particular, this model fits the data better than the alternate model in the spirit of the current state-of-the-art resource-rational model introduced by Callaway *et al.* (2022). This suggests that the improved expanded cost model might better fit data in the Mouselab-MDP task than previously-applied models, especially those with simpler cost models.

I additionally validated the full cost model by showing that simulated data from the model approximately matched participant behavior. I found no evidence that the model fit depends on the degree to which participants understand the task. I will therefore base my method for measuring the cost of planning on the full cost weight model.

3.3 Does the environment structure matter for model fit?

In the previous experiment, I gave people a task where further out nodes had higher variance, which is meant to mirror real-life planning and decision-making. For example, when deciding whether to do some physical activity, the reward or pain one feels on a day-to-day level now is usually at a much lower magnitude than that which one may feel later in life as a consequence of (not) enacting one's plan to stay fit and healthy.

However, what if the higher variance outcomes were at the beginning? What if there was no connection between variance and timescale? Both of these environments may be encountered in decision-making. In fact, in more short-term decisions or decisions in more random (volatile) environments, there may not be a big difference in variance at different time scales. In this section, I set out to test, using an existing dataset, if the model selection results reported in the previous section hold for environments with different variance structures.

Here, I fit participant data from these two participant environments to the same possible cost functions as the previous experiment. The dataset includes data for participants planning in either an environment with no variance difference between node levels or one with decreasing variance with node depth. In both cases, one might expect that the backwards search cost no longer plays a role. In fact, in the environment with higher variance outcomes at the beginning, perhaps people would follow a forward search strategy.

3.3.1 Methods

Participant data were generously provided by Ruiqi He from her work looking at how the development of maladaptive versus adaptive planning strategies varies based on the reward-structure of the environment (He *et al.*, 2021).

3.3.1.1 Participants

As previously reported, 116 participants were recruited on Amazon Mechanical Turk, through CloudResearch (Litman *et al.*, 2017).

3.3.1.2 Materials

Participants played 35 trials of the Mouselab-MDP task. Each participant was randomly assigned to one of two possible reward values: decreasing variance and constant variance. In the decreasing variance condition, the values were inverted from those in Experiment 1. The possible rewards were $\{-48,-24,24,48\}$ at the first step, $\{-8,-4,4,8\}$ at the second step, and $\{-4,-2,2,4\}$ for the third and final step. In the constant variance condition, the possible values were the same at each node ($\{-10,-5,5,10\}$.) Participants received a \$1.50 base payment and a bonus of up to \$5.00, depending on performance (specifically, \$0.002 per game point).

3.3.1.3 Procedures

All participants gave informed consent and were then presented with the task instructions. Similar to Experiment 1 in this chapter, participants were tested on their comprehension of these instructions with a brief instructions quiz. After passing the instructions quiz, participants were allowed to continue with the task. All participants saw 35 Mouselab-MDP trials for their assigned condition.

3.3.2 Results

Similar to the analyses for Experiment 1, I ran, for each condition, a fixed-effect model comparison using the BIC and the random-effect model comparison using BMS. I considered all possible combinations of models as set out in Subsection 3.1.4. As in Experiment 1, I fit the models to only the last 20 trials, for each participant.

3.3.2.1 The full cost model explains participant data best in the decreasing variance condition

For the decreasing variance condition, I found the model with all five cost weights performed best, with a log Bayes factor of approximately 31 in favor of this model compared to the second-best model which did not includes the depth cost weight w_{depth} (see Figure 3.4). The third-best model was the model not including $w_{back.}$, making the second and third best models the same as in Experiment 1, but with a different ordering. For the Bayesian Model Selection, I found the largest group of participants (~ 9.92) were best-explained by the $w_{back.}, w_{forw.}, w_{eff.}$ model (see Table A.6). The Null (Given) model best explained ~ 2.01 participants, while the Null (Random) model best explained 0 participants.

3.3.2.2 The full cost model explains participant data best in the constant variance condition

For the constant variance condition, I found the model with all five cost weights performed best, with an approximate log Bayes factor of at least 48 between all other models (see Figure 3.5). This time the second-best model was that without w_{depth} and the third-best that without $w_{dist.}$. Perhaps due to the fact that there was no difference in variance among levels, I found that many participants were best-explained by submodels containing the $w_{eff.}$ parameter in Bayesian Model Selection (see Table A.7). The largest group of participants (~ 10.30) were best explained by the $w_{dist.}, w_{forw.}, w_{eff.}$ model. The Null (Given) model best explained ~ 0.04 participants, while the Null (Random) model best explained 0.01 participants.

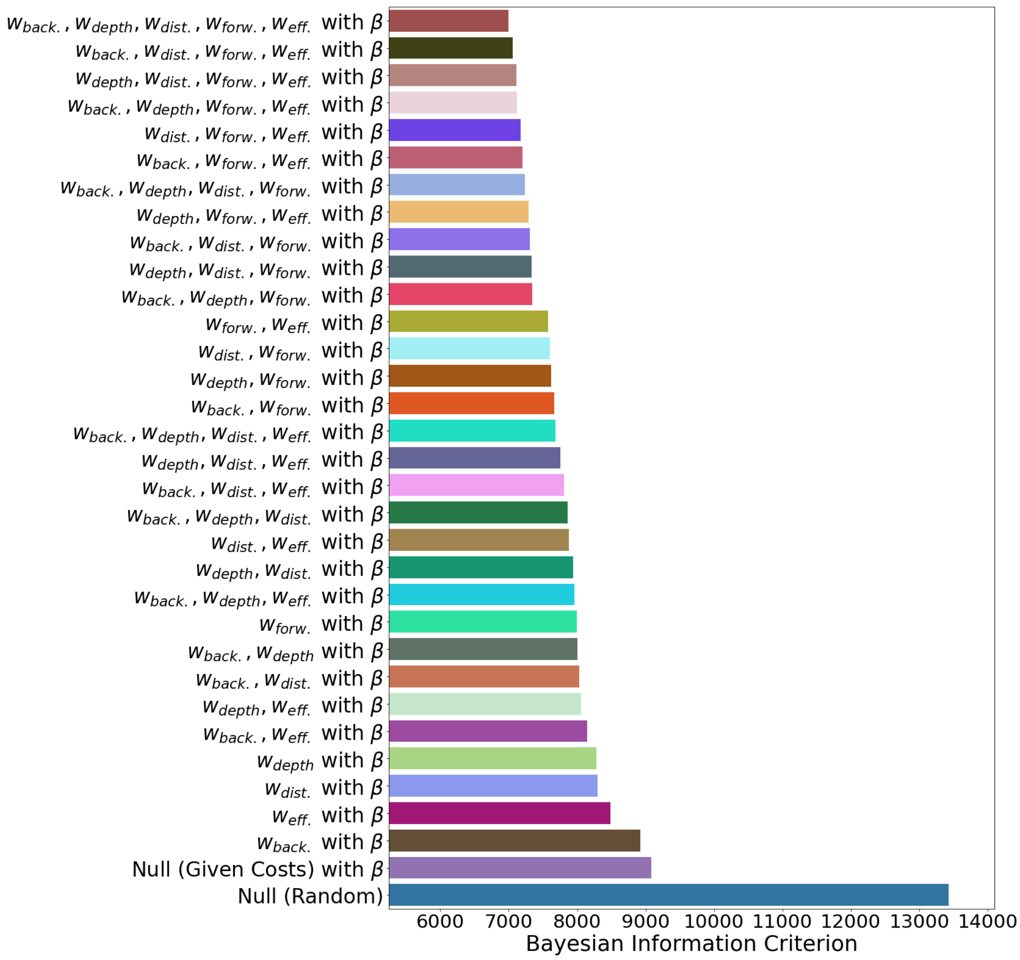


Figure 3.4: The fixed-effect model comparison for the “decreasing variance” environment.

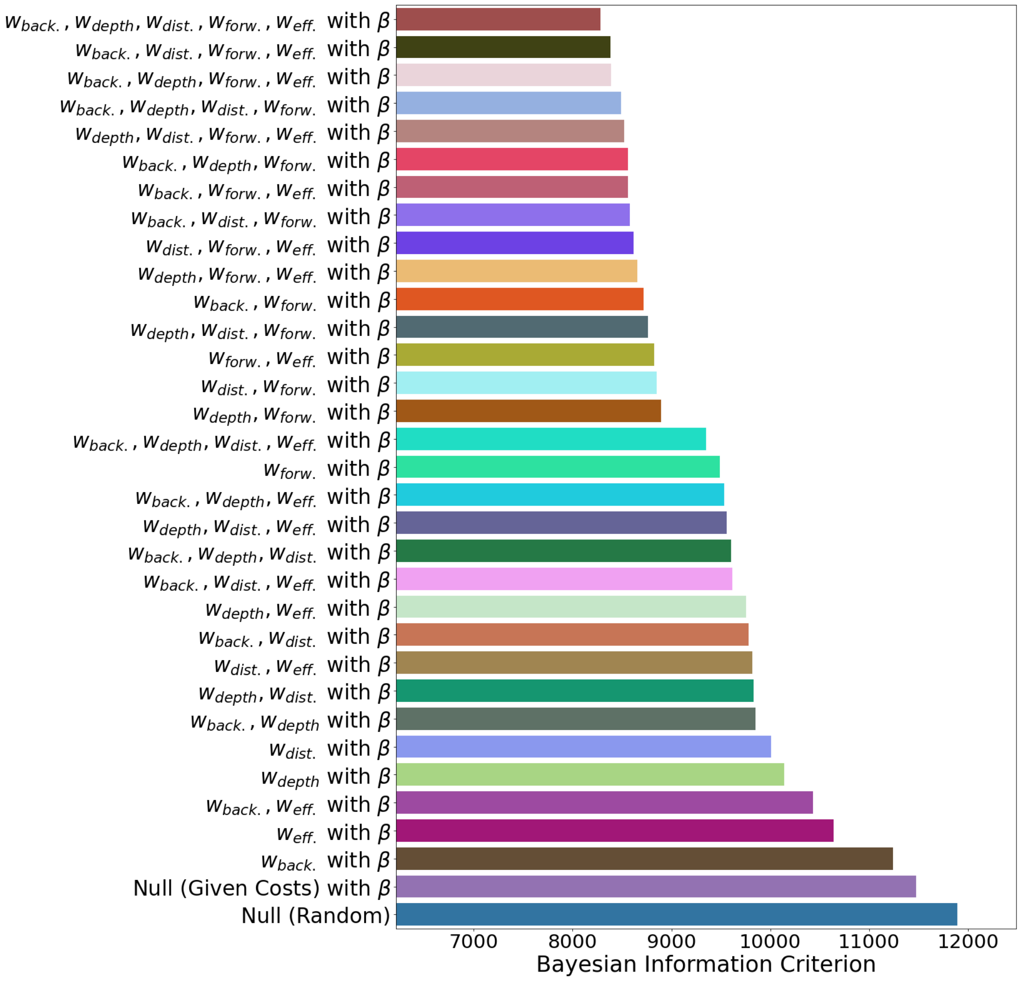


Figure 3.5: The fixed-effect model comparison for the “constant variance” environment.

3.3.2.3 Most cost weights do not differ significantly based on condition.

Next, I ran Mann-Whitney tests to see if the cost weights estimated in these experiments were different from the cost weights estimated in Experiment 1 (i.e., the "high increasing variance" environment). I corrected for multiple comparisons using a Benjamini-Hochberg correction with a false discovery rate of 0.05 (Benjamini and Hochberg, 1995). Between Experiment 1 and the decreasing variance condition, I found that the distance cost weight was higher in Experiment 1 (Mann-Whitney $U = 4425.00$, RBC $= -0.25$, adj.$p = 0.019$, two-sided; $M_{\text{Experiment 1}} = 2.139$, $M_{\text{Decreasing}} = 1.353$). This may reflect the fact that, in these experiments, participants' cursors did not always start at the spider's starting node. Non-backward search cost weight was higher in Experiment 1, mirroring the fact that backward search is a more optimal strategy in Experiment 1 (Mann-Whitney $U = 4694.50$, RBC $= -0.33$, adj.$p = 0.001$, two-sided; $M_{\text{Experiment 1}} = 4.291$, $M_{\text{Decreasing}} = 1.241$). Of a bit of concern, is that the temperature parameter was higher for Experiment 1 than the decreasing variance experiment (Mann-Whitney $U = 4292.50$, RBC $= -0.21$, adj.$p = 0.033$, two-sided; $M_{\text{Experiment 1}} = 7.580$, $M_{\text{Decreasing}} = 3.328$). This implies that participants behaved more randomly in Experiment 1.

I found that the MLEs for the distance cost weight for participants in the low constant dataset were lower (Mann-Whitney $U = 4798.00$, RBC $= -0.36$, adj.$p < 0.001$, two-sided; $M_{\text{Experiment 1}} = 2.139$, $M_{\text{Constant}} = 0.698$). Again, this might be based on a difference in methodology, as participants were not made to click on the spider's starting node to begin a trial. I also found that the non-backward search cost weight was higher in Experiment 1, again mirroring the difference in optimal strategies (Mann-Whitney $U = 4671.50$, RBC $= -0.32$, adj.$p = 0.001$, two-sided; $M_{\text{Experiment 1}} = 4.291$, $M_{\text{Constant}} = 2.095$). I found no significant difference in parameter estimates between the MLEs for the decreasing variance dataset and the constant variance dataset as well as between Experiment 1 and the constant variance dataset.

3.3.3 Conclusion

In this section, I investigated whether and how cost models can explain planning behavior in two very different planning environments: one where forward search is optimal (the decreasing variance dataset), and another where planning is less beneficial and not depth-dependent (the constant variance dataset). I found that the five parameter model which best explained participant data in Experiment 1 also best explained planning behavior in both of these environments based on BIC. Additionally, I found a difference in the non-backward search cost weight between environments which implies that parameter estimates may be influenced by the environment structure. However, I also found a difference in the distance cost weight between environments which may be caused by a difference in experiment design between Experiment 1 and these borrowed datasets. All in all, despite the fact that distance cost weight parameters may not be comparable between these datasets, the results indicate that the full cost parameter model works for other environments.

3.4 Discussion

In this chapter, I presented a new model of individual differences in planning in the Mouselab-MDP task. This model is resource-rational, meaning it assumes optimal planning under some constraints on resources. A model incorporating all five different planning costs was selected as the best choice according to the Bayesian Information Criterion. I also showed that this model was the best choice by Bayesian Information Criterion in two additional task settings with different reward distributions.

This model extends the best previous model (Callaway *et al.*, 2022). While there are differences in the implementation, it better explains participant data than the submodel in the spirit of this model. This generative model is also more parsimonious than the method developed by Jain *et al.* (2022). This model, therefore, provides a compromise between two existing model types of behavior in the task.

One assumption of this model is that participants choose their planning operations approximately optimally. This assumes that people's planning strategies are optimally adapted to the structure of the environment, that is, that they know the distribution of possible rewards at different locations. While this becomes a reasonable assumption after a sufficient amount of experience, this assumption may not hold earlier in the learning process, or in cases where participants fail to discover the true structure of the environment.

Another limitation of my work is that I performed parameter estimation over a rather limited grid of possible values. This was due to the higher computational complexity of calculating the state-action values in a previous implementation. I could not change the inference method because of time constraints. Future work may extend my implementation to use continuous optimization, which would make it easier for other researchers to use the method introduced in the next chapter.

The model assumes that differences in behavior stem only from differences in planning costs. As pointed out by a reviewer, another potential cause of reduced planning may be differences in discount rate and utility function (Tymula *et al.*, 2013). These two additions deal with differences in perceived value, rather than dealing with cost like the model discussed in this chapter. In a supplementary analysis, I found that while augmenting the model with a discount rate and a power utility function with a risk aversion parameter (see Subsection A.1.3) yielded lower model BIC, this new model did not match participant behavior. Further investigating participants with differing values of inferred discount rates and risk aversion parameters may be a fruitful topic for future work.

Despite these limitations, I find the full model can be used to measure individual differences in planning behavior. Measuring individuals' subjective planning cost weights may give insight into their planning. In the next chapter, I will look at combining the model developed in this chapter with Bayesian inference to develop a method which extracts these planning cost weights.

Table 3.1: Bayesian Model Selection Results for Experiment 1.

Model	Expected number of participants best explained by the model	Model probability	Exceedance probabilities
$w_{dist.}$ with β	22.63	0.15	0.58
$w_{eff.}$ with β	19.89	0.13	0.29
$w_{dist.}, w_{forw.}$ with β	15.91	0.11	0.07
$w_{back.}, w_{dist.}$ with β	14.39	0.10	0.04
$w_{dist.}, w_{eff.}$ with β	13.28	0.09	0.02
$w_{back.}, w_{forw.}$ with β	6.86	0.05	0.00
$w_{depth}, w_{dist.}, w_{forw.}$ with β	6.01	0.05	0.00
$w_{depth}, w_{forw.}$ with β	4.13	0.03	0.00
$w_{depth}, w_{forw.}, w_{eff.}$ with β	3.74	0.03	0.00
$w_{back.}, w_{forw.}, w_{eff.}$ with β	2.05	0.02	0.00
$w_{depth}, w_{dist.}$ with β	1.92	0.02	0.00
w_{depth} with β	1.91	0.02	0.00
Null (Random)	1.86	0.02	0.00
$w_{back.}, w_{eff.}$ with β	1.74	0.02	0.00
$w_{dist.}, w_{forw.}, w_{eff.}$ with β	1.48	0.02	0.00
$w_{back.}$ with β	0.82	0.01	0.00
$w_{back.}, w_{depth}$ with β	0.53	0.01	0.00
$w_{back.}, w_{dist.}, w_{eff.}$ with β	0.49	0.01	0.00
$w_{depth}, w_{eff.}$ with β	0.45	0.01	0.00
$w_{back.}, w_{dist.}, w_{forw.}$ with β	0.33	0.01	0.00
$w_{depth}, w_{dist.}, w_{forw.}, w_{eff.}$ with β	0.31	0.01	0.00
$w_{depth}, w_{dist.}, w_{eff.}$ with β	0.19	0.01	0.00
$w_{forw.}, w_{eff.}$ with β	0.19	0.01	0.00
$w_{back.}, w_{depth}, w_{dist.}$ with β	0.17	0.01	0.00
$w_{back.}, w_{depth}, w_{forw.}$ with β	0.15	0.01	0.00
$w_{back.}, w_{depth}, w_{eff.}$ with β	0.11	0.01	0.00
$w_{back.}, w_{dist.}, w_{forw.}, w_{eff.}$ with β	0.11	0.01	0.00
$w_{back.}, w_{depth}, w_{dist.}, w_{forw.}$ with β	0.09	0.01	0.00
$w_{back.}, w_{depth}, w_{forw.}, w_{eff.}$ with β	0.09	0.01	0.00
$w_{forw.}$ with β	0.06	0.01	0.00
$w_{back.}, w_{depth}, w_{dist.}, w_{eff.}$ with β	0.04	0.01	0.00
$w_{back.}, w_{depth}, w_{dist.}, w_{forw.}, w_{eff.}$ with β	0.03	0.01	0.00
Null (Given Costs) with β	0.01	0.01	0.00

Chapter 4

Measuring individual differences in planning cost weights

Given the model selection results reported in the previous chapter, I now set out to show that the combination of the full cost model and Bayesian inverse reinforcement learning can produce meaningful parameter estimates for participants performing the task. This method could then allow researchers to quantify differences in people's planning behavior in a way that, compared to traditional self-report studies, is less susceptible to social-desirability bias and demand characteristics (Orne, 1996). This method would also be more parsimonious than existing methods for classifying individual differences in planning such as the Computational Microscope (Jain *et al.*, 2022).

First, I introduce more details of the method. I show, with simulated data, that the method yields good parameter recovery. I present data that shows that the chosen method accurately quantifies its uncertainty about its parameter estimates. Next, I investigate model recovery between the sub-models. I find that model-recovery is not particularly high, providing further evidence that the method should use the full cost model. After establishing that my method performs well on simulated data, I conduct an experiment with assigned planning costs to test whether the method can accurately and precisely infer experimentally imposed inter-individual differences in planning depth cost weight. Finally, I present work showing that a random effects extension to the method may yield better recovery.

Much of the content of this chapter is from the section "Measuring individual differences in the cost of planning" in my preprint available here: https://psyarxiv.com/xmf3y/.

4.1 Method

The method described in this chapter extends upon the model from the previous chapter, and can be formalized as an application of Bayesian inverse reinforcement learning. Therefore, I provide the necessary additional background in this section before moving onto validating the method.

4.1.1 Applying Bayesian inverse reinforcement learning to the Mouselab-MDP paradigm

We are interested in finding the posterior distribution of the cost function parameters and the temperature parameter given the observed data, that is, $P(\vartheta \mid \tau)$ where $\vartheta = \{\theta, \beta\}$ and τ is a trajectory of state-action tuples $\tau = ((b_1, c_1), \ldots (b_T, c_T))$. According to Bayes rule, the posterior distribution is proportional to the product of the likelihood function $P(\tau \mid \vartheta)$ and the prior distribution $P(\vartheta)$, that is $P(\vartheta \mid \tau) \propto P(\tau \mid \vartheta) \cdot P(\vartheta)$.

To compute the likelihood function, we can use the fact that I am assuming people are selecting planning operations according to a softmax policy, the Markov property, and that the initial belief state is constant and the transition structure is deterministic. Putting these together, the likelihood of a sequence of planning operations is the product of the likelihoods of the individual planning operations:

$$
\begin{aligned}
P(\tau \mid \vartheta) &= P(b_1, c_1 \ldots b_T, c_T \mid \vartheta) \\
&= P(b_1)P(c_1 \mid b_1, \vartheta) \cdot \cdots \cdot P(b_T \mid b_{T-1}, c_{T-1}, \vartheta) \cdot P(c_T \mid b_T, \vartheta) \\
&= \prod_{t=1}^{T} P(c_t \mid b_t, \vartheta)
\end{aligned}
\tag{4.1}
$$

4.1.2 Prior

I estimated the cost function parameters θ and the temperature parameter β via grid search. To construct the grid, I chose the range of possible values for the cost function parameters to capture a reasonable range of planning behavior. I determined this by simulating trajectories under the optimal policy for different cost function parameter combinations. I then looked at the number of different types of clicks made (see Figures A.3- A.6 in the appendix). For the temperature parameter, I chose a range of values that allowed me to explain planning behavior that was almost completely deterministic according to the optimal policy as well as completely random planning behavior. Specifically, $\beta \in \{0.1, 1, 10, 100\}$ and for all cost weights w, $w \in \{0, 1, 2.5, 5, 7.5, 10\}$.

4.1.3 Inference

Given some observed planning behavior, the method then uses the prior distributions and the likelihood function in Equation (4.1) above to compute the posterior probability of each parameter combination on the grid. It then selects the parameter combination with the highest posterior probability as the maximum a posteriori (MAP) estimate of the participant's cost function and temperature parameters, that is $\hat{\vartheta}_{MAP} = \text{argmax}_{\vartheta} P(\tau \mid \vartheta)P(\theta)P(\beta)$. Since I assumed a uniform prior on parameter values, the inferred value is the maximum likelihood (MLE): $\hat{\vartheta}_{MLE} = \text{argmax}_{\vartheta} P(\tau \mid \vartheta)$.

Just like in standard Bayesian inverse reinforcement learning method (Ramachandran and Amir, 2007) the method assumes that each participant uses the same planning strategy throughout all trials. Therefore, it should only be applied to the trials after the learning period, once participants' planning behavior has stabilized.

4.1.4 Quantifying the posterior distributions of individual variables

Applying Bayesian inverse reinforcement learning, as described in Subsection 4.1.1, computes the MLE estimate $\hat{\vartheta}_{MLE}$. This MLE estimate is a tuple which contains the temperature parameter and the cost weights. While using the MLE estimate for the parameters can be useful for characterizing individual differences in the cost weights and the temperature parameter, I am also interested in extracting estimates for individual parameters and quantifying their uncertainty. For example, in a setting where possible values are $0, 1, 2$ two subjects may both have an MLE value of 1 with 55%, but the first subject's second most likely term may be 0 with 45% and the other subject's second most likely term may be 1 with 45%. In this toy example, these two participants have very different profiles despite having the same MLE value.

To find the marginal posterior probability distribution of one parameter, say θ_1, I marginalize out the other parameters. I then computed the highest density interval (HDI) of the posterior distribution for that parameter (e.g., $P(\theta_1 \mid \tau)$) by selecting the most probable parameter value, and then greedily including the maximum not-yet-selected neighboring values until the probability of included values reaches 95% (Held and Bové, 2020). I quantified the spread of the highest posterior density interval I as $\max(I) - \min(I)$ (i.e., the smaller the spread, the more confident the method is in the estimate).

4.2 The fixed effects method yields similar results to random effects methods.

In Chapter 3, the full cost model was favored in fixed effects analyses at the group level. Via random effects analyses, however, the expected number of participants best explained by the model was a mere 0.03. In this section I look at parameter recovery for simulated datasets for three possible method outputs: the MLE parameters of the full cost model, the MLE parameters of the best model per participant and a weighted (by posterior model probabilities) aggregate of the MLE parameters for all models.

4.2.1 Possible method outputs

I previously defined the method for using the MLE parameters of the full cost model (see Subsection 4.1.3). In this subsection, I outline the calculation of the two additional method outputs considered. Since I consider the model fit per participant, I will call these methods “Random effects” method outputs (weighted or non-weighted) versus the traditional “Fixed effects’ method output.

For the non-weighted random effects output, I first run Bayesian model selection on the participant data. I then take, for each participant, the model with the highest posterior belief that the model generated the data. For this possible method output, I then take the MLE parameters for this model. Therefore, for participant i with data d_i I define the best model $\mu_i = \arg\max_{m\in\{1,2,\ldots,M\}} P(\mathrm{m}|d_i)$. Then, the output of the method would be $\hat{\vartheta}_{\mathrm{RE}} = \hat{\vartheta}^{\mathrm{MLE}}_{\mu_i}$.

For the weighted random effects output, I again run Bayesian model selection on the participant data. I then take, for each participant, a weighted aggregate, by the posterior beliefs that the model generated the data, for participant i: $\hat{\vartheta}_{\mathrm{WRE}} = \sum_{\mathrm{m}=1}^{M} P_i(\mathrm{m}|d_i)\,\hat{\vartheta}^{MLE}_{\mathrm{m}}$.

4.2.2 Simulation data

To simulate data, I wanted to use realistic and feasible parameter value combinations. Therefore, I calculated $\hat{\vartheta}_{\mathrm{MLE}}$, $\hat{\vartheta}_{\mathrm{RE}}$ and $\hat{\vartheta}_{\mathrm{WRE}}$ for each participant from Experiment 1. Then, for each participant’s parameter values, I simulated 20 trials.

4.2.3 Results

4.2.3.1 None of the methods is statistically better than the others.

Next, to test if one method yields better parameter recovery than the others, I employed a mixed-effects model. The mixed-effects model included main effects of inferred parameter value and method type, as well as the interaction effects between inferred parameter value and method type when trying to predict the true parameter value. It also included random intercepts and slopes for the parameter type (e.g., depth cost weight) and the dataset the method was applied on.

The outcome variable (the true parameter value) was standardized to have mean of 0 and standard deviation of 1. To control for within group variations, inferred parameter value was similarly standardized for each analysis type.

Neither of the interaction terms (inferred parameter value and non-weighted random analysis or weighted random analysis) were significant (see Table 4.1). However, the main effect of the inferred parameter value was found to significantly predict the true parameter value ($\beta = 0.826$, $t(6.760) = 7.705$, $p < 0.001$). This implies that the inferred parameter value can predict the true parameter value. This is good news as it is something I would hope is true for the method.

4.2.3.2 Conclusion

Base on the results in this section, I continue to use the fixed effects method in the remainder of this dissertation. However, it should also be noted that a fully Bayesian approach would have aggregated the posterior distributions instead of the MLEs, as the weighted random effects method presented here did. Future work should look at comparing this fully Bayesian approach to the methods presented here.

Table 4.1: Variance explained in true parameter value, for reported mixed effects analysis.

Predictor	Estimate	Standard Error	T-statistic	P-value
Intercept	-0.004	0.035	-0.114	p = 0.912
Inferred parameter value	0.826	0.107	7.705	p < 0.001
Method (Non-weighted random)	-0.02	0.047	-0.418	p = 0.693
Method (Weighted random)	0.006	0.05	0.115	p = 0.913
Interaction: Inferred parameter value and method (Non-weighted random)	-0.031	0.087	-0.349	p = 0.739
Interaction: Inferred parameter value and method (Weighted random)	0.153	0.102	1.498	p = 0.183

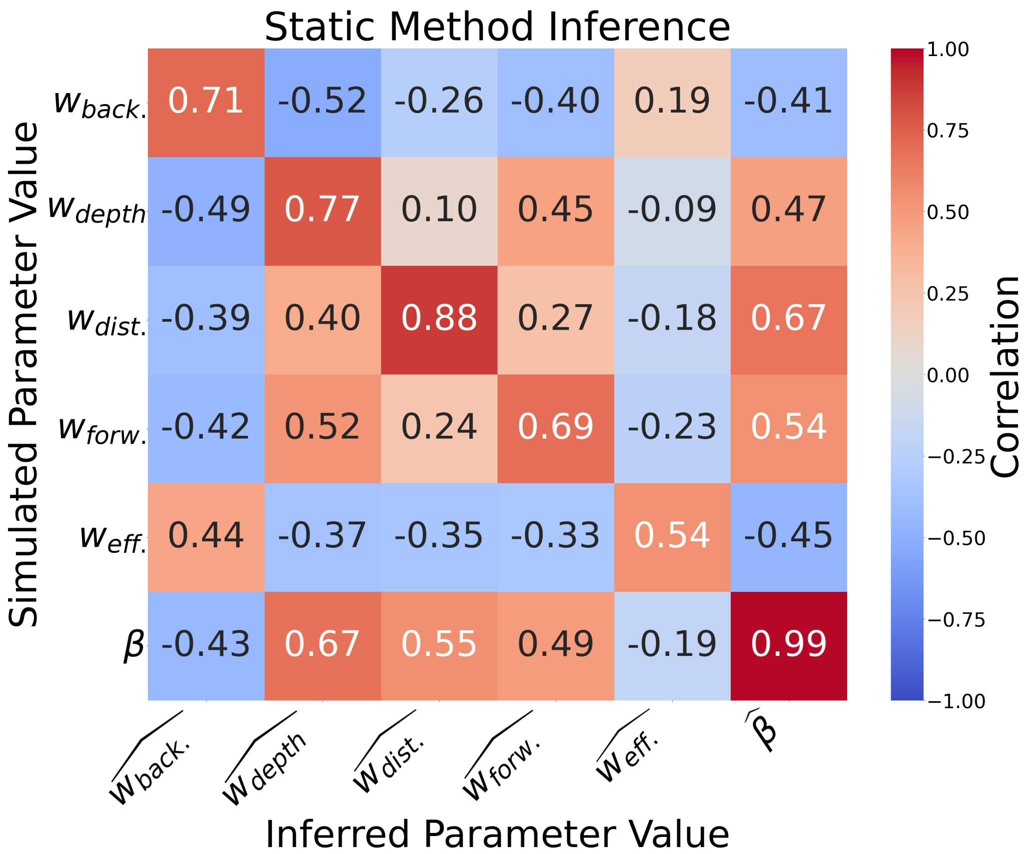


Figure 4.1: Spearman correlations between simulated and inferred parameter values.

4.3 The fixed effects method yields sufficient parameter recovery.

When developing cognitive models, establishing parameter recovery is an important validation step. Therefore, I wanted to establish that we can recover parameter values from data simulated from the best fitting model.

4.3.1 Methods

In order to make sure that the model parameters were in a realistic range, I used the inferred parameter values from Experiment 1. Specifically, I calculated $\hat{\vartheta}_{\mathrm{MLE}}$ for each participant in Experiment 1. Then, for each participant's inferred parameter values, I simulated 20 trials.

4.3.2 Results

4.3.2.1 The method yields good parameter recovery

I calculated the Spearman rank correlations for each model parameter (see Figure 4.1 for a graphical representation). I corrected the p-values with a Benjamini-Hochberg correction with a false discovery rate of 0.05 (Benjamini and Hochberg, 1995). I found that all correlations were significant, even with this correction. The lowest correlation was the effort cost (Spearman's $\rho(122) = 0.54$, $p < 0.001$, 95% C. I. $[0.41, 0.66]$). All other correlation coefficients were greater than 0.69 (non-backward search cost weight: Spearman's $\rho(122) = 0.71$, $p < 0.001$, 95% C. I. $[0.61, 0.79]$; depth cost weight: Spearman's $\rho(122) = 0.77$, $p < 0.001$, 95% C. I. $[0.69, 0.84]$; distance cost weight: Spearman's $\rho(122) = 0.88$, $p < 0.001$, 95% C. I. $[0.83, 0.92]$; non-forward search cost weight: Spearman's $\rho(122) = 0.69$, $p < 0.001$, 95% C. I. $[0.58, 0.77]$; temperature: Spearman's $\rho(122) = 0.99$, $p < 0.001$, 95% C. I. $[0.98, 0.99]$).

4.3.2.2 The highest posterior density intervals accurately capture uncertainty about the method's estimates.

Next, I calculated the 95% highest posterior density intervals for the simulated parameters (see Subsection 4.1.4). The typical spread of the 95% highest posterior density intervals is moderate for most cost parameters (≤ 5). Of note, the distance cost weight and depth cost weight's spreads were relatively low (≤ 3). However, the spread was relatively wide for the non-backward search cost weight parameter (i.e., 8, see Table A.10). The highest posterior density intervals for the temperature parameter also had a low spread ($M : 0.80, SD : 8.15$), although the temperature parameter is measured on a different scale.

Based on the width of the highest posterior density intervals for the cost weights, this points to some uncertainty between MLE values which are close to each other (i.e., 0 versus 1). However, it gives some credence to the method being able to distinguish between cost weight values on a coarser scale. This means the method may be able to distinguish between more aberrant parameter differences such as 0 versus 10.

The spreads of the highest posterior density intervals for all cost weights except the non-backward search cost weight were significantly correlated with the temperature parameter used to generate the data to a moderate degree (non-forward search cost weight:Spearman's $\rho(122) = 0.46$, adj.$p < 0.001$, 95% C. I. $[0.31, 0.59]$; planning depth cost weight: Spearman's $\rho(122) = 0.61$, adj.$p < 0.001$, 95% C. I. $[0.49, 0.71]$; distance cost weight: Spearman's $\rho(122) = 0.79$, adj.$p < 0.001$, 95% C. I. $[0.71, 0.85]$, effort cost weight: Spearman's $\rho(122) = 0.46$, adj.$p < 0.001$, 95% C. I. $[0.31, 0.59]$). The non-backward search cost weight was significantly negatively correlated (Spearman's $\rho(122) = -0.43$, adj.$p <$

0.001, 95% C. I. $[-0.56, -0.27]$). All p-values are with a Benjamini-Hochberg correction for multiple comparisons. This suggests that the cost weights are only coarsely recoverable and that certain cost weight settings are recovered more precisely when participants plan more systematically (i.e., lower decision temperature). This makes sense to a degree, because if a participant has a very high temperature, we expect them to choose possible actions uniformly regardless of the cost parameters. The non-backward search cost result may stem from participants who plan more optimally (higher non-backward search cost) also planning more systematically.

Next, I looked at how often the true parameter values were contained in the 95% highest posterior density intervals. I found the true parameter value was contained in the 95% highest posterior density interval 95 - 100% of the time for all model parameters. This implies that although the method's highest posterior density intervals tend to be a bit wide, they are still relatively well calibrated on simulated data. Nevertheless, users of the method should take the highest posterior density intervals with a grain of salt.

To verify the usefulness of these highest posterior density intervals to distinguish between participants with cost weights on a log scale (e.g., 0 versus 10), I looked at the correlation between the simulated participants' highest posterior density interval spread and the true (simulated) parameter values. I found a moderate correlation for all model parameters except the temperature parameter (see Table A.11.) This suggests that the precision of the planning cost weights' highest posterior density interval is affected by a participant's planning cost weight values, which further suggests that the method should only be able to differentiate between people whose cost weights differ to a significant extent. This might also be due to a ceiling effect for some of the higher parameter values – less (or no) clicking may arise from many different combinations of high parameter values.

I next investigated, for each simulated participant, how the absolute error of the method's estimate (the MLE estimate minus the parameter used to simulate the trajectory) varied with the spread of the HDI. I found that the absolute error and the spread of all estimates were moderately correlated except in the case of the temperature parameter (see Table A.12.) This means that the uncertainty communicated by the widths of the highest posterior density intervals outputted by the method is informative about the accuracy of the corresponding MLE estimate.

4.3.2.3 Conclusion

In conclusion, the outlined method is reasonably accurate at recovering the ground truth model parameters from participant data simulated according to the full cost model. Looking at the highest posterior density intervals for simulated participants, I found that the spread of the highest posterior density intervals was significantly correlated with the simulated value of the temperature parameter, which can be seen as a proxy for model fit. The highest posterior density intervals for the planning depth cost weight and distance cost weights had a smaller spread on average than those for the other cost parameters. This suggests the method is most certain of the values for these two parameters.

4.4 Can we differentiate between simulated data generated from different models?

In the BMS analyses in Subsubsection 3.2.3.2, I found that all participants' data were best explained not by the best model BIC-wise but by models with a subset of cost parameters. In this section, I use simulated data to probe whether we can differentiate when data comes from each of these submodels.

Investigating model recovery is an important validation while doing model comparison (Wilson and Collins, 2019). To do model recovery, one simulates datasets from a potential set of models, and investigates the probability that each model would best explain each of the simulated datasets. We are also interested in looking at reversing the condition and looking at which model is most likely to have generated the data, given the best fitting model.

4.4.1 Methods

Data was generated for the top ten models based on Bayesian Model Selection for Experiment 1 (see Table 3.1). I generated data only from MLE parameters for participants who were best explained by each of the models. Some models best explained only a handful of participants (i.e., 2 to 5). In order to keep the datasets similar in size, as well as similar to the size of Experiment 1, I generated multiple sets of these participants such that each simulated dataset ranged in size from 117 to 126 participants. For each participant, I generated 10 trials.

I also generated data for the top three models based on Bayesian Information Criterion for Experiment 1. Because none of the models had more than one expected participant based on Bayesian Model Selection, I used the MLE parameter for all participants. Therefore, all datasets had 122 simulated participants. For each participant, I generated 10 trials.

For each new dataset, I ran Bayesian Model Selection to get the expectation of the posterior $P(\text{model}|\text{data})$.

This value is used in the posterior matrices. For the inverted posterior matrices, using Bayes' rule, we have:

$$P(\text{simulated model}|\text{inferred model}) = \frac{P(\text{inferred model}|\text{simulated model})P(\text{simulated model})}{P(\text{inferred model})} \tag{4.2}$$

We assume $P(\text{simulated model})$ is uniform, therefore $P(\text{inferred model})$ is found by summing over the columns in the posterior probability matrix.

4.4.2 Results

A sign of non-identifiability is large off-diagonal values in the model recovery matrices. In the matrices shown in Figure 4.3, we see some non-identifiability between these submodels. For the top three models, we see that when the full cost model is inferred to be the best model, the data is almost uniformly likely to have come from any of the possible models (Figure 4.2b). This may be because these models are nested and the additional parameter does not significantly affect fit for most participants.

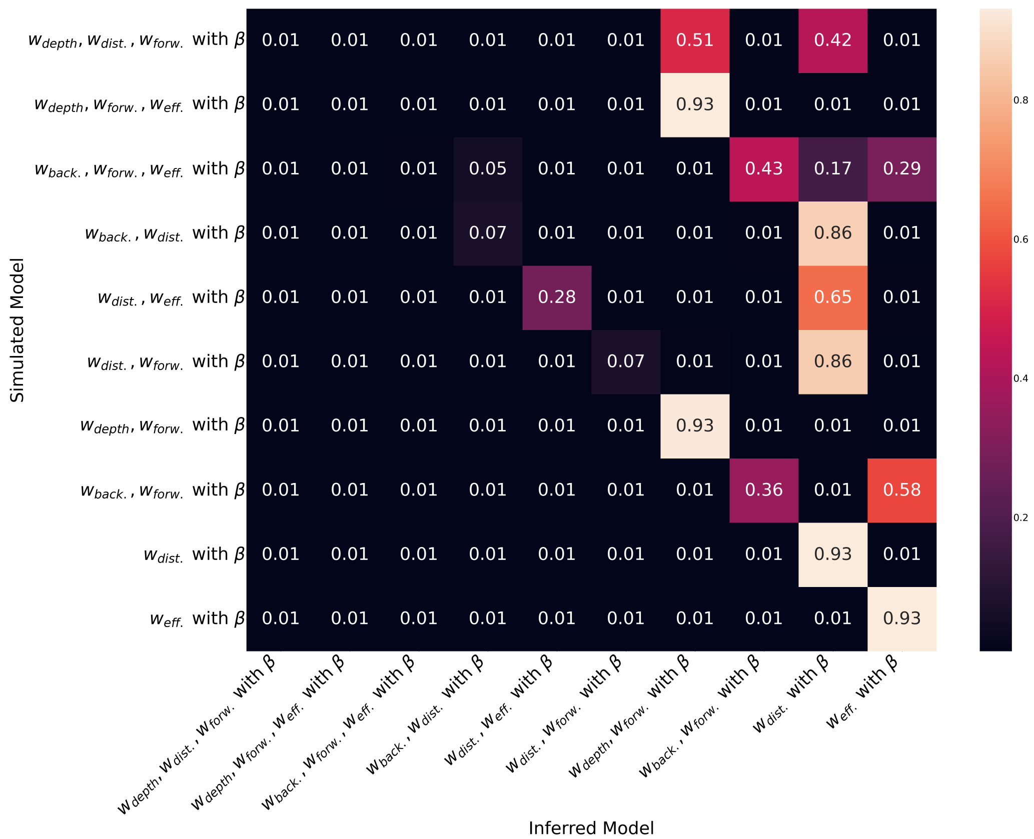
For the random effects analysis shown in Figure 4.2a, only the single cost parameter models, as well as one of the two cost parameter models have diagonal values > 0.80. At the same time, in Figure 4.2b, we see that only one of the two parameter models has a very high $P(\text{simulated model}|\text{inferred model})$. Therefore, we cannot easily differentiate between which simpler model generated the data.

4.4.3 Conclusion

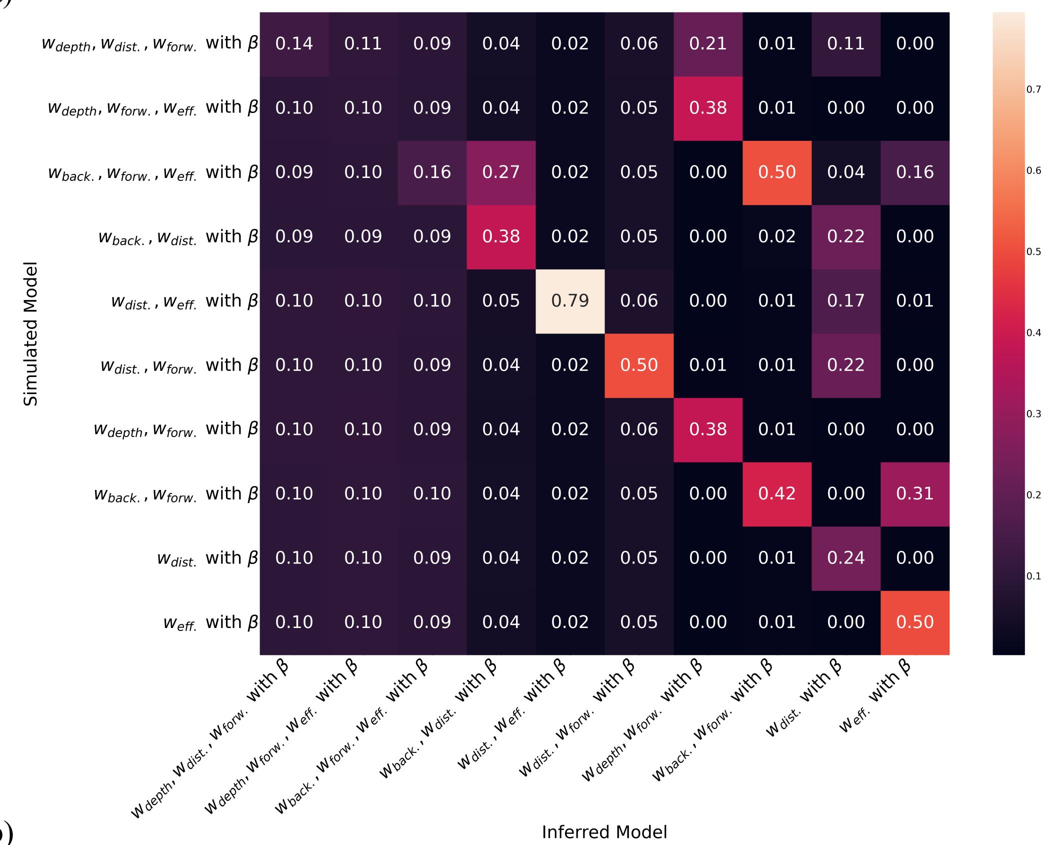
Considering the good parameter recovery as demonstrated in Subsection 4.3, and the fact that some of these models are nested, I will continue to use the full parameter model. However, this poor model recovery draws attention to the fact that the same behavior may be explainable by quite different models. This is a common issue with inverse reinforcement learning and I will go on to discuss this more generally in Chapter 6.

4.5 Experiment 2: Can the method recover induced cost weights in a validation experiment?

Now that I have validated the method on simulated data, I present the results of a validation experiment I devised to test the method's accuracy for human participants. I ran an experiment in which I imposed different levels of additional planning cost (specifically,



(a)



(b)

Figure 4.2: (a) The posterior model probabilities P(inferred model | simulated model), which show the probabilty of a given model being the best for the simulated datasets. (b) The inverted posterior model probabilities P(simulated model | inferred model), which show the probability of a simulated model having generated the data, given that it is the best model.

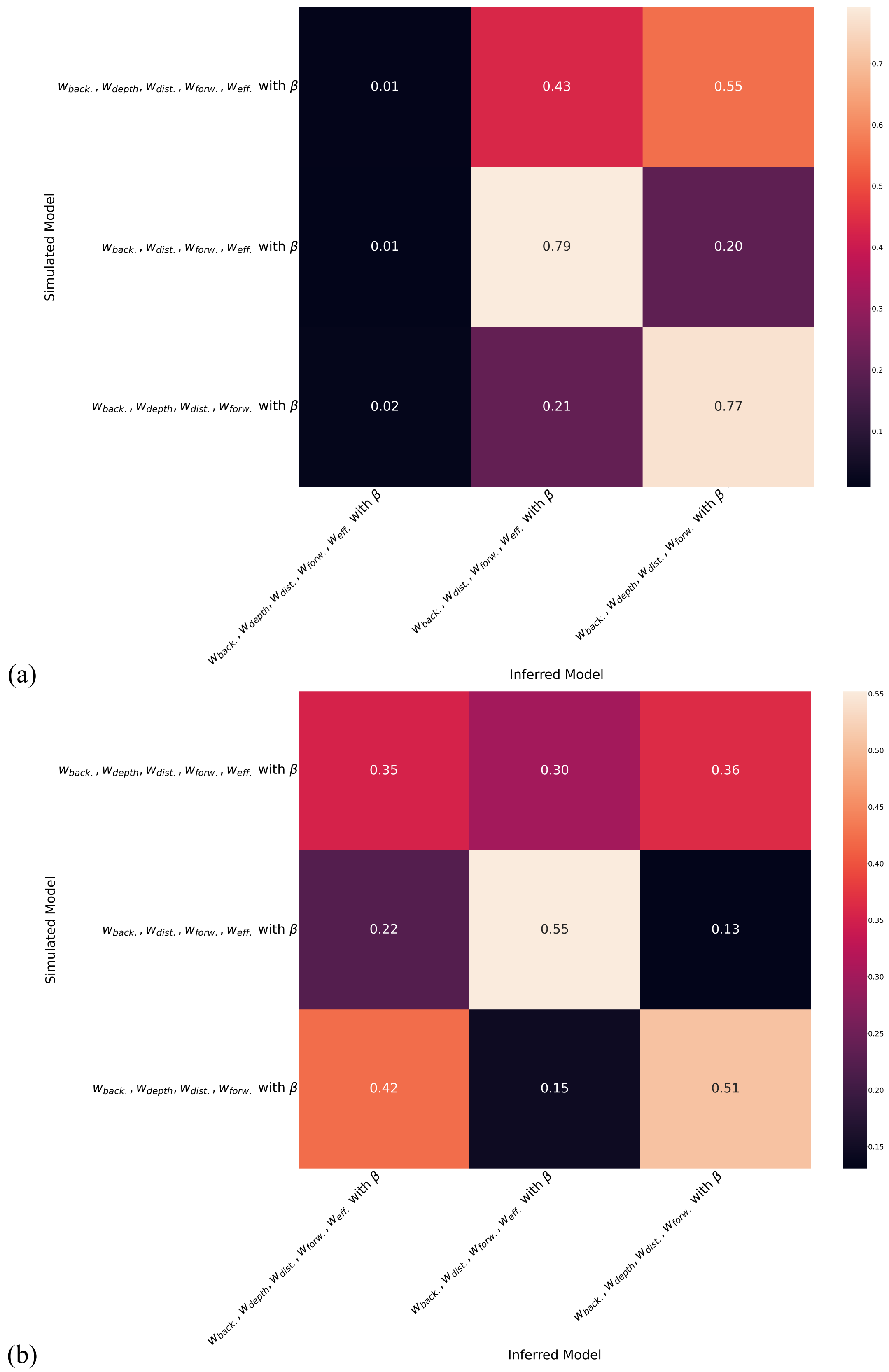


Figure 4.3: (a) The posterior model probabilities P(inferred model | simulated model), which show the probabilty of a given model being the best for the simulated datasets. (b) The inverted posterior model probabilities P(simulated model | inferred model), which show the probability of a simulated model having generated the data, given that it is the best model.

the depth cost weight and the effort cost weight due to ease of manipulation) on different participants to emulate inter-individual differences in the planning depth cost weight and the effort cost weight.

4.5.1 Methods

4.5.1.1 Participants

I recruited 420 participants on Amazon Mechanical Turk, using CloudResearch (Litman *et al.*, 2017). 347 participants completed the experiment; the other participants quit during the instructions phase. Of these 347 participants, 12 participants were unable to pass the instructions quiz. So, I have complete data for 335 participants (192 females, 141 males, 2 nonbinary; median age 40, age range 19-78). Participants completed the task in a median time of 25.65 minutes and earned a median \$4.20 bonus in addition to the \$0.50 base pay.

4.5.1.2 Materials

I created a modified version of the Mouselab-MDP task in which participants have to wait a variable amount of time after inspecting a node before they can inspect a new node or move the spider through the web. As a cover story, participants are told that the node inspector has limited power and that they need to wait for it to recharge in between uses. The amount the node inspector was drained by inspecting a node – and how long it consequently took for it to recharge – varied by condition and node depth. Participants were not told how long the node inspector would take to recharge after any particular node, nor were the nodes marked according to this duration. Instead, participants were only told that the time it takes the node inspector to recharge might vary between nodes. Because Experiment 1 had revealed that people differ in how costly planning is for them at baseline, I added a block where participants were told the node inspector did not need to recharge; the purpose of this block was to measure participants' baseline cost values.

Participants were incentivized with a financial bonus of 1 cent for every \$5 earned in this game. At the beginning of the task, each participant received an endowment. To keep the payment fair across conditions, the endowment participants received at the beginning of the task was chosen to equalize the average wage approximately across all 12 experimental conditions.

4.5.1.3 Experimental Design

I randomly assigned each participant to one of 12 conditions, which were combinations of the following imposed additional costs: $\Delta_{w_{depth}} \in (0.0, 2.0, 5.0, 10.0)$ and $\Delta_{w_{eff.}} \in (2.0, 5.0, 10.0)$.

The amount of time participants had to wait was based on the cost function combining these two parameters; that is participants had to wait $t = \Delta_{w_{depth}} \cdot d_n + \Delta_{w_{eff.}} \cdot 1$ seconds after inspecting a node n at depth d_n. For instance, a participant with $\Delta_{w_{depth}} = 5.0$ and $\Delta_{w_{eff.}} = 2.0$ would have to wait for 7 seconds after inspecting an immediate outcome ($d_n = 1$), 12 seconds after inspecting an outcome that is two steps away ($d_n = 2$), and 17 seconds for inspecting one of the final outcomes ($d_n = 3$).

4.5.1.4 Procedures

Following the experiment's instructions, participants were given a pre-task comprehension quiz that checked their understanding of the task and the possible performance bonus. Participants were again only allowed to move onto the task if they passed the pre-task comprehension quiz within four attempts.

Participants were then presented with the task. In all conditions, I gave participants a block of 20 trials of the modified Mouselab-MDP task with time delays, a block of 15 trials without delays (baseline trials), and a block of 15 trials with time delays. The order of the last two blocks of 15 trials was counterbalanced across participants.

4.5.2 Results

4.5.2.1 Manipulation checks: The manipulated planning depth cost weight but not the effort cost weight has an effect on planning behavior.

To check that the imposed additional costs influenced behavior in the way I expected, I looked at the ratio of late node clicks for the additional planning depth cost weight and the total number of clicks for the additional effort cost weight. I expected the ratio of late node clicks to decrease with higher planning depth cost weight values and the total number of clicks to decrease with higher effort cost weight values. I corrected the four (two per block) p-values with a Benjamini-Hochberg correction with a false discovery rate of 0.05.

The manipulation of the planning depth cost weight appeared to be working, the proportion of late nodes clicked was lower in the baseline block than in the test block ($W = 8935.00$, RBC $= -0.49$, $p < 0.001$, two-sided; baseline: $M = 0.746, SD = 0.352$; test: $M = 0.569, SD = 0.453$). As additional evidence, participants assigned to higher planning depth cost weight conditions clicked less on late level nodes in the test block (Spearman's $\rho(335) = -0.33$, adj.$p < 0.001$, 95% C. I. $[-0.42, -0.23]$, adj.$p < 0.001$). As expected, the planning depth cost weight intervention did not correlate with late clicks in the baseline block (Spearman's $\rho(335) = -0.07$, adj.$p = 0.230$, 95% C. I. $[-0.17, 0.04]$, adj.$p = 0.230$). This implies that the depth cost weight manipulation worked, behaviorally.

The manipulation of the effort cost weight only partially worked: the number of clicks was significantly lower in the test block than in the baseline block ($W = 2867.50$, RBC $= -0.88$, $p < 0.001$, two-sided; baseline: $M = 3.621, SD = 2.272$; test: $M = 1.660, SD = 1.548$), However, there was no significant correlation between the assigned effort cost weight values and the number of clicks in the test block (Spearman's $\rho(335) = -0.07$, $p = 0.198$, 95% C. I. $[-0.18, 0.04]$, adj.$p = 0.230$). This implies that while this manipulation increased clicking, clicking did not seem to increase in a way that correlated with the manipulation amount.

4.5.2.2 The planning depth cost weight but not the effort cost weight manipulation is reflected in block-wise MLE estimates.

To validate the method, I compared the MLE estimates between the test block that imposed an additional cost of planning and the baseline block that did not. I expected to see higher MLE estimates for the planning depth cost weight in the test block compared to the baseline block. For the baseline block, I expected the MLE estimates to be similar to the MLE estimates in Experiment 1. Because the manipulation of the effort cost weight appeared to work in a way different than intended, I expected little difference between inferred values in the test and baseline blocks.

As expected, the MLE estimates of the planning depth cost weight were higher for the test block ($\mathrm{M} = 1.61,\ \mathrm{SD} = 2.78$) than those for the baseline block ($\mathrm{M} = 1.07, \mathrm{SD} = 2.79$; Wilcoxon signed-rank test: $W = 2576.00$, RBC[1] $= 0.40$, adj.$p < 0.001$, two-sided). The MLE estimates of the planning depth cost weight in the test block were also significantly higher than those in Experiment 1 ($\mathrm{M} = 1.16,\ \mathrm{SD} = 2.72$; Mann-Whitney $U = 23060.50$, RBC $= -0.13$, adj.$p = 0.028$, two-sided). As predicted, the MLE estimates of the planning depth cost weight were not significantly different between the baseline block of Experiment 2 and Experiment 1 (Mann-Whitney $U = 18967.00$, RBC $= 0.07$, adj.$p = 0.152$, two-sided). These results are consistent with the hypothesis that the manipulation worked and that my method can recover the imposed cost of planning because the MLE estimates were larger for the test block that imposed an additional cost of planning than for the baseline block of Experiment 1 which did not.

For the effort cost weight, the MLE estimates were also significantly higher for the test block ($\mathrm{M} = 4.01,\ \mathrm{SD} = 3.90$) than for baseline block ($\mathrm{M} = 1.78,\ \mathrm{SD} = 2.74$; Wilcoxon signed-rank test $W = 5032.50$, RBC $= 0.64$, adj.$p < 0.001$, two-sided) even though I was unable to reliably manipulate the effort cost weight. Comparing the MLE estimates for the effort cost weight in Experiment 1 ($\mathrm{M} = 1.51,\ \mathrm{SD} = 2.42$) against the MLE esti-

[1] The rank biserial correlation (RBC) is one way of quantifying the effect size of a Wilcoxon signed-rank test. It is the proportion of data in favor of the hypothesis versus not (Vallat, 2018; Kerby, 2014).

mates in Experiment 2 for the test block, I found a significant difference (Mann-Whitney $U = 28941.50$, RBC $= -0.42$, adj.$p < 0.001$, two-sided). As expected, I did not find a significant difference between the MLE estimates for the effort cost weight in Experiment 1 and the MLE estimates in the baseline block of Experiment 2 (Mann-Whitney $U = 21437.00$, RBC $= -0.05$, adj.$p = 0.506$, two-sided). The difference found between the test block versus the baseline block, and between the test block and Experiment 1 suggests that the method was able to recover the manipulation of the effort cost weight. These results suggest that the effort cost weight may be more reliably recovered and may have, in fact, been manipulated as expected.

Other model parameters appeared to vary between baseline and test blocks. Temperature varied as one would expect (see Subsection A.4.1). The distance cost weight was also higher in the test block than the baseline block (Wilcoxon signed-rank test: $W = 8896.00$, RBC $= 0.20$, adj.$p = 0.024$, two-sided). There was no significant difference between test and baseline blocks for the inferred non-forward search cost weight (Wilcoxon signed-rank test: $W = 8292.00$, RBC $= -0.12$, adj.$p = 0.200$, two-sided). Interestingly, non-backward search cost weight was inferred to be lower for the baseline block than the test block (Wilcoxon signed-rank test: $W = 3729.00$, RBC $= -0.27$, adj.$p = 0.015$, two-sided). This may be due to the removal of high imposed cost weights – suddenly a backward search strategy is not as important as before.

I wanted to see whether I could recover the assigned cost weights with the inference method. However, unlike when I assessed parameter recovery on simulated data, I was missing information on the ground truth parameter values. Therefore, I performed one linear regression per manipulated variable, investigating whether the inferred parameter value in both blocks could explain the variance of each assigned cost weight. I included baseline block presentation as an addition covariate, although it does not seem that this variable affected the MLE parameter values too much (see Appendix Subsection A.4.2.)

For the assigned planning depth cost weight as the dependent variable, the overall regression was statistically significant but with a low R^2 (adj. $R^2 = 0.02$, $F(3,331) = 3.81$, $p = 0.010$). For the effort cost weight as the response, the overall regression was also not statistically significant (adj. $R^2 = -0.01$, $F(3,331) = 0.18$, $p = 0.909$). This implies that we cannot explain a good portion of the variance in either manipulated costs from the inferred parameters values in the test and baseline blocks.

Given the covariance between different parameters with the simulated datasets (see Figure A.7) and the change in MLEs for other model parameters, I decided to add more variables to the independent variables in the linear regression. I performed the same linear regression as above, investigating whether the inferred parameter values (all five cost weight values and the temperature parameter) in both blocks could explain the variance of each assigned cost weight. Again, I included the order of the baseline block presentation as an additional covariate.

Now, for the depth cost weight as the dependent variable, the overall regression was both significant and had a higher (although still rather low) R^2 (adj. $R^2 = 0.13$, $F(13,321) = 4.67$, $p < 0.001$). This suggests, to some degree, that a fair proportion of the variance in the planning depth cost weight can be explained by the inferred parameter values in the test and baseline blocks. However, variance in the effort cost weight was still not explained by this expanded set of variables (adj. $R^2 = -0.02$, $F(13,321) = 0.53$, $p = 0.908$). This might be because, as discussed above, the manipulation of the effort cost weight was only partially working.

The question remains: can the method infer the planning cost weights reliably enough to use them as an individual-differences metric? To answer this question, I examined the root-mean-square error (RMSE) of the linear regression models. I calculated the RMSE between the prediction of the linear regression and the true cost imposed on the participant. To estimate the RMSE for predicting new observations, I performed 10-fold cross-validation, stratified across participant conditions (assigned cost and order of baseline block presentation). The average RMSEs were 3.97 (SD: 1.35) for the planning depth cost weight and 3.54 (SD: 0.58) for the effort cost weight. Since the given parameters ranged from 0 to 10, this means that the method should at least be able to differentiate individual participants' values on a coarse scale, such as "low" versus "high" (it should be able to tell the difference between values $2 \cdot \text{RMSE}$ apart).

I then used the computed RMSEs to test if the amount of variance explained by the model is significantly greater than zero. Concretely, I performed a one-sample t-test to see if participants' squared error (between regression output and instructed cost) was significantly less than the variance of the instructed cost weight. For the planning depth cost weight, I found that the squared error was significantly less than the variance ($t(334) = -3.30$, $p < 0.001$). This implies that the regression model does explain a significant proportion of the variance of the true planning depth cost weight. However, for the effort cost weight, I found that participants' squared error was not significantly different from the variance of the true cost weight ($t(334) = -0.59$, $p = 0.277$). This makes sense given that the number of clicks was not significantly correlated with the imposed effort cost weight.

4.5.2.3 The method can be used to classify individuals as exhibiting a high versus low cost of planning.

Having found that the method is appropriate for measuring the cost of planning at a coarse scale, I then evaluated how accurately it can classify people as behaving according to a high versus low cost of planning. Concretely, I performed median splits on the imposed planning costs[2] and assessed whether the method could predict whether the imposed cost was above versus below the median. The classification was performed by applying a decision stump (one-level decision trees) to the prediction of the linear regression model described above. I used a 5-fold cross-validation procedure in which the decision threshold was optimized on one set of the data and the performance of the resulting classifier was evaluated on the held-out set. The performance of the classifier was measured by the balanced accuracy score (Brodersen *et al.*, 2010).

For the planning depth cost weight, cross-validation had an average balanced accuracy score of 70%. The values the data was split on a narrow range of values from 3.29 to 4.00, depending on the fold. This suggests there is a bit of an offset in the linear regression model output, although it should correct for the participants' baseline parameter values. For the effort cost weight, the cross-validation average balanced accuracy was only 51%. This suggests that we cannot reliably differentiate between coarse values of the effort cost weight.

These results imply that the method can somewhat accurately differentiate between planning depth cost weight values on a coarser scale, but cannot differentiate between values of the effort cost weight. Since this analysis depends on both the linear regression model to correct for baseline cost values and the success of the cost manipulations, the accuracy reported here could be interpreted as a lower bound on the performance of the method.

4.5.2.4 Highest posterior density interval spreads signify uncertainty of method in test block.

I examined the highest posterior density intervals outputted by my method (see details of calculation in Subsection 4.1.4) to investigate the reliability of my method's uncertainty estimates. To do so, I looked at the spread (width) of the highest posterior density intervals which should quantify the confidence of the method in providing a parameter estimate. I also measured the accuracy of the method by looking at how often the values outputted by my linear regression model occur in the highest posterior density intervals.

The spread of the highest posterior density intervals in the test block were higher than in the baseline block for the all planning cost weights (see Table A.9). The higher average temperature parameter in the test block indicates increased variability. This fits the narrative that participants have to adapt to new, external costs, and may react to these external costs in varied ways.

[2] The median was 5.0 for the planning depth cost weight and 5.0 for the effort cost weight.

When measured only in the test block, higher MLE estimates for the temperature parameter were correlated with a lower spread size of the highest posterior density intervals for the non-backward search cost weight only (see Table A.14). This suggests that my method is more confident for people with higher non-backward search cost weight values. I also found that higher MLE estimates for the distance cost weights are strongly correlated with HDI spreads (see Table A.13). This means that the method is less confident for higher values of the distance cost weight.

4.5.2.5 Highest posterior density intervals for the depth cost weight accurately communicate uncertainty in inference.

Next I examined the correlation between the highest posterior density interval spread in the test block and the absolute error between the parameter estimates using the output of the linear regression model and the assigned costs for each participant. In order for the highest posterior density intervals to be useful, the method should report more narrow highest posterior density interval spreads when it can accurately estimate parameter values. There was a significant correlation between the spreads of the 95% highest posterior density intervals and the absolute error per participant for the depth cost weight (Spearman's $\rho(335) = 0.41$, adj.$p < 0.001$, 95% $C.\,I.\,[0.31, 0.49]$) but not the effort cost weight (Spearman's $\rho(335) = 0.06$, adj.$p = 0.255$, 95% $C.\,I.\,[-0.05, 0.17]$).

This implies that when my method cannot estimate the depth cost weight well it communicates that by returning a wider range of possible values. However, the HDIs provided by the method were unable to indicate uncertainty in the effort cost weight estimate. This may also have to do with issues in the manipulation or linear regression model, rather than the method.

4.5.3 Conclusion

This validation experiment tested how accurate and how precise the method is at measuring individual differences in the cost of human planning, particularly the planning depth cost weight. The results show that the method's MLE estimates of the planning depth cost weight are reasonably accurate, even though the accuracy of the corresponding highest posterior density intervals remains unclear. Whether the method can accurately measure the effort cost weight remains unclear because the effort cost weight manipulation did not have the desired effect on the amount of planning. This may have to do with the effort cost weight having the lowest parameter recovery out of all the model parameters. Relevant to potential clinical applications, I found it is possible to detect whether the planning depth cost weight was below or above the population median with a balanced accuracy of 70% based on the MLE parameter estimates. The ability to accurately tell whether an individual participant's planning depth cost weight is larger than the standard value of the planning depth cost weight in the task may help researchers identify individuals with aberrant planning.

4.6 Discussion

In this chapter, I presented a method which uses the model set forth in Chapter 3 to measure individual differences in human planning. I found that parameter recovery is relatively good for a simulated dataset with the MLE parameters found for participants in Experiment 1. I ran a validation experiment where participants' planning depth and effort costs were increased. I found that the effort cost weight manipulation did not work as planned, but that the added planning depth cost weight was recoverable on a coarse scale (median split) with a balanced accuracy score of 70%.

Previous task-based work has focused on behavioral tasks such as the Plan-a-Day task, the towers of London or Hanoi (Simon, 1975; Shallice *et al.*, 1982) or the two-step task (Daw *et al.*, 2011). Most models of these tasks provide insight into whether or how far a person is planning. In contrast, this method provides a mechanistic explanation of planning which encompasses several different planning factors. This model also has some similarities to the Computational Microscope (Jain *et al.*, 2022) which is also designed for the Mouselab-MDP task. This method uses Hidden Markov Models to infer a participant's strategy per trial, whereas my model is more parsimonious: only five parameters are inferred. In contrast, the Computational Microscope classifies each trial as having arisen from one of 79 defined strategies. This means that where the Computational Microscope might infer someone to follow one of several forward search strategies, I would expect my model to just output a high non-forward search cost weight.

One drawback of my Bayesian inverse reinforcement learning algorithm is that it assumes that a person's planning strategy never changes. Despite this limitation, I found that my method allows one to draw broad inferences about individual differences in the subjective cost of planning. A possible extension would be to extend my approach to settings where people's planning strategies change over time; this could be accomplished by using Bayesian inverse reinforcement learning for learning agents (Kubala *et al.*, 2019). This work could draw on recent models of how people learn planning strategies (Jain *et al.*, 2019; He *et al.*, 2021). One drawback of such a method would be the computational cost required for performing inference over the potentially large number of parameters for the model of how people learn planning strategies.

Another limitation is that we do not know how constant the inferred cost weights are.

Are they characteristic of an individual, remaining stable over time? How much are they influenced by environmental context? Would a person with higher depth cost weight in the high increasing variance environment also have a higher depth cost weight in the constant variance environment? Future work should be done to see whether and by how much the cost weights vary over time within people, and whether and how individual differences in the parameters are influenced by the experiment structure.

Despite this, the method is nevertheless a first step for inferring individual differences in cognitive costs in a process-tracing paradigm. In the next chapter, I will explore whether inferred cost weights correlate with real-world planning and psychiatric symptomatology.

Chapter 5

Exploring the relation between inferred planning cost weights and self-report measures

Preliminary evidence exists for a connection between planning ability and increased well-being (MacLeod and Conway, 2005; MacLeod *et al.*, 2008), as well as a connection between psychiatric symptomatology and decreased well-being (Grant *et al.*, 2013). Planning deficits are also a commonality across multiple psychiatric disorders (East-Richard *et al.*, 2020). However, much is still unclear about the relation between planning ability and psychiatric symptomatology, especially in real-world situations and across diagnostic groupings.

In this chapter, I investigate whether there is a relation between planning behavior in the Mouselab-MDP task and psychiatric symptomatology. To do so, I adopt an empirical approach, pairing the measurement of planning costs as presented in Chapter 4 with a battery of self-report measures. These self-report measures assess various personal characteristics, including psychiatric symptomatology and dimensions of real-world planning (such as self-reported consideration of future consequences and planning behavior in the financial and time-management domains.)

Instead of focusing on recruiting participants with one disorder, I take a subclinical transdiagnostic approach similar to that in Gillan *et al.* (2016). Using such an approach, I do not specifically recruit participants diagnosed with a one disorder. Instead, I recruit participants from the general population on an online platform. I then investigate how planning differs across (normally) subclinical variations in psychiatric symptoms.

While this approach overlooks extremes in disorders, it yields two important advantages. First, it is not focused on one disorder in particular and can look at more gen-

The methods section of this chapter is based on my preregistration found here: https://osf.io/tnhpv.

eral transdiagnostic categories or symptoms common to multiple disorders. For example, anhedonia is defined as "the inability to enjoy experiences or activities that normally would be pleasurable" in the American Psychological Association Dictionary of Psychology (VandenBos, 2007). While it is a central component of Major Depressive Disorder, anhedonia is also often present in schizophrenia and a multitude of neurological disorders. Taking a transdiagnostic approach may lead to a better understanding of commonalities between disorders in differences in planning.

The second advantage of taking such an approach is that it allows us to recruit from online platforms rather than traditional psychiatric study pools, since there is no need to recruit participants who meet clinical diagnostic criteria for a specific disease (Gillan and Daw, 2016). This approach has been applied successfully to study transdiagnostic differences in many other behavioral paradigms in recent years (e.g., Rouault *et al.*, 2018; Patzelt *et al.*, 2019; Wise and Dolan, 2020.)

The present work also acts as validation for the process-tracing task as a measure of real-world planning. This process-tracing task may act as a valuable tool for ascertaining planning behavior for those who may not be good at self-reporting their own planning styles. However, up until now, it is not known if behavior in the task is correlated with real-world planning behavior.

One potential drawback that other potential planning tasks may face is the "reliability paradox". Tasks have often been selected for their reliable group-level effects, which mean lower between-participant variance (e.g., people do not behave that differently from each other). However, this lower between-participant variance in fact leads to lower test-retest reliability of individual differences (Hedge *et al.*, 2018). This may stem from a mismatch between the goals of experimental psychology and individual differences research (Karvelis *et al.*, 2023; Zorowitz and Niv, 2023). Enkavi *et al.* (2019) probed test-retest reliability for self-regulation measures by pairing a large battery of tasks with self-report surveys. They found that surveys have better test-retest reliability than tasks, on average. However, the self-regulation tasks used were tasks developed for experimental psychology and not individual differences research. Since behavior in the Mouselab-MDP often varies a lot per participant (Jain *et al.*, 2022), the Mouselab-MDP task may be a better candidate task for individual differences research.

The work in this chapter lays the groundwork for developing a model-based understanding of planning deficits and the personalized selection of interventions that help people overcome those deficits. First, I will give an overview of the large experiment I ran (Experiment 3). Then I will report on the results of some preregistered hypotheses established before running the experiment. Finally, I will report on further exploratory analyses I performed on the data.

5.1 Methods

The methods reported here were preregistered in Felso and Lieder (2023). The hypotheses were organized in two levels. The first level of hypotheses explored whether we can explain the variance in self-report questionnaires with people's inferred cost parameter weights (above the explained variance by demographic variables and IQ.) The second level, if the data supported the previous level hypothesis, examined whether certain cost parameter weights (particularly the depth cost weight) contributed significantly to explaining the variance in the self-report measures. The hypotheses were grouped in seven categories according to the dependent variables in question: psychiatric measures, planning and future thinking measures, life regrets and life satisfaction, impulsivity and risk-taking measures, intolerance of uncertainty measure, distress tolerance, and cognitive styles (rationality). I will go through each set of categories here in the methods section. For a list of hypotheses, see Subsection A.6.1 or Felso and Lieder (2023).

5.1.1 Participants

Eight hundred seventy participants were recruited using Prolific (www.prolific.co; Palan and Schitter, 2018). Of those participants, 123 were excluded due to not passing the pre-session quiz in the maximal number of tries (four) or otherwise leaving the experiment. Of the remaining 747 participants, 37 were excluded for straightlining[1] on at least one page of questionnaires (where there were at least five questions and one question was reverse-coded.) The final sample of 710 participants consisted of 279 self-reported women, 415 self-reported men and 16 people who were non-binary or did not provide their gender. The median age of participants was 29 (range 19 to 78). Age data were not provided by 8 participants. All participants gave informed consent. I recruited participants until the pre-registered financial stopping rule of 9750 GBP was met.

5.1.2 Materials

Participants were given 40 trials of the Mouselab-MDP task with the high increasing variance settings, as described in Experiment 1 in Chapter 3. As in Experiment 1, participants were given 7 seconds per trial and were told that they would earn a 0.002 GBP bonus per game point. Participants completed the task in a median time of 64.78 minutes, and earned a median 3.18 GBP bonus in addition to the 7.00 GBP base pay.

[1]Providing the same response for all questions on a page, as if going down in a line without reading the questions.

5.1.2.1 Self-report measures

Participants were given a battery of self-report measures to complete as part of the experiment. These self-report measures are sorted into psychiatric measures, planning and future thinking measures, life regret and life satisfaction measures, impulsivity and risk-taking measures, and two measures of personality traits that are thought to lead to psychiatric dysfunction and which have been linked to less mental effort avoidance (Patzelt *et al.*, 2019).

§1 Psychiatric measures The psychiatric group of questionnaires comprised of 85 out of the 395 items presented to subjects (including the Cognitive Reflection Task and an intelligence measure). Participants were given the battery of psychiatric questionnaires introduced by Wise and Dolan (2020), which are a subset of the questionnaire battery first introduced in (Gillan *et al.*, 2016) (reducing 209 questions to 85). This reduced subset allows one to measure the three transdiagnostic factors ("anxious-depression", "compulsive behavior and intrusive thought", and "social withdrawal") from (Gillan *et al.*, 2016) in a more efficient way (with 76 items versus 233 items). I used the weights provided by Wise and Dolan (2020) to calculate the three transdiagnostic factors/ I included all items on the trait anxiety scale in order to test specific hypotheses related to anxiety.

§1.1 Alcohol Use Disorders Identification Test Only one item was included from the Alcohol Use Disorders Identification Test (AUDIT; Saunders *et al.*, 1993), which was "How often do you have a drink containing alcohol?". The possible responses ranged on a five-point scale from "Never" to "Four or more times a week".

§1.2 Apathy Evaluation Scale Four items from the Apathy Evaluation Scale (AES; Marin *et al.*, 1991) were included. Apathy is a lack of motivation or drive, one example item would be the question "I have motivation". Participants had to endorse (or not endorse) questions on a four-point Likert scale. Cronbach's alpha for the scale was 0.847 (95% C. I. $[0.828, 0.865]$).

§1.3 Barratt Impulsiveness Scale Eleven items of the Barratt Impulsiveness Scale (BIS-10; Patton *et al.*, 1995) were included. This scale measures how impulsive people act in their day-to-day life. One example item is "I act 'on impulse' ". Participants had to endorse (or not endorse) questions on a four-point Likert scale. Cronbach's alpha for the scale was 0.723 (95% C. I. $[0.692, 0.753)$.

§1.4 Eating Attitudes Test Four items from the Eating Attitudes Test (EAT-26; Garner *et al.*, 1982) were included. This scale asks participants about their eating habits. Participants must respond on a scale from "Always" to "Never". A sample item is "I am terrified about being overweight". Cronbach's alpha for the scale was 0.879 (95% C. I. $[0.863, 0.893]$).

§1.5 Liebowitz Social Anxiety Scale Twenty-six items from the Liebowitz Social Anxiety Scale (LSAS; Liebowitz, 1987) were included. Each situation was presented twice, once as a question of how fearful a participant would feel (4-point scale from "None" to "Severe") and next the degree of avoidance (4-point scale from "Never (0%)" to "Usually (67 - 100%)"). A sample item is "Giving a prepared oral talk to a group". Cronbach's alpha for the scale was 0.945 (95% C. I. $[0.939, 0.951]$).

§1.6 Obsessive-Compulsive Inventory-Revised Eleven items from the Obsessive-Compulsive Inventory-Revised (OCI-R; Foa *et al.*, 2002) were included. This scale measures people's obsessive-compulsive behaviors and/or attitudes. Participants endorse items on a 4-point Likert scale ranging from "Rarely/Never" to "Almost always/Always". A sample item is "I check things more often than necessary". Cronbach's alpha for the scale was 0.870 (95% C. I. $[0.856, 0.884]$).

§1.7 Self-Rating Depression Scale Eight items from the Self-Rating Depression Scale (SDS; Zung, 1965) were included. This scale screens for symptoms of depression. Participants must respond on a 4-point Likert scale from "A little of the time" to "Most of the time". A sample item (reverse-coded) item is "I feel that I am useful and needed". Cronbach's alpha for the scale was 0.864 (95% C. I. $[0.848, 0.879]$).

§1.8 Trait portion of the State-Trait Anxiety Inventory (STAI) Participants responded to the full trait portion of the State-Trait Anxiety Inventory (STAI; Spielberger *et al.*, 1983), which consists of twenty items. This scale measures the degree which participants have trait anxiety. Trait anxiety can be seen as part of their personality, as opposed to "state" anxiety which would be anxiety a participant feels due to the current situation (Schmitt and Blum, 2020). Participants must respond on a 4-point Likert scale from "Almost never" to "Almost always". A sample (reverse-coded) item would be "I am 'calm, cool, and collected'." Cronbach's alpha for the scale was 0.864 (95% C. I. $[0.937, 0.949]$).

§2 Planning and future thinking measures The planning and future thinking group of questionnaires comprised of 51 out of the 365 items presented to subjects. Participants were given three self-report questionnaires that focus on their planning behavior and future-thinking.

§2.1 Consideration of Future Consequences Propensity to consider future consequences was measured by the Consideration of Future Consequences scale (CFC; Strathman *et al.*, 1994). The scale consists of twelve items which are rated on a Likert scale from “Extremely uncharacteristic” to “Extremely characteristic”. A sample item is “Often I engage in a particular behavior in order to achieve outcomes that may not result for many years”. Cronbach’s alpha for the scale was 0.837 (95% C. I. $[0.818, 0.854]$).

§2.2 Future Orientation Scale Future orientation as measured by the Future Orientation Scale (FOS; Steinberg *et al.*, 2009). The scale consists of fifteen items, which consist of two (opposing) statements each. Participants must identify which statement most applies to them, and if this statement only applies slightly or strongly. A sample item is “Some people make decisions and then act without making a plan BUT Other people usually make plans before going ahead with their decisions”. Cronbach’s alpha for the scale was 0.825 (95% C. I. $[0.805, 0.843]$).

§2.3 Propensity to Plan The Propensity to Plan scale (PTP; Lynch *et al.*, 2010) measures how much people plan on two timescales (days versus weeks) and two planning domains (financial versus time.) Each timescale and domain subscale consists of six items, for a total of twenty-four items. Participants must endorse the statement on a seven-point Likert scale from “Strongly disagree” to “Strongly agree”. A sample item from the short-term time-domain subset is “I consult my planner to see how much time I have left for the next few days.” Cronbach’s alpha for the scale was 0.961 (95% C. I. $[0.956, 0.965]$).

§3 Life regret and life satisfaction measures The life regret and life satisfaction group of questionnaires comprised of 29 out of the 365 items presented to subjects.

§3.1 Brief Inventory of Thriving Thriving was measured by the Brief Inventory of Thriving (BIT; Su *et al.*, 2014) The scale consists of ten items which measure psychological well-being (e.g., positive feelings towards life such as feelings of purpose, belonging.) Participants must endorse the statement on a five-point Likert scale from “Strongly disagree” to “Strongly agree”. A sample item is “What I do in life is valuable and worthwhile”. Cronbach’s alpha for the scale was 0.920 (95% C. I. $[0.911, 0.929]$).

§3.2 Life Regrets Scale Regrets in life as measured by the Life Regrets Scale (Pethtel and Chen, 2014). The scale consists of nine items which ask about regrets in different areas of life (e.g., finances, relationships.) Participants respond using a five-point Likert scale which ranges from "No regret" to "Very strongly regret". A sample item is "Looking back on your life, how much do you regret the way you pursued your LEISURE? (e.g., not traveling enough, not pursuing a hobby, etc.)". Cronbach's alpha for the scale was 0.848 (95% C. I. $[0.831, 0.864]$).

§3.3 Regrets Scale Propensity to regret was measured by the Regrets Scale (Schwartz *et al.*, 2002). Five questions are asked about how often a participant feels regrets related to their decisions in life. Participants must endorse the statement on a seven-point Likert scale from "Strongly disagree" to "Strongly agree". A sample item is "If I make a choice and it turns out well, I still feel like something of a failure if I find out that another choice would have turned out better." Cronbach's alpha for the scale was 0.764 (95% C. I. $[0.735, 0.790]$).

§3.4 Satisfaction with Life Scale Satisfaction with life was measured by the Satisfaction with Life Scale (SWLS; Diener *et al.*, 1985). Five questions gauge how satisfied a person is with their life. Participants must endorse the statement on a seven-point Likert scale from "Strongly disagree" to "Strongly agree". A sample question is "I am satisfied with my life". Cronbach's alpha for the scale was 0.912 (95% C. I. $[0.901, 0.922]$).

§4 Impulsivity and risk-taking measures The impulsivity and risk-taking group of questionnaires comprised of 162 out of the 365 items presented to subjects.

§4.1 UPPS-P Impulsive Behavior Scale Impulsivity was measured by the five-factor "Lack of premeditation as measured by the Urgency, Premeditation (lack of), Perseverance (lack of), Sensation Seeking, Positive Urgency, Impulsive Behavior" scale (UPPS-P; Whiteside *et al.*, 2005; Lynam *et al.*, 2006). The scale consists of 59 items which are 4-point Likert scale ranging from "Agree Strongly" to "Disagree Strongly". A sample question in the lack of premeditation subscale is "I have a reserved and cautious attitude toward life". Cronbach's alpha for the scale was 0.934 (95% C. I. $[0.926, 0.94)$.

§4.2 Domain-Specific Risk-Taking Scale Risk-taking behavior, perceptions and expected benefits was measured by the Domain-Specific Risk-Taking (DOSPERT; Blais and Weber, 2006). Thirty activities were presented, in three different phases. The first time an item was presented with the questions, participants were asked how likely they would engage in a particular activity. Participants rated the likelihood on a 7-point Likert scale from "Extremely Unlikely" to "Extremely likely". Next, participants were asked to rate how risky they found each activity on a 7-point Likert scale from "Not at all risky" to "Extremely risky". Finally, participants had to rate the benefits each activity could provide them using a 7-point Likert scale ranging from "No benefits at all" to "Great benefits". An example item is "Moving to a city far away from your extended family". Cronbach's alpha for the likelihood was 0.872 (95% C. I. $[0.858, 0.886]$), 0.879 for the perceived risk (95% C. I. $[0.8660.892]$) and 0.897 for the expected benefit (95% C. I. $[0.8860.908]$)

In addition, I measured the (ir)rationality of the ratings provided. This was a novel measure proposed in the preregistration for this experiment (Felso and Lieder, 2023). For each situation, the irrationality index defaults to 0. For risks that people say they are "Somewhat likely", "Moderately likely" or "Extremely likely" to engage in, the "irrationality" of the risk is calculated as $\max(\text{risk perception} - \text{risk benefit}, 7 - \text{risk benefit})$. This means that someone who sees a lot of benefit but no risk in engaging in a risk would be assigned a low irrationality score for engaging in the situation. At the same time, if someone sees no benefit but a lot of risk in a situation they would be given a very high irrationality score for engaging in the situation.

§5 Intolerance of Uncertainty Scale Intolerance uncertainty was measured with the Intolerance of Uncertainty Scale (IUS-12; Carleton *et al.*, 2007). The IUS-12 is a two-factor scale, with factors "Prospective Intolerance of Uncertainty" (e.g. "I should be able to organize everything in advance") and "Inhibitory Intolerance of Uncertainty" ("analysis paralysis", e.g. "When it's time to act, uncertainty paralyses me."). Participants must rate their agreement with twelve statements on a five-point Likert scale from "Not at all characteristic of me" to "Entirely characteristic of me". Cronbach's alpha for the scale was 0.916 (95% C. I. $[0.907, 0.925]$).

§6 Distress Tolerance Scale Distress tolerance was measured with the Distress Tolerance Scale (DTS; Simons and Gaher, 2005). The scale consists of fifteen questions which measure how well people can tolerate distress. Participants must rate their agreement on a five-point Likert scale from "Strongly agree" to "Strongly disagree". A sample item is "When I feel distressed or upset, all I can think about is how bad I feel". Cronbach's alpha for the scale was 0.935 (95% C. I. $[0.928, 0.942]$).

5.1.2.2 Cognitive Reflection Test

Participants were given an expanded version of the Cognitive Reflection Test (CRT; Frederick, 2005; Toplak *et al.*, 2014). This version of the Cognitive Reflection Tests consists of 7 questions. This expanded version of the task was developed, in part, because of participant's potential familiarity with the Cognitive Reflection Task and proved to be a highly reliable measure of rational thinking (Toplak *et al.*, 2014). A sample item is "A bat and a ball cost $1.10 in total. The bat costs a dollar more than the ball. How much does the ball cost?". A participant responding intuitively might respond with "10 cents", but a participant who is able to override this intuitive response will pay attention to the "more than" wording and respond with "5 cents".

5.1.2.3 Measure of Cognitive Ability

Participants were given the Sample Test of the International Cognitive Ability Resource (ICAR-16; Condon and Revelle, 2014) to measure their cognitive abilities. The assessment probes participants' ability to solve problems of four subtypes: Letter and Number Series, Matrix Reasoning, Verbal Reasoning and Three-Dimensional Rotation. Participants responded to four questions in each of these four categories.

5.1.3 Procedures

All participants gave informed consent. Similar to Experiment 1, participants then went through a set of instructions for the Mouselab-MDP task and took a quiz on their comprehension of the instructions.[2] The quiz tested their knowledge of the environment structure and the bonus payment scheme.

After passing the instructions quiz, participants played 40 trials of the Mouselab-MDP task. Participants then took part in a post-task quiz where they were asked about the available values along the spider web as well as about the bonus rate and then, finally, about their motivation during the task. The mean reported motivation was 1.751 (SD: 0.600), rated on a scale from -2 ("Very unmotivated") to 2 ("Very motivated").

Participants were then given all the self-report measures, in a randomized order. Afterward, they were presented with the Cognitive Reflection Test and the ICAR-16. Finally, participants reported their demographics and overall effort and were given the chance to give us feedback. Sixteen of the seven hundred and ten final participants did not fill this section out due to technical issues, and one participant marked that they were unsure. The average overall effort was 2.921 (SD: 0.271), rated on a scale from 0 ("No effort (e.g., randomly clicking")) to 3 ("A lot of effort").

[2] The instructions quiz was almost identical to Experiment 1. During data collection, I added one line to the instructions in order to clarify a more ambiguous question.

Table 5.1: Mann-Whitney tests for parameters MLEs between Experiment 1 and Experiment 3. All tests are two-sided.

Parameter	Exp. 3 Median	Exp. 1 Median	Exp. 3 Mean Rank	Exp. 1 Mean Rank	Mann-Whitney U	adj. p-value	RBC
$w_{back.}$	2.50	1.75	420.35	394.07	40574.00	0.280	0.06
w_{depth}	0.00	0.00	412.07	442.25	46452.00	0.249	-0.07
$w_{dist.}$	1.00	1.00	410.83	449.50	47335.50	0.249	-0.09
$w_{forw.}$	1.00	1.00	415.17	424.25	44255.50	0.687	-0.02
$w_{eff.}$	1.00	1.00	420.47	393.39	40490.50	0.280	0.07
β	1.00	1.00	411.93	443.08	46553.00	0.280	-0.07

5.2 Results

5.2.1 Participant data was best fit by the cost model with all five cost weights.

As preregistered, I first performed model comparison on the participant data, considering all 32 (i.e., 2^5, different models for all combinations, with and without each one of the five cost parameters) possible cost models. The model containing all five cost weights explained the data best, according to the Bayesian Information Criterion. This result matches the model comparison presented in Chapter 3. In fact, the approximate log Bayes factor between the two models is even higher (i.e., ~ 1097 versus ~ 156 in Experiment 1) due to the larger number of participants.

5.2.2 The inferred parameter values were similar to those in Experiment 1.

As a secondary (not preregistered) check, I looked at if the MLE and the HDI spread differed between this experiment and Experiment 1. I compared each measure with a Mann-Whitney test for each cost parameter. I corrected the p-values for both sets of tests with a Benjamini-Hochberg correction with a false discovery rate of 0.05. I found no difference for any of the parameters for the MLE (see Table 5.1). I also found no difference for the HDI spread (see Table 5.2). The mean and standard deviations of each parameter in Experiment 3 can be found in Table 5.3.

Table 5.2: Mann-Whitney tests for parameters HDI spreads between Experiment 1 and Experiment 3. All tests are two-sided.

Parameter	Exp. 3 Median	Exp. 1 Median	Exp. 3 Mean Rank	Exp. 1 Mean Rank	Mann-Whitney U	adj. p-value	RBC
$w_{back.}$	7.50	5.00	420.87	391.08	46411.00	0.709	-0.07
w_{depth}	1.00	1.00	417.73	409.37	44180.00	0.846	-0.02
$w_{dist.}$	0.00	0.00	413.94	431.42	41490.00	0.709	0.04
$w_{forw.}$	2.50	2.50	418.94	402.28	45044.50	0.709	-0.04
$w_{eff.}$	2.50	2.50	418.93	402.35	45036.50	0.709	-0.04
β	0.00	0.00	416.36	417.34	43208.00	0.945	0.00

Table 5.3: Average Model Parameters for Experiment 3.

Parameter Name	Mean (Standard Deviation)
$\mathrm{w}_{back.}$	4.937 (4.670)
w_{depth}	0.744 (2.110)
$\mathrm{w}_{dist.}$	1.835 (2.826)
$\mathrm{w}_{forw.}$	2.600 (3.645)
$\mathrm{w}_{eff.}$	1.756 (2.589)
β	5.145 (14.052)

5.2.3 Inferred model parameters could not explain the variance in the self-report measures.

Before investigating the preregistered hypotheses regarding whether the R^2 increased with the addition of the model parameters, I checked for multicollinearity between the model parameters and demographic parameters. If there is multicollinearity between some parameters, this might lead to issues with identifying each individual variables' contributions to the regression model. I found the variance inflation factor was less than five for all parameters (see Table 5.7 for more information). This suggests that the independent variables in the preregistered regressions do not correlate with one other to a problematic degree.

Table 5.4: Variance explained in dependent variable by the model parameters.

Dependent Variable	f^2	F-test ($F(6,683)$)	p-value
Compulsive Behavior and Intrusive Thought	0.008	0.910	0.746
Anxious Depression	0.004	0.405	0.876
Social Withdrawal	0.008	0.891	0.746
Trait anxiety (STAI)	0.008	0.878	0.746
Consideration of Future Consequences	0.018	2.052	0.426
Propensity to Plan	0.010	1.125	0.746
Future Orientation Scale	0.014	1.540	0.746
Life Regrets	0.012	1.414	0.746
Life Satisfaction	0.005	0.521	0.850
Brief Inventory of Thriving	0.005	0.528	0.850
Risk-taking	0.007	0.830	0.746
Lack of Premeditation (UPPS-P)	0.011	1.255	0.746
Intolerance of Uncertainty	0.005	0.614	0.850
Distress Tolerance Scale	0.008	0.886	0.746
Cognitive Reflection Task	0.025	2.901	0.127

Next, for each hypothesis, I performed two multiple linear regressions, one with just the demographic variables as independent variables and another with the demographic variables as well as the model parameters. I then compared the increase in R^2 with the model parameters included in the linear regression via an F-test. I found that my model parameters did not significantly increase the proportion of explained variance (R^2) for any of the questionnaires (see Table 5.4). Because of this result, I did not perform any of the pre-registered follow-up analyses on whether individual regression weights were significantly greater than zero.

The results of the preregistered analyses suggest that individuals' cost model parameters cannot explain the variance in their responses to the questionnaires above and beyond simpler demographic variables and a measure of intelligence. As preregistered, I corrected all p-values of the tested hypotheses using a Benjamini-Hochberg correction with false discovery rate of 0.05 (Benjamini and Hochberg, 1995). In fact, the closest measure to being explained by model parameters was the Cognitive Reflection Test (Cohen's $f^2 = 0.025$, $F(6,683) = 2.90$, adj.$p = 0.127$). There has been recent work showing that tasks are more likely to correlate with each other than self-report surveys (Eisenberg *et al.*, 2019; Dang *et al.*, 2020). Therefore, this near-explainability may be due in part to the fact that the Cognitive Reflection Test is also a task and not a self-report survey.

5.2.3.1 The achieved power of the performed analyses was very high.

Since I stopped recruiting participants due to a preregistered financial stopping rule, I then looked at post-hoc power achieved. I wanted to rule out the case that the analysis might have had a much lower power than intended, which would mean we have a higher chance of Type II error (false negative, i.e., failure to find an effect present in the data).

To calculate the achieved power, I used G*Power (Faul *et al.*, 2009) as in the preregistration. I tested for "Linear multiple regression: Fixed model, R^2 increase". This time I entered an effect size of $f^2 = 0.05$, as preregistered, and a total sample size of 694 as achieved in these regression analyses. Since none of the cost-specific hypotheses were tested, I entered a p-value that was Bonferroni-corrected for 15 hypotheses, $0.05/15$. The total number of predictors was the number of model parameters (6) plus the demographics, age, gender and intercept (4), while only the model parameters themselves were tested. This post-hoc analysis results in a power of 0.97264 being achieved, so the chance of a false negative (Type II error) is quite low.

Note that the Bonferroni correction is even stricter than the Benjamini-Hochberg correction, but since we cannot reject the null hypothesis for any of the tests, we could not calculate an effective p-value threshold to use in the power analysis. The chance of a Type I error (false positive) is probably lower than 5%, although it is not relevant given the negative results presented here.

5.2.4 Simpler behavioral measures can explain the variance in the Cognitive Reflection Test.

Next, as an exploratory analysis, I performed the same analyses as those in the above section, but instead of the model parameters I inputted a participants' average number of clicks on immediate, intermediate and final outcomes. I expected to see a correlation between more optimal planning behavior (more clicks on the later versus earlier nodes) and planning and future thinking measures.

I found that a statistically significant portion of the variance of participants' scores on the Cognitive Reflection Test is explained by the number of clicks on each level of the web, above and beyond the variance explained by the demographic variables and IQ (Cohen's $f^2 = 0.021, F(3,686) = 4.84, p = 0.037$). None of the other analyses yielded a significant portion of variance explained by adding the clicking behavior into the regression (see Table 5.5)

As a follow-up analysis similar to the preregistered analyses, I then tested whether any of the coefficients, particularly the number of clicks on each of the three levels (i.e., first, second, and third) were significantly greater than zero. I corrected the p-values for multiple comparisons using a Benjamini-Hochberg correction with a false discovery rate of 0.05. I found that the coefficient for the number of clicks on intermediate outcomes ($t(686) = -3.20$, adj.$p = 0.003$, $\beta = -0.17$) was significantly less than zero whereas the coefficient of the number of clicks on final outcomes ($t(686) = 3.10$, adj.$p = 0.003$, $\beta = 0.11$) was significantly greater than zero. I also found that some of the demographic and IQ coefficients significantly differed from zero (IQ: $t(686) = 17.22$, adj.$p < 0.001$, $\beta = 0.54$; male gender: $t(686) = 5.09$, adj.$p < 0.001$, $\beta = 0.32$; intercept: $t(686) = -3.88$, adj.$p < 0.001$, $\beta = -0.19$).

These results suggest that people who follow more optimal planning strategies in the Mouselab-MDP paradigm (i.e., higher prioritization of the most informative final outcomes) are better able to override their intuitive responses on the Cognitive Reflection Test. This gives credence to the use of this version of the Cognitive Reflection Task as a measure for rationality (Toplak *et al.*, 2014). However, it should be cautioned that it appears that IQ itself contributes just as much to the Cognitive Reflection Test score.

5.2.5 Inferred strategy could not explain variance in the self-report measures.

Next, I looked at the questionnaire scores versus the proportion of participants' trials classified as each strategy according to the computational process tracing method by (Jain *et al.*, 2022). I considered only those strategies classified as the most frequent strategy for 5% of participants. This ended up amounting to three strategies: the optimal strategy (Strategy 21), a depth-first strategy (Strategy 31), and a strategy which sparsely explore final outcomes (Strategy 57). Again, I replicated the preregistered analyses done with the planning cost weights. However, I found none of the performed tests reached the significance threshold after applying the Benjamini-Hochberg correction (see Table 5.6).

Table 5.5: Variance explained in dependent variable by behavioral measures.

Dependent Variable	f^2	F-test ($F(3,686)$)	p-value
Compulsive Behavior and Intrusive Thought	0.001	0.335	0.809
Anxious Depression	0.003	0.640	0.718
Social Withdrawal	0.006	1.317	0.655
Trait anxiety (STAI)	0.004	0.909	0.684
Consideration of Future Consequences	0.010	2.371	0.520
Propensity to Plan	0.004	0.871	0.684
Future Orientation Scale	0.005	1.098	0.655
Life Regrets	0.008	1.813	0.538
Life Satisfaction	0.006	1.346	0.655
Brief Inventory of Thriving	0.008	1.942	0.538
Risk-taking	0.003	0.589	0.718
Lack of Premeditation (UPPS-P)	0.001	0.322	0.809
Intolerance of Uncertainty	0.005	1.140	0.655
Distress Tolerance Scale	0.003	0.696	0.718
Cognitive Reflection Task	0.021	4.838	0.037

5.2.6 Planning cost weights still correlate with questionnaire measures.

In the previous three analyses, I have found that task variables cannot explain variance in the questionnaire measures above and beyond demographic variables. This lack of explanation means we cannot use the task variables in a diagnostic, or predictive way. However, the question remains: do the variables in the task have anything to do with the questionnaire measures? In order to investigate this, as an exploratory analysis I looked at the Spearman correlation between each task variable and questionnaire measure.

Note that this data is not really well-equipped for answering the question of whether or not these correlations really exist, especially for lower correlation levels. Even for an uncorrected α level, the achieved power for detecting a 0.10 correlation is ~ 0.76. When correcting for all 34 tested hypotheses with a Bonferroni correction, the achieved power would be ~ 0.30. Therefore, I report tests without correcting p-values, and caution the reader that any correlations here may just be false positives. Only the correlations between inferred distance cost parameter weight and IQ or CRT score would survive a Benjamini-Hochberg correction with false discovery rate 0.05 (we would be able to reject the null hypothesis that they are not correlated). I mark these correlations with an asterisk (*) below.

Another point of caution is that for most of the psychiatric questionnaires, participants were only presented a portion of the items. At the most extreme, only one item was

Table 5.6: Variance explained in dependent variable by most-frequent strategy.

Dependent Variable	f^2	F-test ($F(3,686)$)	p-value
Compulsive Behavior and Intrusive Thought	0.003	0.757	0.814
Anxious Depression	0.002	0.484	0.814
Social Withdrawal	0.004	0.877	0.814
Trait anxiety (STAI)	0.002	0.391	0.814
Consideration of Future Consequences	0.013	3.070	0.205
Propensity to Plan	0.008	1.889	0.650
Future Orientation Scale	0.007	1.587	0.717
Life Regrets	0.002	0.484	0.814
Life Satisfaction	0.002	0.560	0.814
Brief Inventory of Thriving	0.003	0.674	0.814
Risk-taking	0.002	0.435	0.814
Lack of Premeditation (UPPS-P)	0.003	0.693	0.814
Intolerance of Uncertainty	0.001	0.166	0.920
Distress Tolerance Scale	0.002	0.500	0.814
Cognitive Reflection Task	0.014	3.099	0.205

presented from the AUDIT questionnaire.

5.2.6.1 Higher non-backward search cost weights correlate with increased propensity to regret.

I found the non-backward search cost weight was correlated with scores on the regrets scale ($\rho(700) = 0.09$, $p = 0.019$, 95% C. I. $[0.01, 0.16]$), scores on the Cognitive Reflection Test (Spearman's $\rho(697) = 0.08$, $p = 0.028$, 95% C. I. $[0.01, 0.16]$), and the post-task structure-knowledge score (Spearman's $\rho(700) = 0.07$, $p = 0.049$, 95% C. I. $[0.00, 0.15]$). This implies that participants who follow a backwards search strategy might also experience slightly more regret in their day-to-day life. They also may understand the task slightly more than others (which makes sense, as the optimal strategy is backwards search.) See Figure A.15 for a plot showing the relation between all tested variables and the non-backward search cost weight.

5.2.6.2 Lower planning depth cost weight correlates with increased SW and CIT factor scores.

I found several psychiatric measures were correlated with the depth cost weight (Liebowitz Social Anxiety Scale: Spearman's $\rho(700) = -0.08$, $p = 0.038$, 95% C. I. $[-0.15, 0.00]$; Apathy Evaluation Scale: Spearman's $\rho(700) = 0.11$, $p = 0.003$, 95% C. I. $[0.04, 0.18]$; Self-Rating Depression Scale: Spearman's $\rho(700) = -0.08$, $p = 0.032$, 95% C. I. $[-0.15, -0.01]$;

Obsessive-Compulsive Inventory-Revised: Spearman's $\rho(700) = -0.09$, $p = 0.021$, 95% C. I. $[-0.16, -0.01]$). This seems to suggest that people with more symptoms of social anxiety, depression and OCD have less cost for planning further into the future. However, those with more apathy have a higher planning depth cost weight.

Similarly, there was a negative correlation for the Compulsive Behavior and Intrusive Thought Factor (Spearman's $\rho(700) = -0.09$, $p = 0.013$, 95% C. I. $[-0.17, -0.02]$) and the Social Withdrawal Factor (Spearman's $\rho(700) = -0.08$, $p = 0.036$, 95% C. I. $[-0.15, -0.01]$).

I also found that participants with more life regrets seem to have lower planning depth costs (Spearman's $\rho(700) = -0.07$, $p = 0.050$, 95% C. I. $[-0.15, 0.00]$) as did participants with more propensity to regret (Spearman's $\rho(700) = -0.08$, $p = 0.026$, 95% C. I. $[-0.16, -0.01]$). Additionally, I found that older participants had higher planning depth costs (Spearman's $\rho(697) = 0.11$, $p = 0.005$, 95% C. I. $[0.03, 0.18]$). This will be explored more in depth in Subsection 5.2.7. See Figure A.17 for a plot showing the relation between all tested variables and the depth cost weight.

5.2.6.3 Higher distance cost weights correlate with lower IQ and CRT scores.

I found that the distance cost weight was correlated with IQ* (Spearman's $\rho(694) = -0.13$, $p < 0.001$, 95% C. I. $[-0.21, -0.06]$) and scores on the Cognitive Reflection Test* (Spearman's $\rho(697) = -0.15$, $p < 0.001$, 95% C. I. $[-0.22, -0.08]$). Additionally, there was a slight negative correlation between distance cost weight and the propensity to regret (Spearman's $\rho(700) = -0.09$, $p = 0.018$, 95% C. I. $[-0.16, -0.02]$) and understanding of the task in the post-task score (Spearman's $\rho(700) = -0.10$, $p = 0.009$, 95% C. I. $[-0.17, -0.02]$). See Figure A.16 for a plot showing the relation between all tested variables and the distance cost weight.

5.2.6.4 The non-forward search cost weight correlates with IQ and CRT.

Higher scores in the forward search cost weight, meaning more forward search, correlated positively with scores in the Future Orientation Scale (Spearman's $\rho(700) = -0.11$, $p = 0.003$, 95% C. I. $[-0.18, -0.04]$) and Regrets Scale (Spearman's $\rho(700) = -0.08$, $p = 0.036$, 95% C. I. $[-0.15, -0.01]$). Lower scores in the forward search cost weight correlated negatively with scores in the UPPS-P Lack of Premeditation factor (Spearman's $\rho(700) = 0.10$, $p = 0.008$, 95% C. I. $[0.03, 0.17]$). See Figure A.14 for a plot showing the relation between all tested variables and the non-forward search cost weight.

5.2.6.5 Lower effort cost weights correlate with higher IQ and CRT scores.

I found that the effort cost weight was correlated with IQ (Spearman's $\rho(694) = -0.10$, $p = 0.011$, 95% C. I. $[-0.17, -0.02]$) and scores on the Cognitive Reflection Test (Spearman's $\rho(697) = -0.09$, $p = 0.019$, 95% C. I. $[-0.16, -0.01]$). See Figure A.13 for a plot showing the relation between all tested variables and the effort cost weight.

5.2.7 Forward search is associated with increased age.

Previous work shows that older adults generally plan less optimally than younger adults. For example, older adults have been shown to plan more noisily and to lower depths than younger adults (Steffen *et al.*, 2023) as well as in a more model-free way (Bolenz *et al.*, 2019). Particularly, in the Mouselab-MDP Task, younger adults are more likely to employ the optimal strategy than older adults, who are more likely to use a depth-first search technique (Das *et al.*, 2019). Therefore, I looked at the correlation between age and the model parameters.

I found that age was indeed correlated with the inferred depth cost weight ($\rho(697) = 0.11$, adj.$p = 0.032$, 95% C. I. $[0.03, 0.18]$) but not the non-forward search cost weight ($\rho(697) = 0.06$, adj.$p = 0.373$, 95% C. I. $[-0.02, 0.13]$). No other model parameters were significantly related to age. This suggests that older participants are more prone to following a (less than optimal) strategy, focusing on the earlier nodes which don't provide as much information as near-optimal strategies. I confirmed that the age of participants using forward search strategies (as classified by the Computational Microscope) was higher than for other strategies ($M_{forward} = 36.379$, $M_{other} = 31.614$; Mann-Whitney $U = 11612.00$, RBC $= -0.20$, $p = 0.035$, greater). This further supports the narrative that older adults are planning in a different way than younger adults.

But could this be due to a difference in understanding the task? To answer this question, I assessed whether older participants were likely to answer all three of the post-test questions on the possible reward distributions behind the nodes. I found that there was a small negative correlation between age and structure understanding (Spearman's $\rho(697) = -0.09$, $p = 0.017$, 95% C. I. $[-0.16, -0.02]$). These results suggest that older adults may follow more of a forward-search strategy, and plan to lower depths due to impaired structure understanding.

This apparent focus on planning forward from the start and the apparent increased cost for planning further into the future, may be related to some older participants having a misunderstanding of how the task works, such as thinking they had to explore nodes starting from where the spider started. To test this hypothesis, I removed all 41 participants who planned in a purely forward-search manner in the last twenty test trials and performed the same analysis as above. With these participants removed, there was no longer a correlation between age and depth cost weight (Spearman's $\rho(657) = 0.07$, adj.$p = 0.338$, 95% C. I. $[0.00, 0.15]$). Again, no other model parameters were significantly related to age. This suggests that the correlation between non-forward search cost weight and age was driven mostly by individuals who plan only in a forward search manner. Future work might look at whether older adults continue to plan in a forward search manner, even when instructed about the exact reward distributions and given examples in the instructions phase where it is apparent they do not have to click in the same order that the spider walks.

Variable	Variance Inflation Factor
$w_{back.}$	1.29
w_{depth}	1.37
$w_{dist.}$	1.47
$w_{forw.}$	1.62
$w_{eff.}$	1.17
β	1.34
const	1.70
age	1.04
IQ	1.07
female	1.03
other	1.01

Table 5.7: Variance Inflation Factor for Independent Variables.

5.3 Discussion

The cost parameters outputted by my method could not explain the variance in the individual self-report measures, above and beyond demographic variables and IQ. However, I did find in exploratory analyses that there were some smaller correlations between planning cost weights and the self-report measures. Additionally, I replicated previous findings related to differences in planning as people age.

The results suggest that the measured cost weights do not capture trait-like individual differences in planning and future-thinking measures. However, there were still some smaller correlations between planning cost weights and self-report variables. While exploratory, I nevertheless reported these correlations in case they could be useful to researchers trying to understand planning as it relates to mental disorders. Future work is needed to investigate these exploratory results in a more rigorous manner.

One possible explanation for the lack of predictive results using the cost model is that there is no evidence that the inferred parameters are trait-like (Schmitt and Blum, 2020). If someone is inferred to have a high backwards search cost and low planning depth cost weight today, will the same be true if they are tested in thirty minutes, tomorrow, or in a year (controlling for any learning or fatigue)? One major weakness of this study is that I compared these cost parameters to (more reliable and stable) self-report measures before knowing much about the parameters' reliability.

Focusing on a subclinical, broad spectrum of psychopathology could have also had an impact on the predictive power of the planning cost weights. Perhaps if I had recruited people with clinical levels of mental disorders, there would have been more variance in the dependent variables that could've been explained by the planning cost weights. For example, people diagnosed with OCD may show lower effort costs due to decreased integration of past states/actions which may lead to impaired structure learning (Seow *et al.*, 2021; Fradkin *et al.*, 2020). Future work could look into how inferred planning costs vary for people with clinical levels of specific key disorders, like OCD.

One other reason we may not have been able to explain variance in the self-report surveys above and beyond demographic variables was that these relationships may not be linear. For example, there may be a quadratic relationship between some of these variables, where two extremes of symptomatology would both have similar aberrant planning costs. Tying into the point above, the relationship could also be more like a step function, with the relationship only becoming apparent for extreme values of psychiatric symptomatology or planning costs. Future work could look at investigating this in a more principled manner.

Finally, a minor weakness may be in the selected psychiatric questionnaires and the factor weights used. Since preparing and running the experiment, an improved reduction of questionnaire items has been proposed and validated with a larger, more diverse dataset in Hopkins *et al.* (2022). Since the weights and reduction of questionnaire items I used were based on limited data, there may have been more noise in the transdiagnostic factors than had I not reduced the questionnaire items.

Further, in the years since Gillan's seminal work (Gillan *et al.*, 2016), there has been work on developing new transdiagnostic classifications, such as the Hierarchical Taxonomy of Psychopathology (HiTOP; Kotov *et al.*, 2017). The HiTOP Clinical Translation Workgroup has developed a different battery of questionnaires (Simms *et al.*, 2022). This battery of questionnaires does not yield three factors, but instead six broad spectra of psychopathology. Perhaps using this model of transdiagnostic categories would have yielded different results.

Another factor that may play a role is that self-report measures themselves are imperfect. While the self-report measures on planning and future-thinking do seem to correlate with real-world outcomes (see Chapter 2), it is possible that my method is measuring real individual differences in planning that are not captured by these self-report measures.

In this chapter, I found that the cost parameters inferred by my method do not seem to be predictive of psychiatric symptomatology or dimensions of real-world planning. However, I did find, in exploratory analyses, that some of the cost weights may be slightly correlated to different questionnaire measures. For example, higher scores in the "Compulsive Behavior and Intrusive Thought" factor and "Social Withdrawal" factor correlate with lower planning depth cost. While my method for inferring costs in the Mouselab-MDP task may not work well as a diagnostic tool for planning ability, it is possible that it could be useful for understanding individual differences in planning that aren't captured by the self-report measures.

Chapter 6

Conclusions and Future Directions

In this dissertation, I developed a method for inferring planning costs and investigated how the inferred costs related to psychiatric symptomatology and decision-making in the real world. In this chapter, I summarize the contributions of each chapter and outline limitations and future directions.

6.1 Summary of contributions

In Chapter 3, I introduced a cognitive model of individual differences in planning using planning costs, which serves as the backbone for the method to infer individual differences in planning costs. In fact, this cognitive model explains planning behavior better than an alternative model tested which includes cost parameters which mirror features of the previous best fitting model in the task. Additionally, I systematically evaluated this model compared to 32 models: 31 submodels as well as a random null model. I found that simulated behavior from this model matches participant behavior in the task. This suggests that human planning is, to some degree, rational according to some individual constraints.

In Chapter 4, I then presented a method to infer planning costs using this cognitive model, and validated it on simulated and human planning data. I found that the method can reliably recover planning cost weights from simulated data, with moderate to strong correlations for all model parameters. This method is also more parsimonious than the existing method for extracting individual differences in planning, the Computational Microscope. Rather than classifying each trial as arising from one of 79 defined strategies my method outputs five cost weight values and a temperature value.

Finally, in Chapter 5, I presented a large preregistered study pairing the task with questionnaires and the Cognitive Reflection Test (CRT). I found that the inferred cost weights could neither predict scores on self-reported planning and future thinking measures nor psychiatric measures. However, in exploratory analyses, I did find that some of the inferred cost weights were slightly correlated with several of the measures.

6.2 Limitations & future directions

In this section, I discuss the limitations of the work presented in the previous chapters and outline future work that could be done to further validate and improve my method. I also lay out how the results presented in this dissertation could, despite the drawbacks outlined in this dissertation, nevertheless be used in developing personalized planning interventions.

6.2.1 Inverse problems

Inferring subjective planning cost weights from human behavior is an inverse problem. To be "well-posed", an inverse problem must have a unique solution as well as satisfy a stability constraint, meaning the solution should not change too much based on small perturbations to inputs (Groetsch, 2015). Unfortunately in this case, and in general in the standard inverse reinforcement learning setting, this is not the case. The problem is "ill posed": many different reward functions may lead to the same behavior.

While this is unfortunately not something that can be fixed, in some ways, the way I constructed the reward function may have accentuated its non-identifiability. For example, since I wanted to distinguish between motor effort and cognitive effort when it comes to clicking on the further-out nodes, I included the distance cost weight. However, these cost parameters do not look all so different when it comes to clicking on futher-out nodes due to the placement of the nodes in geometric space. In fact, someone following the optimal strategy should have both a low depth cost weight and low distance cost weight. Additionally, Bustamante *et al.* (2023) found, by augmenting a patch foraging task to measure both cognitive and physical effort, that cognitive and physical effort are highly correlated. This is notable, as the two types of effort are not as confounded in this patch foraging task as in the Mouselab-MDP task.

Future work could look at trying to disentangle these two elements of the cost of planning, perhaps by coming up with a setting of the Mouselab-MDP task where the two cost components are not correlated. One such example would be placing the nodes in more of a tree, where the nodes at each level are at the same height on a webpage. Another more out-of-the-box idea might be to get rid of the differences in motor cost entirely, and allow users to uncover nodes in another way (e.g., spoken).

Finally, since I only looked at the later trials, the cost parameter weights were not constrained by information about behavior in the earlier trials. Adapting the method to take into account learning, as discussed in Chapter 4 might allow us to capture more information about the underlying behavior of participants in the task. Using data from the learning period may help to pinpoint which possible combination of cost parameter weights drove the observed behavior.

6.2.2 Investigation of larger scale environments

Future work should also look at larger, less toy-like Mouselab-MDP environments. The work in this dissertation was limited to a smaller environment, because of a previous implementation of computing the state-action values which required dynamic programming. Now that the method does not rely on a dynamic programming solution to the Meta-level MDP, but instead uses approximations of the state-action values, it might make sense to look at larger environments and whether inferred cost parameters in these larger environments are more related to real-world planning than those in a 3-step environment.

In a larger environment, people may rely less on their working memory and more on heuristics due to computational intractability (Gigerenzer, 2008). Using larger environments with more node levels may also help with the separability of different cost components. For example, in an environment with more branches and node depths, it may be easier to differentiate between the non-forward search cost weight and the depth cost weight.

Finally, using larger environments might help to elicit differences in planning that may be driven by demographic and personality factors. More recent work looking at Big-Five personality traits and gender in the information-acquisition stage of planning (using the Mouselab task) found that individual differences only emerged in task configurations with a larger search space (Le and Jang, 2023). Real-world planning is often more complex, with more potential outcomes to consider on longer time scales. Differences in individual differences in planning costs may only become more apparent as the size of the planning problem increases. Perhaps extending the method to larger environments could then lead to a measurable predictive relationship between the method outputs and the planning and future thinking scores.

6.2.3 Continuous optimization

As discussed in Chapter 3, the previous implementation of the state-action values necessitated that the optimization be performed via grid-search. This was due to the fact that the dynamic programming solution itself took quite a while, so I cached the results for all seen state-action pairs for different values of the cost parameter values on a grid. Now that approximations of state-action values are used, this is not needed. Due to the constraints of needing to wrap up this dissertation, I could not switch the optimization method used here.

Using continuous optimization would most likely speed up model fitting for anyone planning on applying this method to their data. More importantly, parameter estimates would also be more precise, making the method more sensitive to individual differences. It may also help identify promising sections of the parameter space that were not explored in the present work.

6.2.4 Are cost weights trait-like?

One weakness of the method presented here is that we cannot be sure whether the inferred cost parameters are trait-like or just a product of the present moment. This means we cannot be sure if the cost parameters inferred now will remain the same some time after the experiment, when they are re-tested. Future work could look into how stable these inferred cost parameters are over time.

6.2.5 Do cost weights vary by context?

Even if these cost parameter weights are trait-like, they may depend on context. For example, with a social framing, people higher in social anxiety might be inferred to have different cost weights. Inducing stress has also been shown to cause differences in model-free versus model-base behavior for people with higher levels of depression (Heller *et al.*, 2018). Therefore, future work might look at how planning in the task differs in various contexts. After all, real-world planning itself often occurs in social and stressful situations.

6.2.6 Executing plans

Scholnick and Friedman (1993) have outlined the five steps of planning: representing the problem, goal-setting, choosing to plan, formulating a plan and finally, execution and monitoring. Two of these steps in particular are of interest when looking at the Mouselab-MDP paradigm: choosing to plan and formulating a plan. It is very easy to make a plan, but creating a plan which can be stuck to is more difficult. For example, developing a plan for marathon training may be easy, but actually running the marathon relies on not immediately giving up on a rainy week (“execution”) and adapting the plan when life happens (“monitoring”). The Mouselab-MDP paradigm could be extended to include more of the “Execution and monitoring” portion of planning in order to investigate how participants deal with this. This may allow researchers to study planning in a more ecologically valid way, further isolating planning versus mathematical aptitude.

Planning as studied with the current version of Mouselab-MDP paradigm could be seen as more of a problem-solving task. This may explain why it seemed to be so predictive of CRT scores. Some people who excel at the problem-solving aspect of the task may nevertheless not plan in their real lives, especially if they know they will not stick to or adapt their plans. Therefore, extending the task to capture these parts of planning may help in developing a more robust measure-based method for inferring individual differences in planning propensity.

6.2.7 Individualized planning interventions

In Chapter 1, I mentioned that understanding individual differences in planning could help in selecting (and developing) personalized planning interventions. Despite the drawbacks of my method, I believe it could still help in selecting personalized planning interventions. Here, I lay out the motivation and how this method could be used in the selection of personalized planning interventions.

The resource-rational model of planning introduced in Callaway *et al.* (2022) and expanded here can provide some insight on how to intervene on people's planning in cases where the potential rewards or losses are much larger than the present ones. There are two direct ways to intervene on planning: reducing the perceived cost of planning or increasing the perceived benefit to planning. Several planning interventions have already been implemented, both in the Mouselab-MDP task and in other planning tasks. Here, I briefly introduce both types of planning interventions and discuss previous work intervening on planning in the Mouselab-MDP task.

§1 Cost Interventions Cost interventions make planning easier, and are seen often in the real-world. People such as financial planners, career counselors or even a friend with more experience or tools such as decision support tools or flowcharts can alleviate the burden of planning. A person might already see the value in putting effort into a plan or decision, but feel overwhelmed or ill-equipped to do so themselves.

In the Mouselab-MDP task, a cost intervention could be a tool that works on the specific costs that participants might experience: voice-activated node inspection in the case of motor-costs, already-revealed nodes, or even a running total of how much each path is worth. To my knowledge, none of these methods have been tested. The experiment I ran in Chapter 4, where I manipulated participants' costs, is close to what could be considered a cost intervention.

§2 Value Interventions One interesting thought brought up by Gabaix and Laibson (2017) is that apparent myopic planning may be caused by imperfect information about the future. Misalignment of values can be based on not just a steep discount rate (see Subsection A.1.3) but also the inability to forecast into the future. This could be seen in discounting how important it is to start contributing to a retirement account at a young age or how important it is to continue to engage in physical activity as one ages. It can be hard to forecast just how important such actions are, without having direct experience to draw inspiration from.

Value interventions are common in education, in the context of expectancy-value theory (Eccles, 1983). The idea here is that students' motivation is a product of both their expectation of success and how much they care about success. Value could also be intervened on, for example via education about the utility of math and science education on students' future prospects (Harackiewicz *et al.*, 2012). However, one may also intervene on motivation by increasing the expectation of success via reframing (e.g., educating that success is within the student's control) or ingraining a growth mindset (Hulleman *et al.*, 2016).

In the Mouselab-MDP task, people may also have imperfect information about the task, and how rewarding it can be. For example, an online respondent may assume that the task is not very rewarding: perhaps most experiments they took part in were not rewarding. A value intervention could educate participants about how rewarding the task could be, if they put effort into planning, by telling participants how much they could reasonably expect (in real money) to earn on each trial, or how their performance compares to the optimal participant. This may help to improve planning for demotivated online participants, particularly those who assume bonuses are random or uniform. Since I assumed participants had perfect information about the reward distribution behind nodes, this is not directly related, but future work could either make sure participants do have perfect knowledge of this reward distribution, or query participants' knowledge and expectations before attempting one of these interventions.

Becker *et al.* (2023) paired the Mouselab-MDP task with "reflection prompts". While completing trials of the Mouselab-MDP planning task, participants were given modified reflection prompts by Wolfbauer *et al.* (2020). I consider this to be similar to a value intervention, because through self-reflection participants may reflect on the value of planning and the efficacy of their actions. Becker *et al.* (2023) found that participants who engaged in these reflection prompt rounds engaged in more future-thinking planning strategies (as measured by the Computational Microscope), compared to participants in a control group with no reflection prompts. Another way to intervene on value is to instruct participants on the most valuable action in the current state. This is similar to the work done in Lieder *et al.* (2019b), where participants in a Mouselab-MDP experiment were told which node they should have inspected.

§3 Implications of my findings for personalizing interventions to improve planning

While the parameters uncovered by my method should not be treated as trait-like or stable, they may still be able to provide a snapshot of subjective planning costs at the moment. Future work might test how individuals with different parameter combinations respond to different interventions. In particular, it would be interesting to investigate whether cost interventions, informed by individuals' pre-intervention cost weights are more effective than value interventions.

In Chapter 5, I found exploratory evidence for a slight correlation between the non-forward search cost weight and scores in the UPPS-P Lack of Premeditation factor. I also found a predictive relationship between the CRT and number of clicks per node, as well as smaller correlations between CRT scores and several of the cost parameters. Following these exploratory results, reflection prompts targeted at specific types of poor performance could be developed. These reflection prompts could be further personalized based on inferred cost weights in some baseline trials. Specifically, some reflection prompts could focus on future consequences (targeted at those with higher non-forward search cost weights), while other reflection prompts could focus on reflecting about what is the best choice (targeted at those with higher distance and effort cost weights).

6.3 Conclusions

Taken together, the work in this dissertation highlights the importance of individual differences in the study of human planning and decision-making. Specifically, I found that a significant proportion of individual differences in planning can be explained as a rational adaptation to subjective planning costs. While the subjective planning costs outputted by the method cannot predict psychiatric symptomatology or future thinking and real-world planning (as measured by the chosen self-report questionnaires), in exploratory work, I did find smaller correlations between several self-report measures and inferred subjective planning costs. Taken together, this work underscores the need to carefully establish that a model measures what we think it does before moving onto transdiagnostic work and real-world conclusions. I hope that the work presented here can act as a case study on what may or may not work in model development, and that my exploratory findings in Chapter 5 can inform future studies of human planning.

Appendix A

Appendix

A.1 Chapter 3

A.1.1 Participant strategies for Experiment 1

As described in Subsubsection 3.2.3.1, I used the Computational Microscope as a measure of when participants have reached a stable policy and stopped learning. The Computational Microscope gives us the most likely strategy a participant was using for each trial. I then calculated the most common strategy for the last 20 trials of the experiment.

74 out of 122 participants operated via Strategy 21 of the Computational Microscope. This strategy approximates the optimal strategy by exploring the final outcomes first. The next most frequently used strategies were 31 and 37, with 7 participants each. Strategy 31 approximates depth first search, while strategy 37 inspects the nodes closer to the start and then follows a best-first-search strategy.

A.1.2 Simulated data from the best model matches human behavior just as well as data generated from alternative models

As in Subsubsection 3.2.3.4, I simulated 20 trials for each parameter combination demonstrated by participants. I computed the number of clicks per level, per trial for each simulated participant for all of the possible cost models and computed the average node-level click rate per cost weight combination. I then performed a Friedman test for each click category to see if there was a difference in clicking behavior between the participants and the simulated optimal participants between cost models.

I found no significant difference between the individual levels of clicks in total ($F(1.98, 240.02) = 2.631$, $p = 0.075$) or for all levels of nodes (early: $F(1.98, 240.02) = 2.383$, $p = 0.095$; middle: $F(1.98, 240.02) = 0.419$, $p = 0.657$; late: $F(1.98, 240.02) = 0.381$, $p = 0.682$). Since the model without the depth cost weight had a higher correlation for later nodes (see Table A.1), I nevertheless conducted pair-wise Friedman tests between each pairs of models for each type of clicks. I corrected the p-values of these pair-wise tests using a Benjamini-Hochberg correction with a false discovery rate of 0.05. There were no significant differences between any of the pairs of models for any types of click. These results suggest that simulated behavior from all three models match participant data just as well as others.

A.1.3 Does the addition of a discount rate and a concave utility function explain the data better than the depth parameter?

Neglecting to plan into the future could arise from both experienced cost for planning into the future (depth cost weight) as well as decreased value for planning. Two supported factors which may lead to a decreased value of planning are a discount rate and people having concave utility function Tymula *et al.* (2013). These two parameters capture a general propensity to discount future values ($4 now might matter more than $24 in two time steps if the experiment ends) and that participants might vary in their marginal utility for earning rewards.

I augmented my cost model to include a discount rate and power utility function in the model. For the discount rate γ, I modeled the state rewards as being discounted by γ for each step. I assumed first level node values were drawn with equal probability from the values shown to participants, the second level node values were modeled as having been drawn with equal probability from the shown possible values multiplied by γ and the third level node values were drawn with equal probability from the shown possible values multiplied by γ^2. Specifically I transform the reward r_n by $U(r_n) = r_n \cdot \gamma^{d_n - 1}$ whether $d(n)$ is the depth of node n, as defined above. This step-by-step discount rate might relate to uncertainty of whether the task would continue and reward would be experienced in the next step.

Likewise, I modeled the state rewards as being transformed by a power utility function $U(r_n) = \text{sign}(r_n) * r_n^{\kappa}$. The value of 0 for the risk aversion parameter κ might be interpreted as a complete indifference to the task bonus, while a risk aversion parameter of 1 would be interpreted as experiencing the rewards of the task as they were set up.

I varied κ and $\gamma \in \{0.10, 0.30, 0.50, 0.75, 0.90, 1\}$.

First, I looked at model comparison using BIC to see whether the addition of these parameters was useful in the model. This new seven parameter model had a much lower BIC than the five parameter model, with an approximate log Bayes factor of at least 63 between all other models (see Figure A.1).

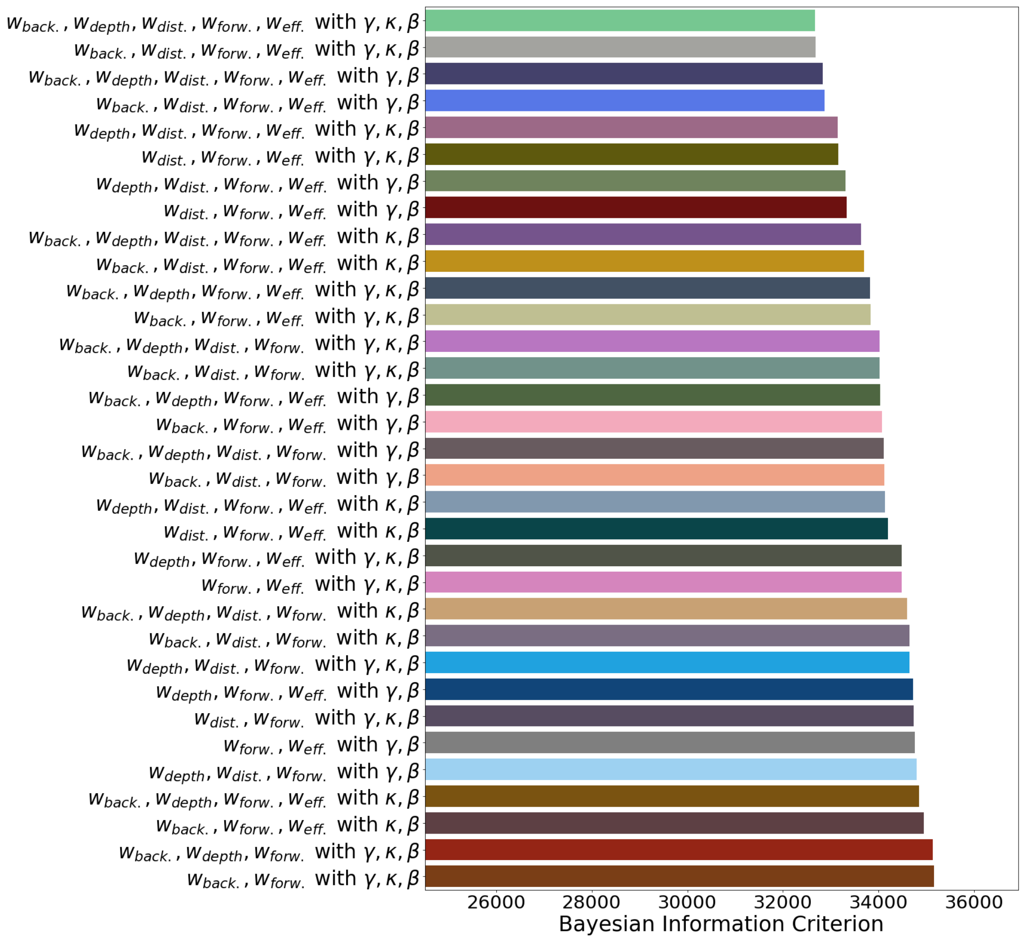


Figure A.1: The fixed-effect model comparison with the expanded range of models (not all models pictures here).

However, when looking at behavior from participants simulated with this seven parameter model versus real participant behavior, I found that behavior simulated from this model often did not match participant behavior when it came to the number of clicks on late nodes (Spearman's $\rho(122) = 0.23$, $p = 0.011$, 95% C. I. $[0.05, 0.39]$). Therefore, I conclude that, while there are some merits to including this discount factor and a concave utility function, I will focus on the cost-only model for this dissertation.

A.2 Tables for Chapter 3

A.3 Figures for Chapter 3

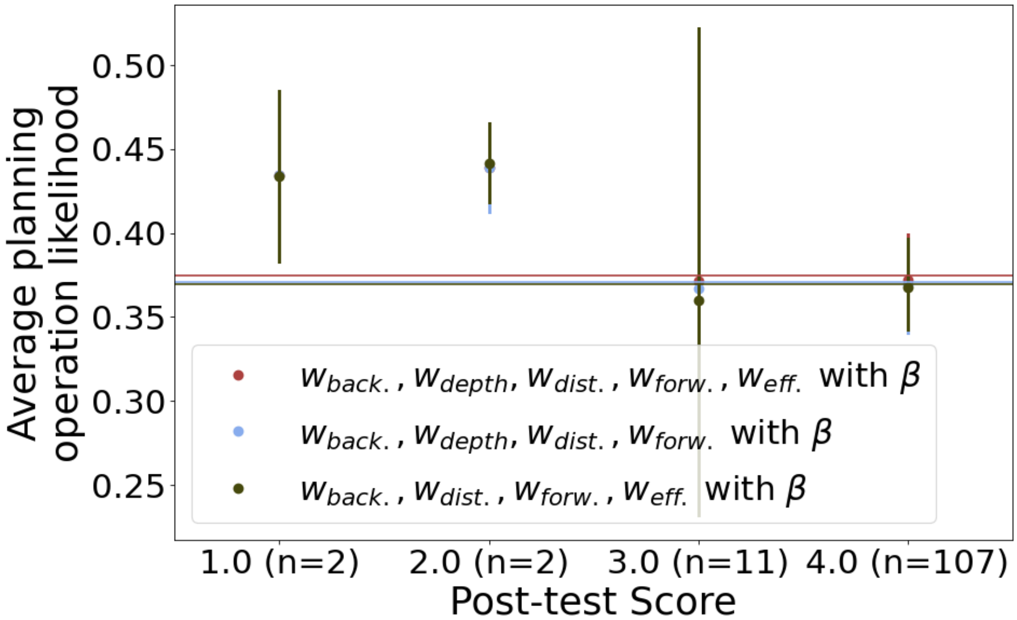


Figure A.2: Absolute fit versus post-test questionnaire.

Table A.1: Posterior-Predictive Check: Correlations for click type for the top three models by BIC.

Model Name	Click Type	Correlation	p-value	95% C. I.
$w_{back.}, w_{depth}, w_{dist.}, w_{forw.}, w_{eff.}$ with β	early	Spearman's $\rho(122) = 0.76$	$p < 0.001$	$[0.67, 0.82]$
$w_{back.}, w_{depth}, w_{dist.}, w_{forw.}, w_{eff.}$ with β	middle	Spearman's $\rho(122) = 0.65$	$p < 0.001$	$[0.54, 0.74]$
$w_{back.}, w_{depth}, w_{dist.}, w_{forw.}, w_{eff.}$ with β	late	Spearman's $\rho(122) = 0.59$	$p < 0.001$	$[0.46, 0.69]$
$w_{back.}, w_{depth}, w_{dist.}, w_{forw.}, w_{eff.}$ with β	clicks	Spearman's $\rho(122) = 0.58$	$p < 0.001$	$[0.44, 0.68]$
$w_{back.}, w_{depth}, w_{dist.}, w_{forw.}$ with β	early	Spearman's $\rho(122) = 0.75$	$p < 0.001$	$[0.67, 0.82]$
$w_{back.}, w_{depth}, w_{dist.}, w_{forw.}$ with β	middle	Spearman's $\rho(122) = 0.65$	$p < 0.001$	$[0.54, 0.75]$
$w_{back.}, w_{depth}, w_{dist.}, w_{forw.}$ with β	late	Spearman's $\rho(122) = 0.57$	$p < 0.001$	$[0.44, 0.68]$
$w_{back.}, w_{depth}, w_{dist.}, w_{forw.}$ with β	clicks	Spearman's $\rho(122) = 0.55$	$p < 0.001$	$[0.41, 0.66]$
$w_{back.}, w_{dist.}, w_{forw.}, w_{eff.}$ with β	early	Spearman's $\rho(122) = 0.76$	$p < 0.001$	$[0.68, 0.83]$
$w_{back.}, w_{dist.}, w_{forw.}, w_{eff.}$ with β	middle	Spearman's $\rho(122) = 0.66$	$p < 0.001$	$[0.55, 0.75]$
$w_{back.}, w_{dist.}, w_{forw.}, w_{eff.}$ with β	late	Spearman's $\rho(122) = 0.61$	$p < 0.001$	$[0.49, 0.71]$
$w_{back.}, w_{dist.}, w_{forw.}, w_{eff.}$ with β	clicks	Spearman's $\rho(122) = 0.69$	$p < 0.001$	$[0.59, 0.78]$

Table A.2: Average Planning Operation Likelihood per Trial for each Model.

Model	Mean	SD
$w_{back.}, w_{depth}, w_{dist.}, w_{forw.}, w_{eff.}$ with β	0.374	0.151
$w_{back.}, w_{depth}, w_{dist.}, w_{forw.}$ with β	0.371	0.153
$w_{back.}, w_{dist.}, w_{forw.}, w_{eff.}$ with β	0.369	0.155

SD = Standard Deviation

Table A.3: Average Model Parameters for Experiment 1.

Parameter Name	Mean (Standard Deviation)
$w_{back.}$	4.291 (4.539)
w_{depth}	1.160 (2.706)
$w_{dist.}$	2.139 (2.845)
$w_{forw.}$	2.799 (3.784)
$w_{eff.}$	1.508 (2.415)
β	7.580 (19.555)

Table A.4: Average Model Parameter for Participants in Decreasing Variance Environment.

Parameter Name	Mean (Standard Deviation)
$w_{back.}$	1.241 (2.377)
w_{depth}	0.767 (2.242)
$w_{dist.}$	1.353 (2.580)
$w_{forw.}$	4.388 (4.650)
$w_{eff.}$	1.284 (2.754)
β	3.328 (13.065)

Table A.5: Average Model Parameter for Participants in Constant Variance Environment.

Parameter Name	Mean (Standard Deviation)
$w_{back.}$	2.095 (3.869)
w_{depth}	0.638 (1.473)
$w_{dist.}$	0.698 (1.252)
$w_{forw.}$	3.500 (4.472)
$w_{eff.}$	2.655 (4.063)
β	2.893 (3.832)

Table A.6: Bayesian Model Selection Results for Participants in Decreasing Variance Environment.

Model	Expected number of participants best explained by the model	Model probability	Exceedance probabilities
$w_{dist.}, w_{forw.}, w_{eff.}$ with β	10.30	0.12	0.52
$w_{depth}, w_{eff.}$ with β	9.04	0.11	0.33
$w_{back.}, w_{forw.}, w_{eff.}$ with β	6.38	0.08	0.09
$w_{back.}, w_{forw.}$ with β	4.99	0.07	0.03
$w_{forw.}, w_{eff.}$ with β	3.58	0.05	0.01
w_{depth} with β	3.30	0.05	0.01
$w_{depth}, w_{forw.}, w_{eff.}$ with β	2.65	0.04	0.00
$w_{dist.}$ with β	2.16	0.03	0.00
$w_{back.}, w_{eff.}$ with β	1.79	0.03	0.00
$w_{eff.}$ with β	1.68	0.03	0.00
$w_{back.}, w_{dist.}, w_{eff.}$ with β	1.33	0.03	0.00
$w_{dist.}, w_{forw.}$ with β	1.12	0.02	0.00
$w_{back.}, w_{dist.}$ with β	1.03	0.02	0.00
$w_{dist.}, w_{eff.}$ with β	0.88	0.02	0.00
$w_{back.}, w_{dist.}, w_{forw.}, w_{eff.}$ with β	0.85	0.02	0.00
$w_{forw.}$ with β	0.79	0.02	0.00
$w_{back.}, w_{depth}$ with β	0.75	0.02	0.00
$w_{depth}, w_{forw.}$ with β	0.71	0.02	0.00
$w_{back.}, w_{dist.}, w_{forw.}$ with β	0.61	0.02	0.00
$w_{back.}, w_{depth}, w_{eff.}$ with β	0.59	0.02	0.00
$w_{back.}, w_{depth}, w_{forw.}, w_{eff.}$ with β	0.52	0.02	0.00
$w_{depth}, w_{dist.}, w_{forw.}, w_{eff.}$ with β	0.48	0.02	0.00
$w_{back.}, w_{depth}, w_{forw.}$ with β	0.39	0.02	0.00
$w_{depth}, w_{dist.}$ with β	0.37	0.02	0.00
$w_{depth}, w_{dist.}, w_{eff.}$ with β	0.35	0.01	0.00
$w_{back.}$ with β	0.28	0.01	0.00
$w_{depth}, w_{dist.}, w_{forw.}$ with β	0.24	0.01	0.00
$w_{back.}, w_{depth}, w_{dist.}, w_{eff.}$ with β	0.24	0.01	0.00
$w_{back.}, w_{depth}, w_{dist.}$ with β	0.22	0.01	0.00
$w_{back.}, w_{depth}, w_{dist.}, w_{forw.}, w_{eff.}$ with β	0.19	0.01	0.00
$w_{back.}, w_{depth}, w_{dist.}, w_{forw.}$ with β	0.14	0.01	0.00
Null (Given Costs) with β	0.04	0.01	0.00
Null (Random)	0.01	0.01	0.00

Table A.7: Bayesian Model Selection Results for Participants in Constant Variance Environment.

Model	Expected number of participants best explained by the model	Model probability	Exceedance probabilities
$w_{dist.}, w_{forw.}, w_{eff.}$ with β	10.30	0.12	0.52
$w_{depth}, w_{eff.}$ with β	9.04	0.11	0.33
$w_{back.}, w_{forw.}, w_{eff.}$ with β	6.38	0.08	0.09
$w_{back.}, w_{forw.}$ with β	4.99	0.07	0.03
$w_{forw.}, w_{eff.}$ with β	3.58	0.05	0.01
w_{depth} with β	3.30	0.05	0.01
$w_{depth}, w_{forw.}, w_{eff.}$ with β	2.65	0.04	0.00
$w_{dist.}$ with β	2.16	0.03	0.00
$w_{back.}, w_{eff.}$ with β	1.79	0.03	0.00
$w_{eff.}$ with β	1.68	0.03	0.00
$w_{back.}, w_{dist.}, w_{eff.}$ with β	1.33	0.03	0.00
$w_{dist.}, w_{forw.}$ with β	1.12	0.02	0.00
$w_{back.}, w_{dist.}$ with β	1.03	0.02	0.00
$w_{dist.}, w_{eff.}$ with β	0.88	0.02	0.00
$w_{back.}, w_{dist.}, w_{forw.}, w_{eff.}$ with β	0.85	0.02	0.00
$w_{forw.}$ with β	0.79	0.02	0.00
$w_{back.}, w_{depth}$ with β	0.75	0.02	0.00
$w_{depth}, w_{forw.}$ with β	0.71	0.02	0.00
$w_{back.}, w_{dist.}, w_{forw.}$ with β	0.61	0.02	0.00
$w_{back.}, w_{depth}, w_{eff.}$ with β	0.59	0.02	0.00
$w_{back.}, w_{depth}, w_{forw.}, w_{eff.}$ with β	0.52	0.02	0.00
$w_{depth}, w_{dist.}, w_{forw.}, w_{eff.}$ with β	0.48	0.02	0.00
$w_{back.}, w_{depth}, w_{forw.}$ with β	0.39	0.02	0.00
$w_{depth}, w_{dist.}$ with β	0.37	0.02	0.00
$w_{depth}, w_{dist.}, w_{eff.}$ with β	0.35	0.01	0.00
$w_{back.}$ with β	0.28	0.01	0.00
$w_{depth}, w_{dist.}, w_{forw.}$ with β	0.24	0.01	0.00
$w_{back.}, w_{depth}, w_{dist.}, w_{eff.}$ with β	0.24	0.01	0.00
$w_{back.}, w_{depth}, w_{dist.}$ with β	0.22	0.01	0.00
$w_{back.}, w_{depth}, w_{dist.}, w_{forw.}, w_{eff.}$ with β	0.19	0.01	0.00
$w_{back.}, w_{depth}, w_{dist.}, w_{forw.}$ with β	0.14	0.01	0.00
Null (Given Costs) with β	0.04	0.01	0.00
Null (Random)	0.01	0.01	0.00

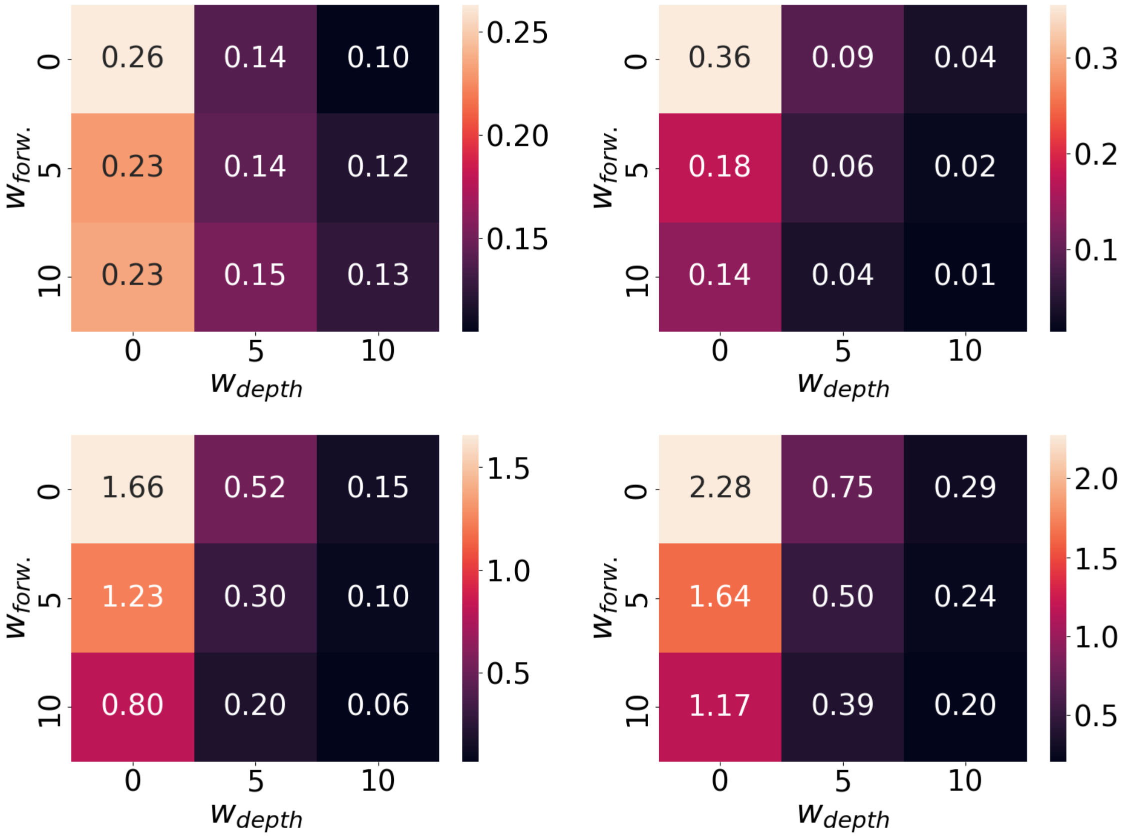


Figure A.3: Average number of clicks per level, simulated for different combinations of the depth cost weight and the non-forward search cost weight.

A.4 Chapter 4

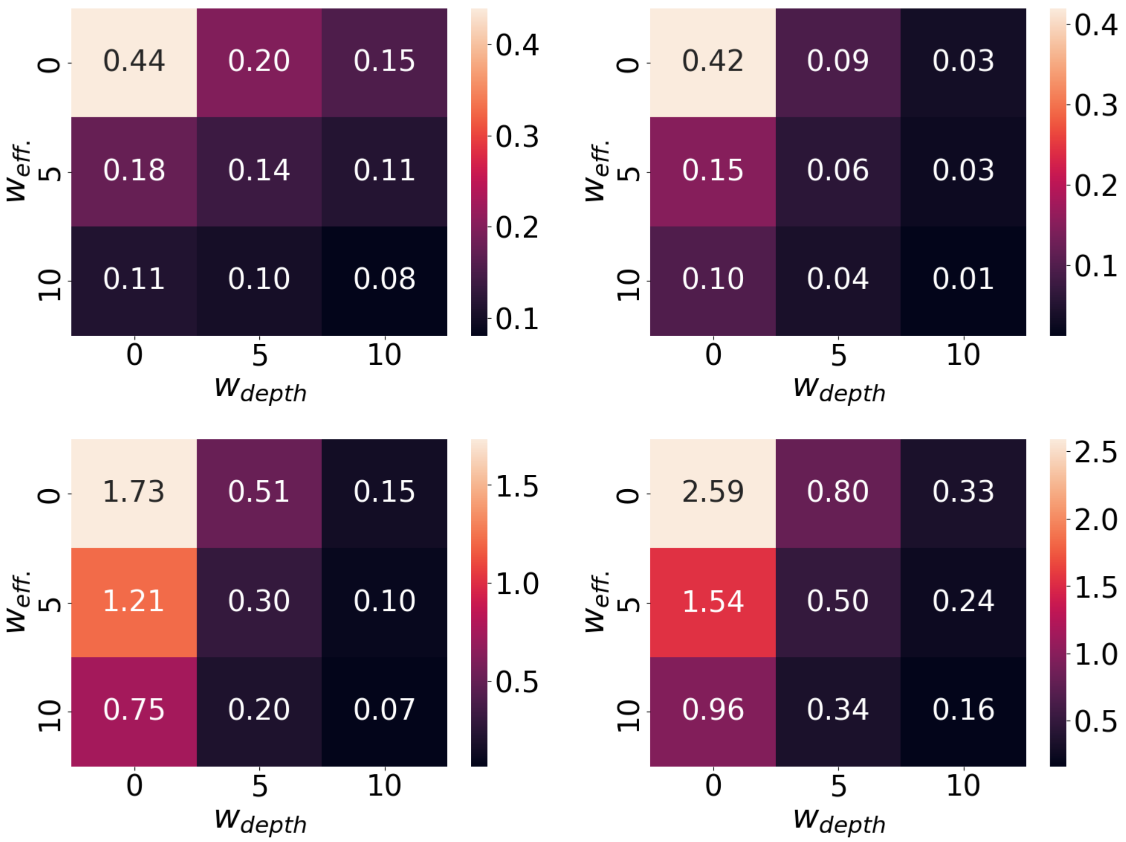


Figure A.4: Average number of clicks per level, simulated for different combinations of the depth cost weight and the effort cost weight.

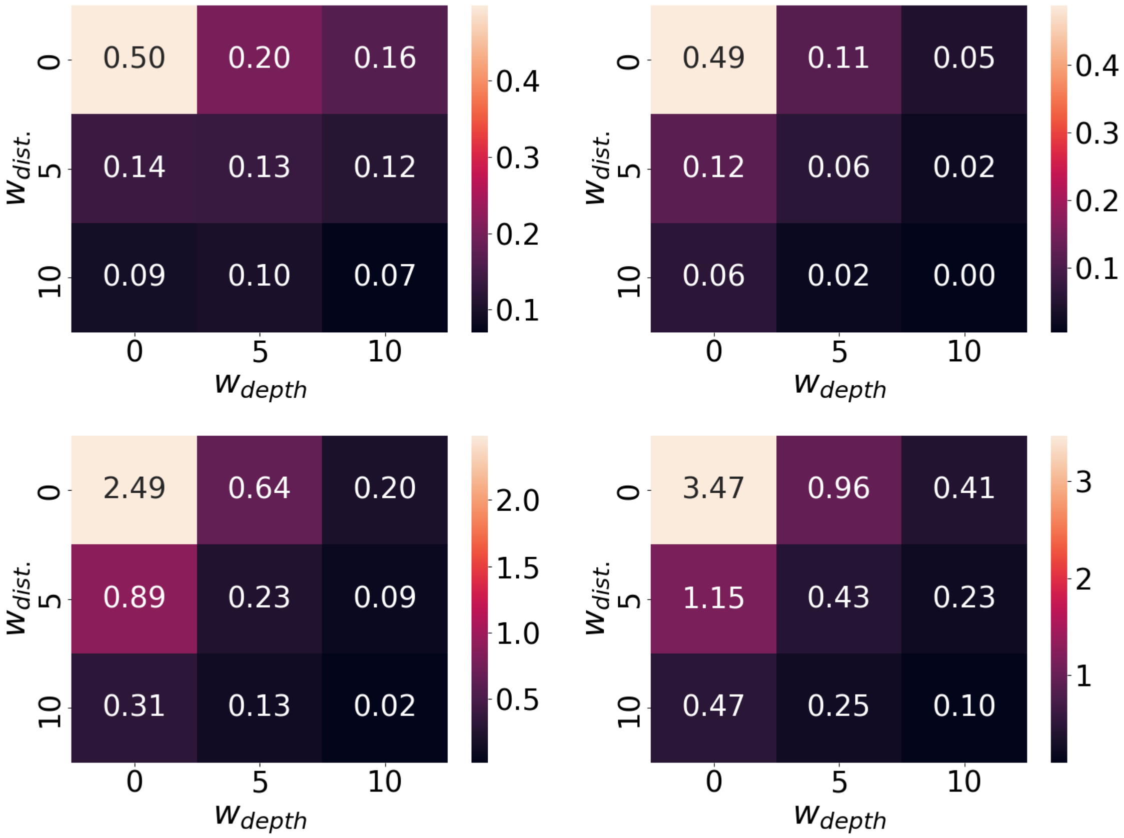


Figure A.5: Average number of clicks per level, simulated for different combinations of the depth cost weight and the distance cost weight.

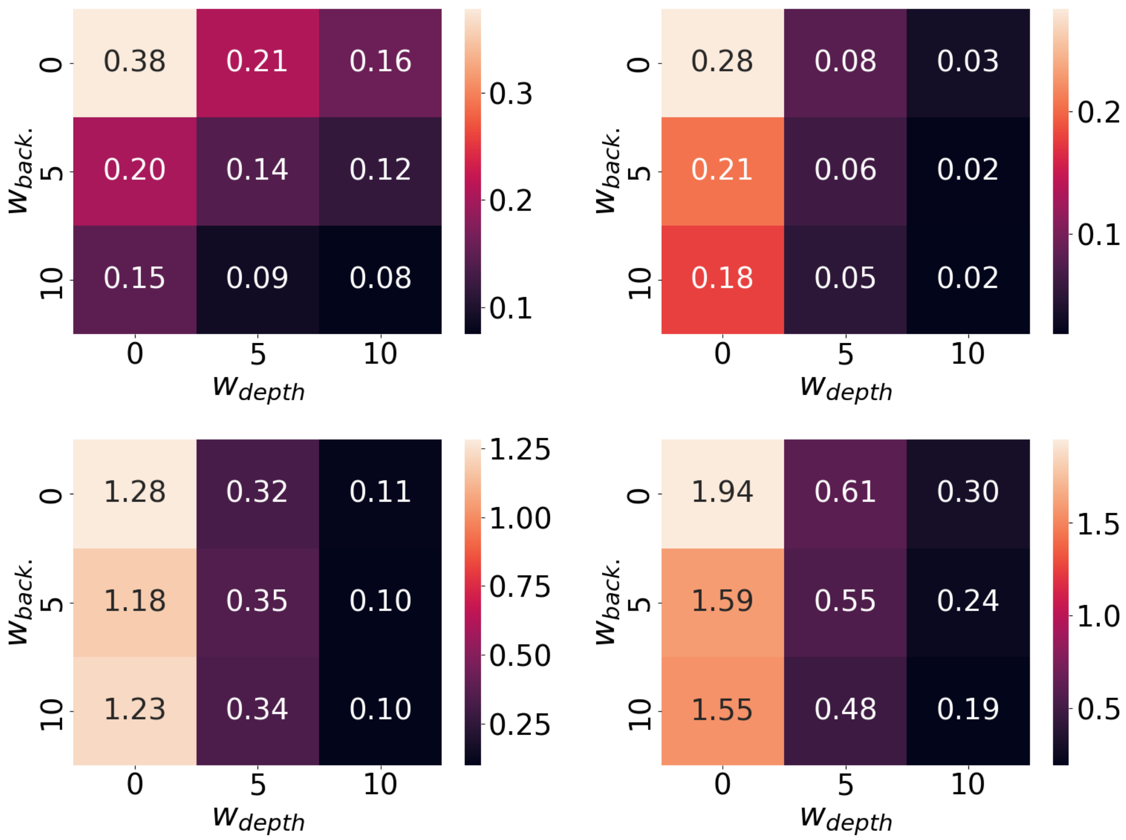


Figure A.6: Average number of clicks per level, simulated for different combinations of the depth cost weight and the non-backward search cost weight.

A.4.1 Both blocks in Experiment 2 are fit just as well as the test block in Experiment 1

To see how well the model fit participants' behavior in each block, I looked at the difference in the temperature parameter estimates between the test versus the baseline block. The temperature parameter MLE estimates were significantly lower for the test block ($\mathrm{M} = 2.41, \mathrm{SD} = 6.38$) than for the baseline block ($\mathrm{M} = 5.08, \mathrm{SD} = 12.45$; Wilcoxon signed-rank test $W = 2872.00$, $\mathrm{RBC} = -0.60$, $p < 0.001$, two-sided). Since the temperature parameter can be seen as an index of model fit, this suggests that participants were better fit by the model in the test block, when costs were higher.

Next, I tested the MLE estimates in each block against the estimates in Experiment 1 to see if the model fits participants in this experiment just as well. I found no significant difference between MLE estimates of the temperature parameter in Experiment 1 ($\mathrm{M} = 7.58, \mathrm{SD} = 19.64$) and the estimates in the baseline block (Mann-Whitney $U = 19835.00$, $\mathrm{RBC} = 0.03$, adj.$p = 0.672$, two-sided). I did find a significant difference between these MLE estimates and the estimates in the test block (Mann-Whitney $U = 14303.50$, $\mathrm{RBC} = 0.30$, adj.$p < 0.001$, two-sided). This is consistent with the interpretation that participant behavior is fit better with higher cost weights than when there are no imposed costs, such as in the baseline block and in Experiment 1.

A.4.2 The order of the baseline block has little effect on behavior or inferred parameter values

Recall that I counter-balanced the baseline and test blocks to correct for two possible confounds: fatigue and increased task knowledge. To test if these order effects were present, I compared behavior and MLE estimates between participants in each condition. I tested whether either the baseline or test MLE values were significantly different for the participants who were shown the baseline block first versus those who were shown it last. I corrected for multiple comparisons using a Benjamini-Hochberg correction with a false discovery rate of 0.05 Benjamini and Hochberg (1995).

I found that whether the test block was presented first or second had no significant effect on the MLE parameter values, in either block. Please see Table A.8 for details on Mann-Whitney tests for each parameter. This suggests that the participants who saw the baseline block first did not gain any new knowledge from the costs being temporarily reduced.

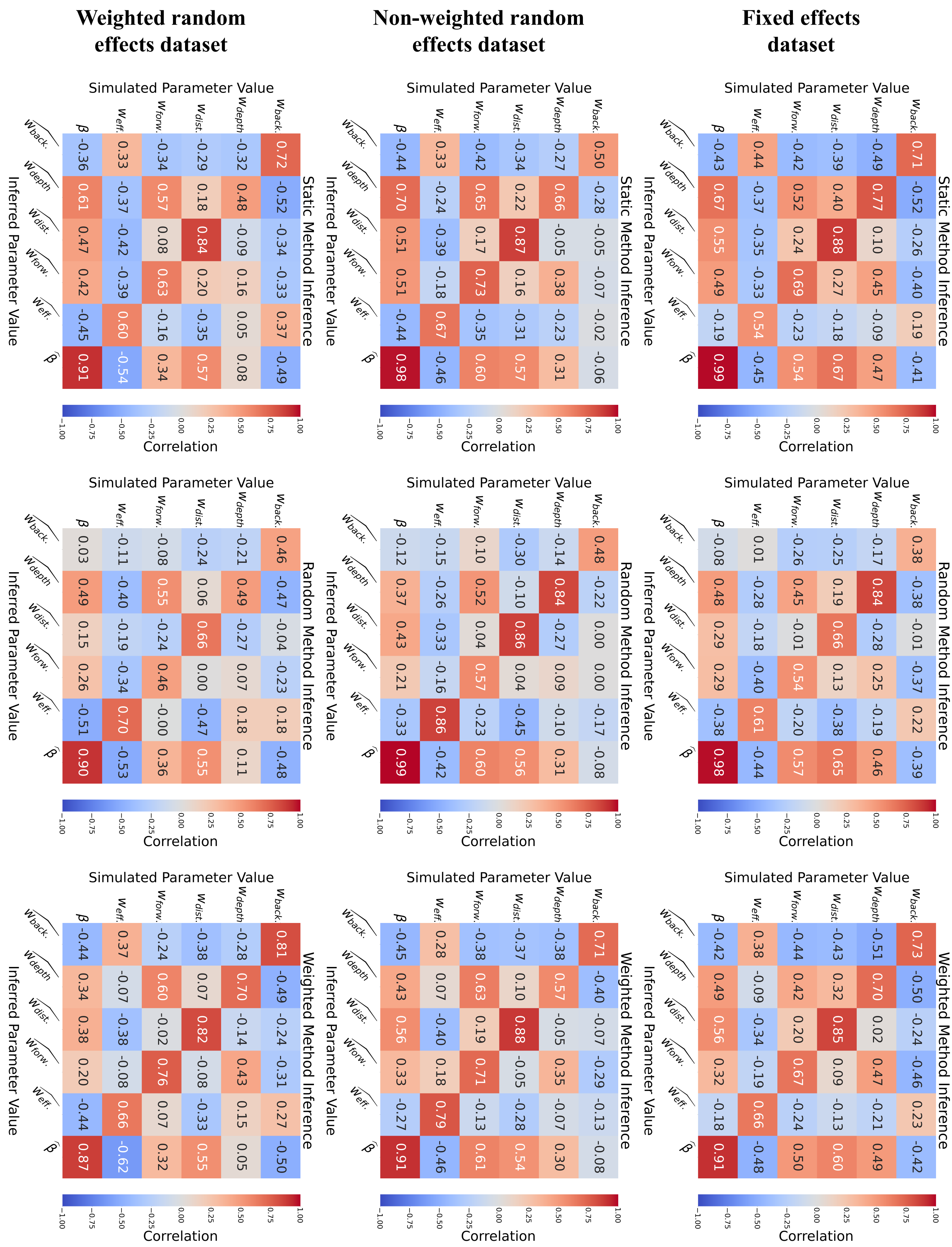


Figure A.7: Parameter recovery under all three methods, for datasets simulated under all three methods.

A.5 Tables for Chapter 4

Table A.8: Experiment 2: Effect on baseline block ordering on MLEs, per block.

Block	Cost parameter	Statistic
baseline	$w_{forw.}$	Mann-Whitney $U = 14832.00$, RBC $= -0.06$, adj.$p = 0.678$
test	$w_{forw.}$	Mann-Whitney $U = 14950.50$, RBC $= -0.07$, adj.$p = 0.677$
baseline	w_{depth}	Mann-Whitney $U = 15109.00$, RBC $= -0.08$, adj.$p = 0.677$
test	w_{depth}	Mann-Whitney $U = 14418.00$, RBC $= -0.03$, adj.$p = 0.712$
baseline	$w_{eff.}$	Mann-Whitney $U = 13485.00$, RBC $= 0.04$, adj.$p = 0.712$
test	$w_{eff.}$	Mann-Whitney $U = 14542.50$, RBC $= -0.04$, adj.$p = 0.712$
baseline	β	Mann-Whitney $U = 13504.00$, RBC $= 0.04$, adj.$p = 0.712$
test	β	Mann-Whitney $U = 14329.00$, RBC $= -0.02$, adj.$p = 0.712$
baseline	$w_{dist.}$	Mann-Whitney $U = 13707.50$, RBC $= 0.02$, adj.$p = 0.712$
test	$w_{dist.}$	Mann-Whitney $U = 12985.50$, RBC $= 0.07$, adj.$p = 0.677$
baseline	$w_{back.}$	Mann-Whitney $U = 12885.50$, RBC $= 0.08$, adj.$p = 0.677$
test	$w_{back.}$	Mann-Whitney $U = 14872.00$, RBC $= -0.06$, adj.$p = 0.677$

Table A.9: Experiment 2: Effect on baseline block ordering on MLEs, per block.

Parameter	Test Block Mean (SD)	Baseline Block Mean (SD)	W	RBC
$w_{back.}$	8.88 (2.46)	6.68 (3.58) 4021.50	-0.69	$p < 0.001$
w_{depth}	4.79 (3.70)	1.69 (2.77) 1444.50	-0.91	$p < 0.001$
$w_{dist.}$	4.29 (4.35)	2.39 (3.34) 3419.00	-0.68	$p < 0.001$
$w_{forw.}$	7.43 (3.25)	3.98 (3.19) 1398.00	-0.91	$p < 0.001$
$w_{eff.}$	7.59 (3.07)	4.09 (3.49) 846.00	-0.94	$p < 0.001$
β	2.12 (7.46)	1.56 (9.90) 2866.00	-0.62	$p < 0.001$

Table A.10: Simulations: HDI Statistics for Each Cost Parameter.

Cost parameter	Statistics
$w_{back.}$	6.64 (3.04)
w_{depth}	2.26 (2.73)
$w_{dist.}$	2.01 (3.14)
$w_{forw.}$	4.61 (3.35)
$w_{eff.}$	4.45 (3.23)
β	0.80 (8.15)

Table A.11: Simulations: Correlation Between HDIs and True Parameter Values.

Cost parameter	Correlation	adj. p-value	95% C. I.
$w_{eff.}$	Spearman's $\rho(122) = 0.12$	adj.$p = 0.193$	$[-0.06, 0.29]$
w_{depth}	Spearman's $\rho(122) = 0.56$	adj.$p < 0.001$	$[0.43, 0.67]$
$w_{back.}$	Spearman's $\rho(122) = 0.34$	adj.$p < 0.001$	$[0.18, 0.49]$
$w_{forw.}$	Spearman's $\rho(122) = 0.43$	adj.$p < 0.001$	$[0.28, 0.57]$
$w_{dist.}$	Spearman's $\rho(122) = 0.74$	adj.$p < 0.001$	$[0.64, 0.81]$
β	Spearman's $\rho(122) = -0.32$	adj.$p < 0.001$	$[-0.47, -0.15]$

Table A.12: Simulations: Correlation Between HDIs and Absolute MLE Error.

Cost parameter	Correlation	adj. p-value	95% C. I.
$w_{eff.}$	Spearman's $\rho(122) = 0.19$	adj.$p = 0.041$	$[0.01, 0.36]$
w_{depth}	Spearman's $\rho(122) = 0.55$	adj.$p < 0.001$	$[0.41, 0.66]$
$w_{back.}$	Spearman's $\rho(122) = 0.13$	adj.$p = 0.155$	$[-0.05, 0.30]$
$w_{forw.}$	Spearman's $\rho(122) = 0.20$	adj.$p = 0.041$	$[0.02, 0.36]$
$w_{dist.}$	Spearman's $\rho(122) = 0.66$	adj.$p < 0.001$	$[0.55, 0.75]$
β	Spearman's $\rho(122) = 0.45$	adj.$p < 0.001$	$[0.30, 0.58]$

Table A.13: Experiment 2: Correlation between HDIs and MLE estimates in test block.

Cost parameter	Correlation	adj. p-value	95% C. I.
β	Spearman's $\rho(335) = -0.30$	adj.$p < 0.001$	$[-0.40, -0.20]$
$w_{back.}$	Spearman's $\rho(335) = -0.06$	adj.$p = 0.311$	$[-0.16, 0.05]$
$w_{forw.}$	Spearman's $\rho(335) = -0.08$	adj.$p = 0.200$	$[-0.18, 0.03]$
$w_{eff.}$	Spearman's $\rho(335) = 0.25$	adj.$p < 0.001$	$[0.15, 0.35]$
$w_{dist.}$	Spearman's $\rho(335) = 0.68$	adj.$p < 0.001$	$[0.62, 0.73]$
w_{depth}	Spearman's $\rho(335) = 0.21$	adj.$p < 0.001$	$[0.11, 0.31]$

Table A.14: Experiment 2: Correlation between HDIs and inferred temperature values in test block.

Cost parameter	Correlation	adj. p-value	95% C. I.
β	Spearman's $\rho(335) = -0.30$	adj.$p < 0.001$	$[-0.40, -0.20]$
$w_{back.}$	Spearman's $\rho(335) = -0.58$	adj.$p < 0.001$	$[-0.65, -0.51]$
$w_{forw.}$	Spearman's $\rho(335) = -0.18$	adj.$p < 0.001$	$[-0.28, -0.08]$
$w_{eff.}$	Spearman's $\rho(335) = -0.21$	adj.$p < 0.001$	$[-0.31, -0.10]$
$w_{dist.}$	Spearman's $\rho(335) = 0.02$	adj.$p = 0.663$	$[-0.08, 0.13]$
w_{depth}	Spearman's $\rho(335) = -0.21$	adj.$p < 0.001$	$[-0.31, -0.10]$

A.6 Chapter 5

A.6.1 Hypotheses

These were the preregistered hypotheses. Please see Felso and Lieder (2023) for more details on the preregistration.

- H1: Psychiatric measures
  - H1.1:
    a) "Compulsive Behavior and Intrusive Thought" factor scores are associated with the inferred model parameters.
    b) "Compulsive Behavior and Intrusive Thought" factor scores are associated with planning depth costs (undirected).
  - H1.2: "Anxious Depression" factor scores are associated with the inferred model parameters.
  - H1.3: "Social Withdrawal" factor scores are associated with the inferred model parameters.
  - H1.4: Trait anxiety scores are associated with the inferred model parameters.
- H2: Planning and future thinking measures
  - H2.1:
    a) Consideration of future consequences scores are associated with the inferred model parameters.
    b) Consideration of future consequences scores are associated with lower planning depth costs (directed).
  - H2.2:
    a) Propensity to plan is associated with the inferred model parameters.
    b) Propensity to plan is associated with lower planning depth costs (directed).
  - H2.3:
    a) Future orientation is associated with the inferred model parameters.
    b) Future orientation is associated with lower planning depth costs (directed).
- H3: Life regrets and life satisfaction measures
  - H3.1:
    a) Life regrets are associated with the inferred model parameters.

b) Life regrets are associated with higher planning depth costs (directed).
c) The effect of depth cost on life regrets is moderated by propensity to feel regret (directed).

- H3.2:
  a) Life satisfaction is associated with the inferred model parameters.
  b) Life satisfaction is associated with lower planning depth costs (directed).
- H3.3:
  a) Thriving is associated with the inferred model parameters.
  b) Thriving is associated with lower planning depth costs (directed).

- H4: Impulsivity and risk-taking measures
  - H4.1:
    a) Risk-taking despite high perceived risk and low perceived risk benefit (i.e., irrational risk taking) is associated with the inferred model parameters.
    b) Risk-taking despite high perceived risk and low perceived risk benefit (i.e., irrational risk taking) is associated with higher planning depth costs (directed).
  - H4.2: Lack of pre-meditation is associated with the inferred model parameters.
- H5: Intolerance of uncertainty measure
  - H5.1: Intolerance of uncertainty is associated with the inferred model parameters.
- H6: Distress intolerance measure
  - H6.1: Distress intolerance is associated with the inferred model parameters.
- H7: Cognitive styles (rationality)
  - H7.1:
    a) The ability to override intuitive cognitive styles (correct responses in the Cognitive Reflection Test) is associated with the inferred model parameters.
    b) The ability to override intuitive cognitive styles (correct responses in the Cognitive Reflection Test) is associated with lower planning depth costs (directed).

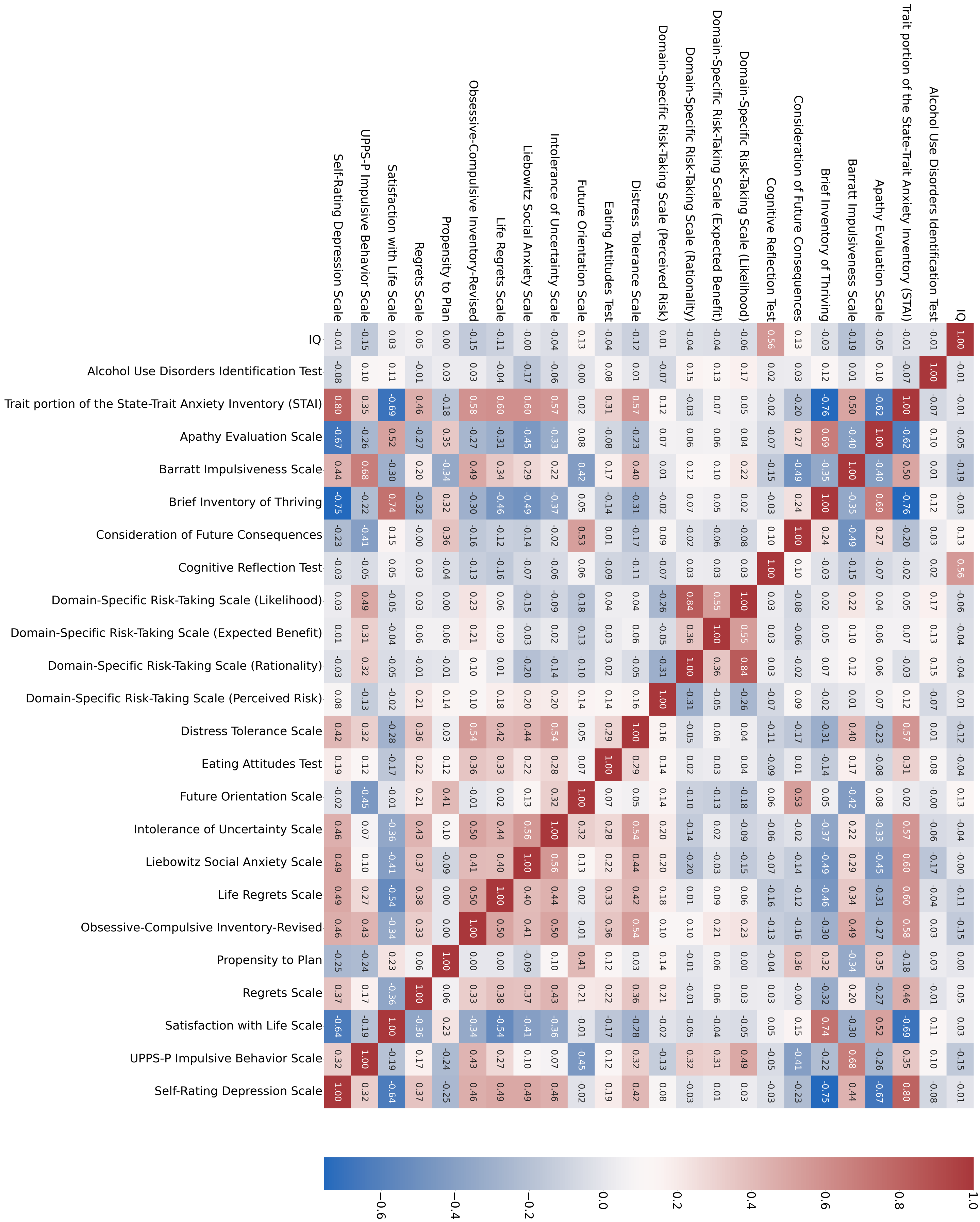


Figure A.8: Correlation between questionnaire measures.

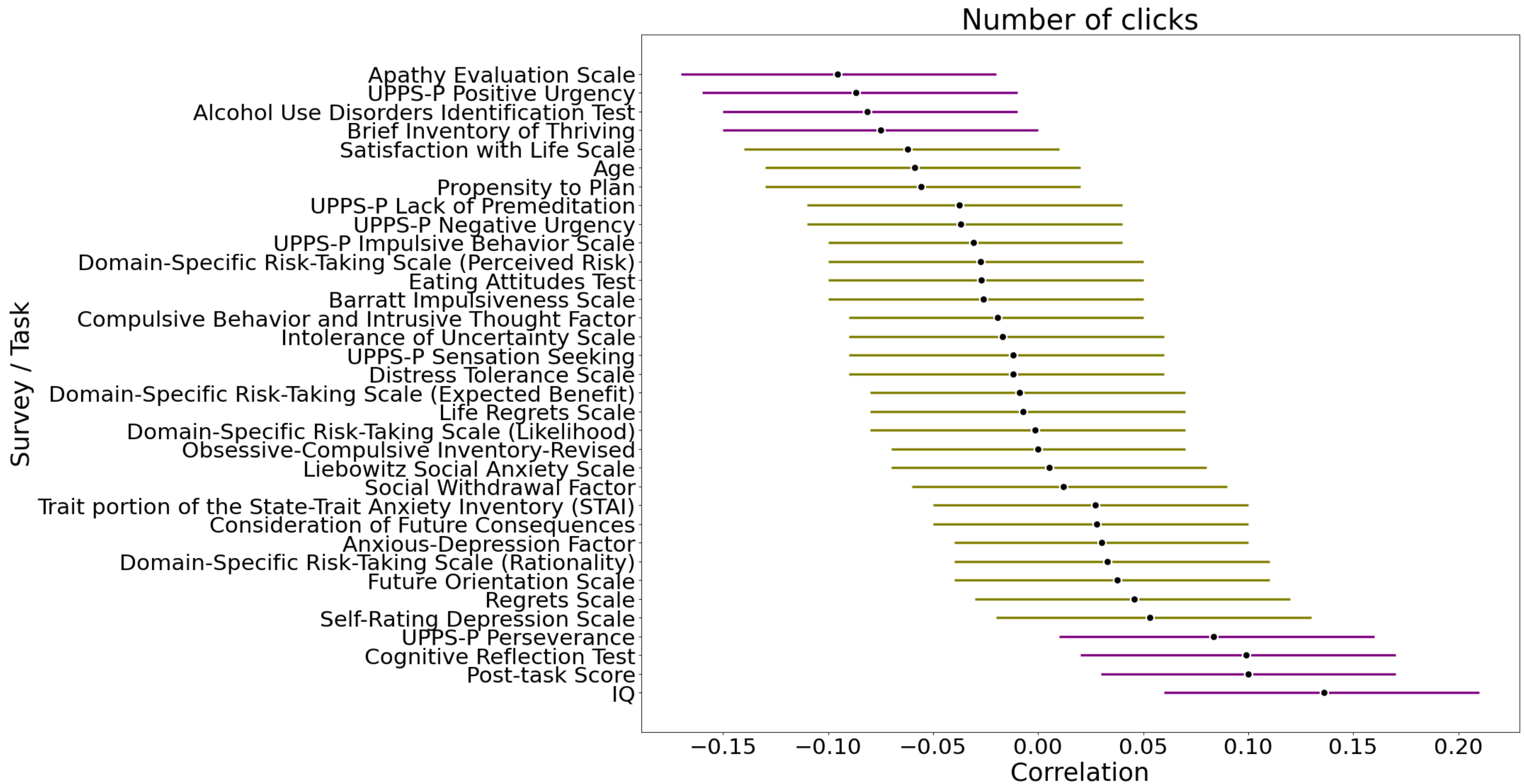


Figure A.9: Spearman's rank correlation coefficient between questionnaires and total number of clicks. Bars denote 95% confidence interval. Purple bars denote p-values below 0.05.

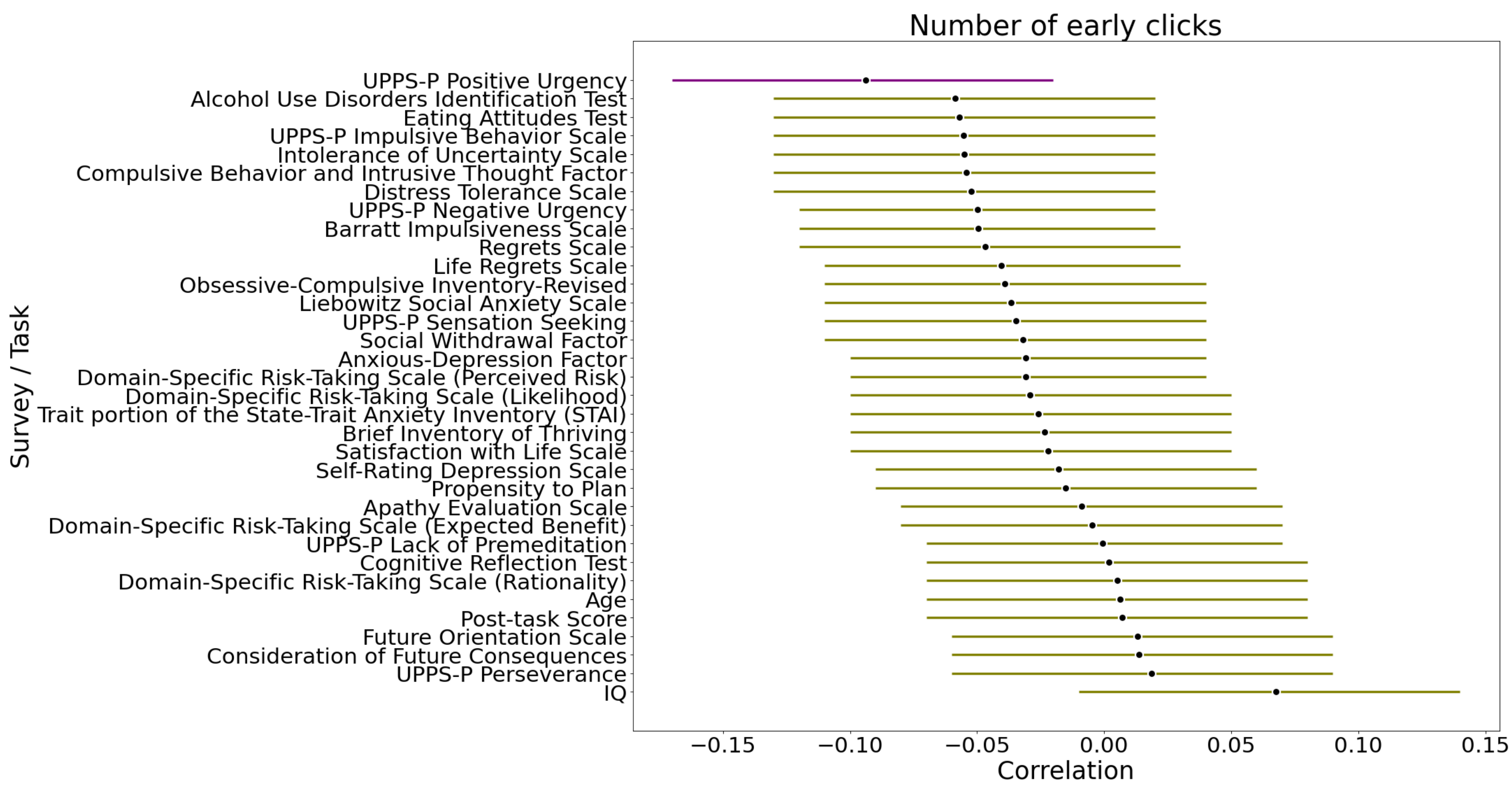


Figure A.10: Spearman’s rank correlation coefficient between questionnaires and number of early clicks. Bars denote 95% confidence interval. Purple bars denote p-values below 0.05.

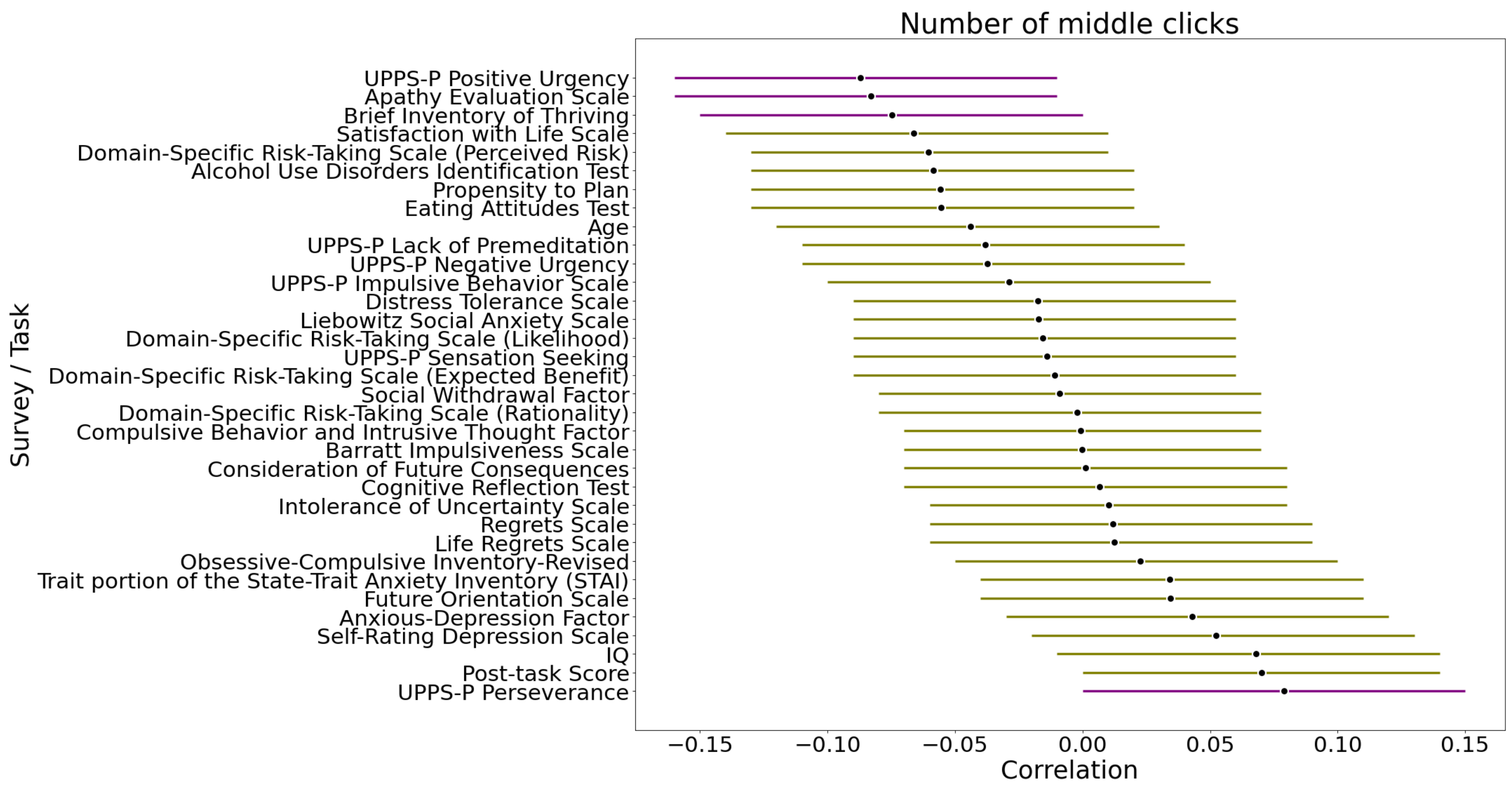


Figure A.11: Spearman's rank correlation coefficient between questionnaires and number of middle clicks. Bars denote 95% confidence interval. Purple bars denote p-values below 0.05.

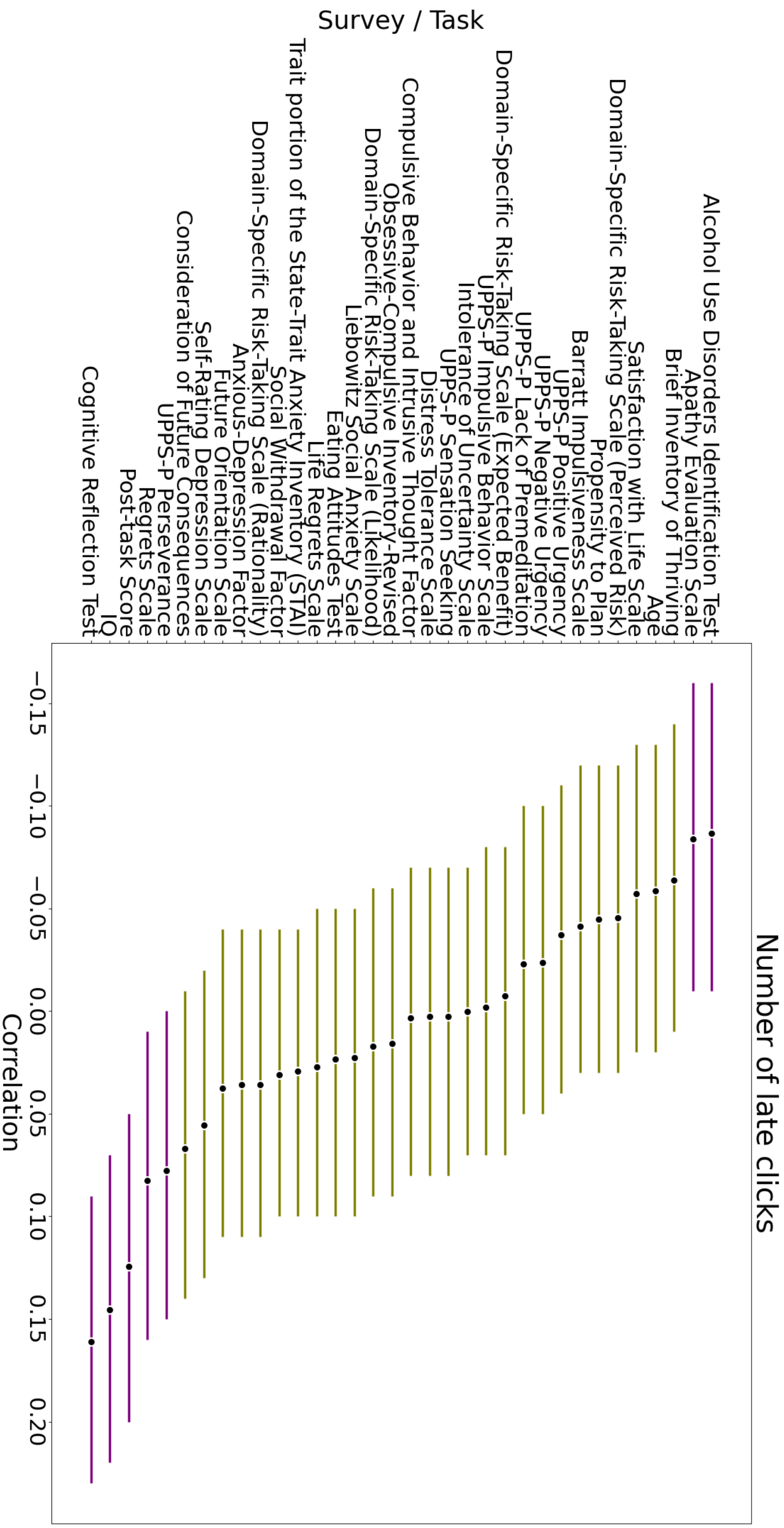


Figure A.12: Spearman's rank correlation coefficient between questionnaires and number of late clicks. Bars denote 95% confidence interval. Purple bars denote p-values below 0.05.

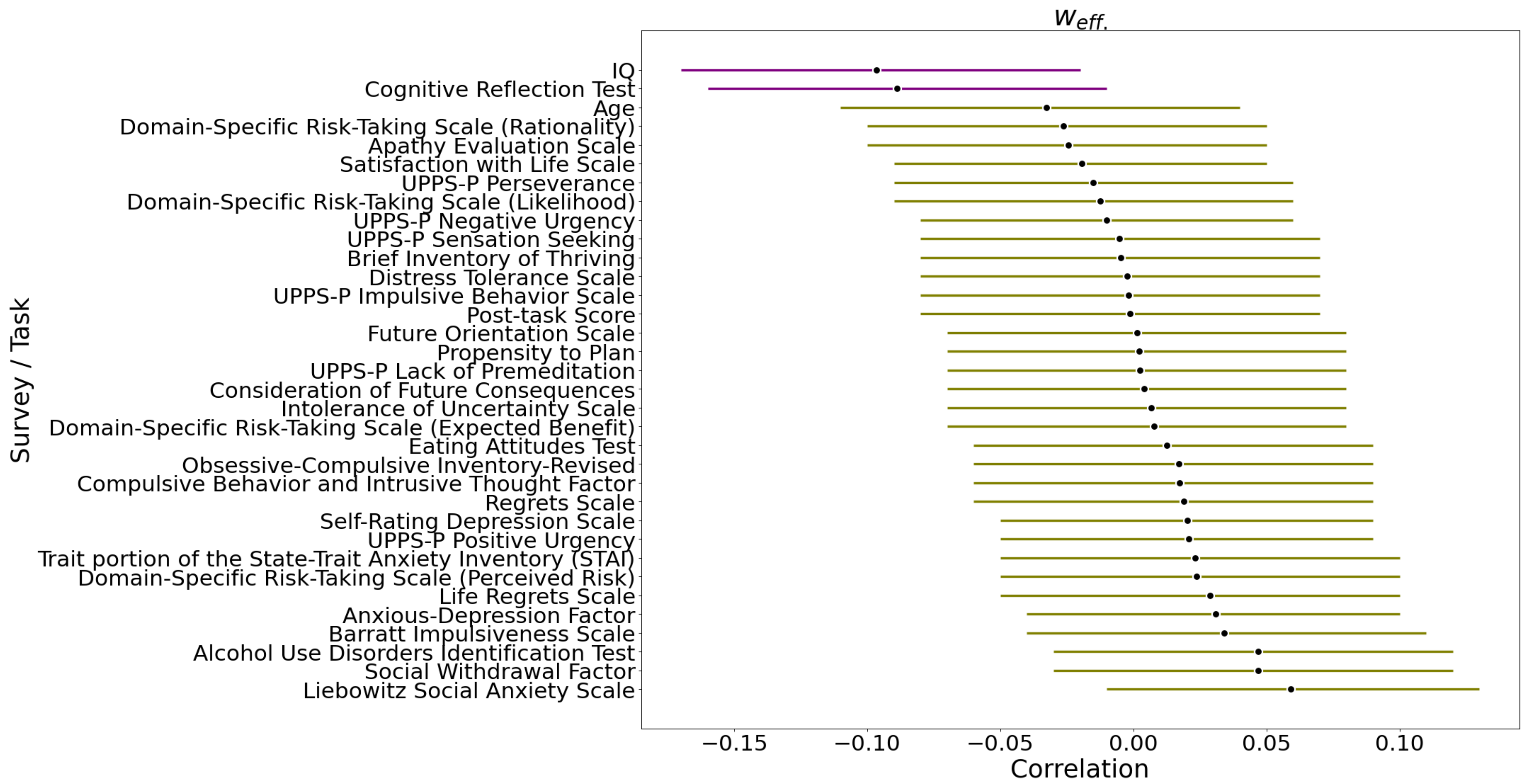


Figure A.13: Spearman's rank correlation coefficient between questionnaires and inferred effort cost weight. Bars denote 95% confidence interval. Purple bars denote p-values below 0.05.

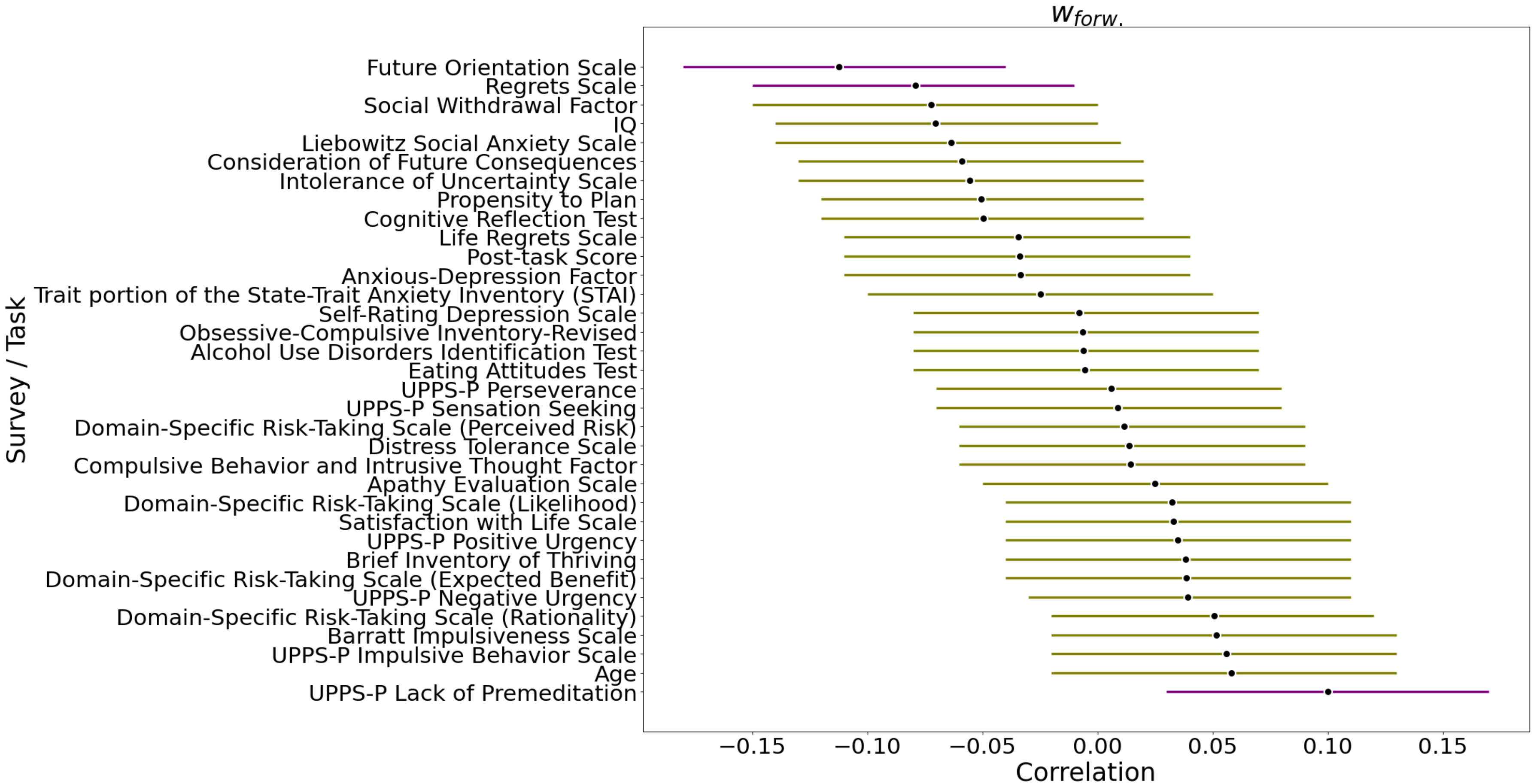


Figure A.14: Spearman's rank correlation coefficient between questionnaires and inferred non-forward search cost weight. Bars denote 95% confidence interval. Purple bars denote p-values below 0.05.

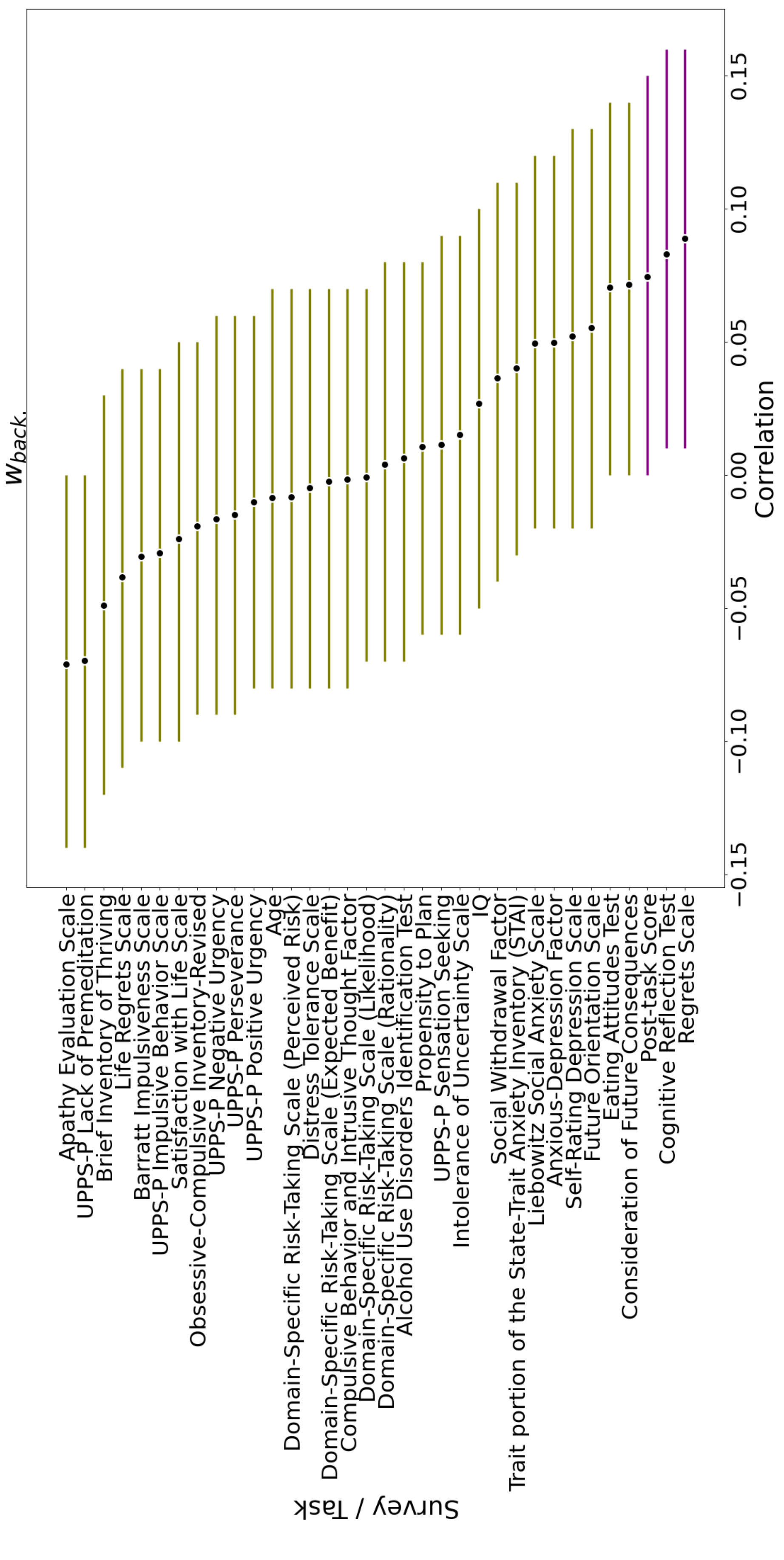


Figure A.15: Spearman's rank correlation coefficient between questionnaires and inferred non-backward search cost weight. Bars denote 95% confidence interval. Purple bars denote p-values below 0.05.

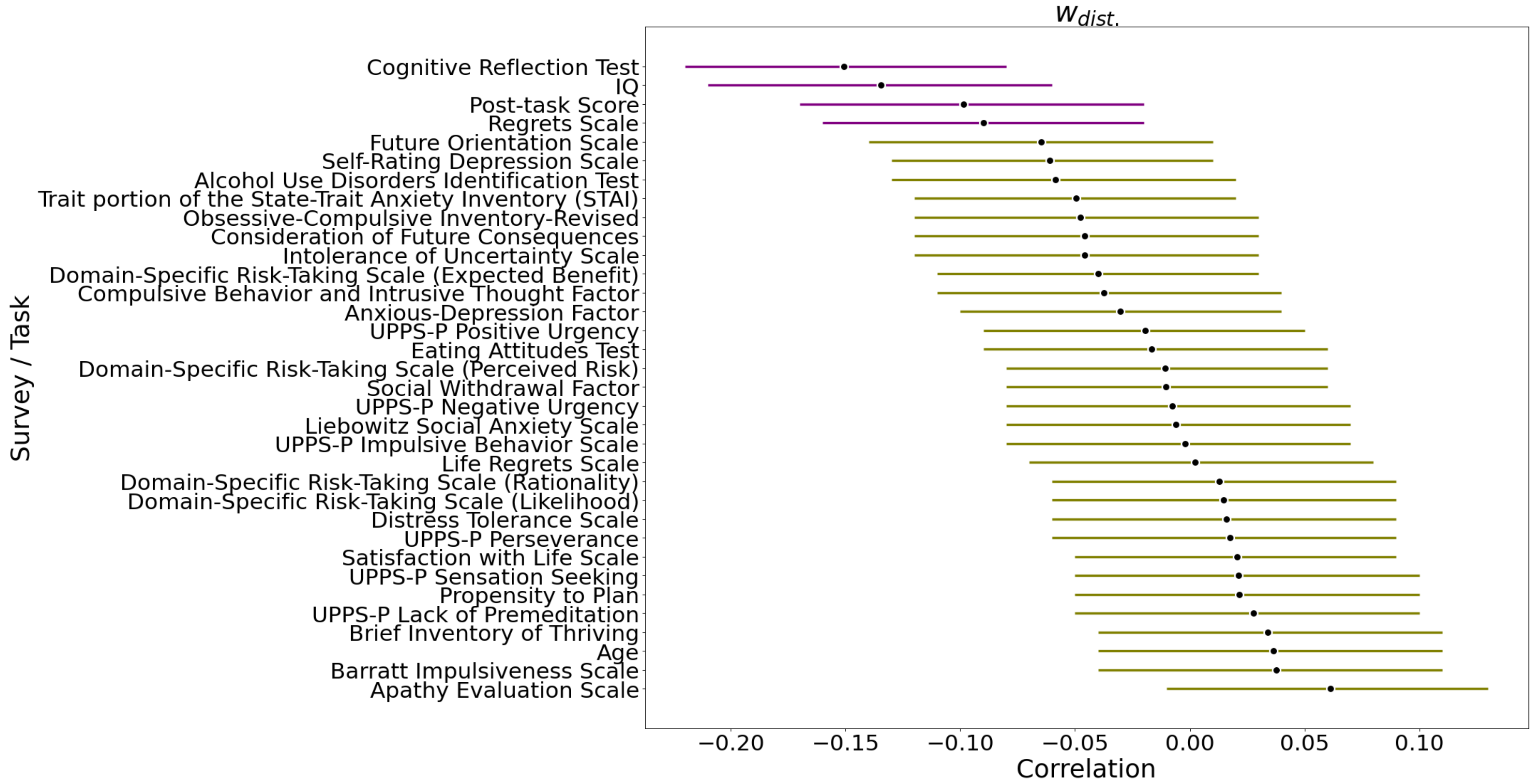


Figure A.16: Spearman's rank correlation coefficient between questionnaires and inferred distance cost weight. Bars denote 95% confidence interval. Purple bars denote p-values below 0.05.

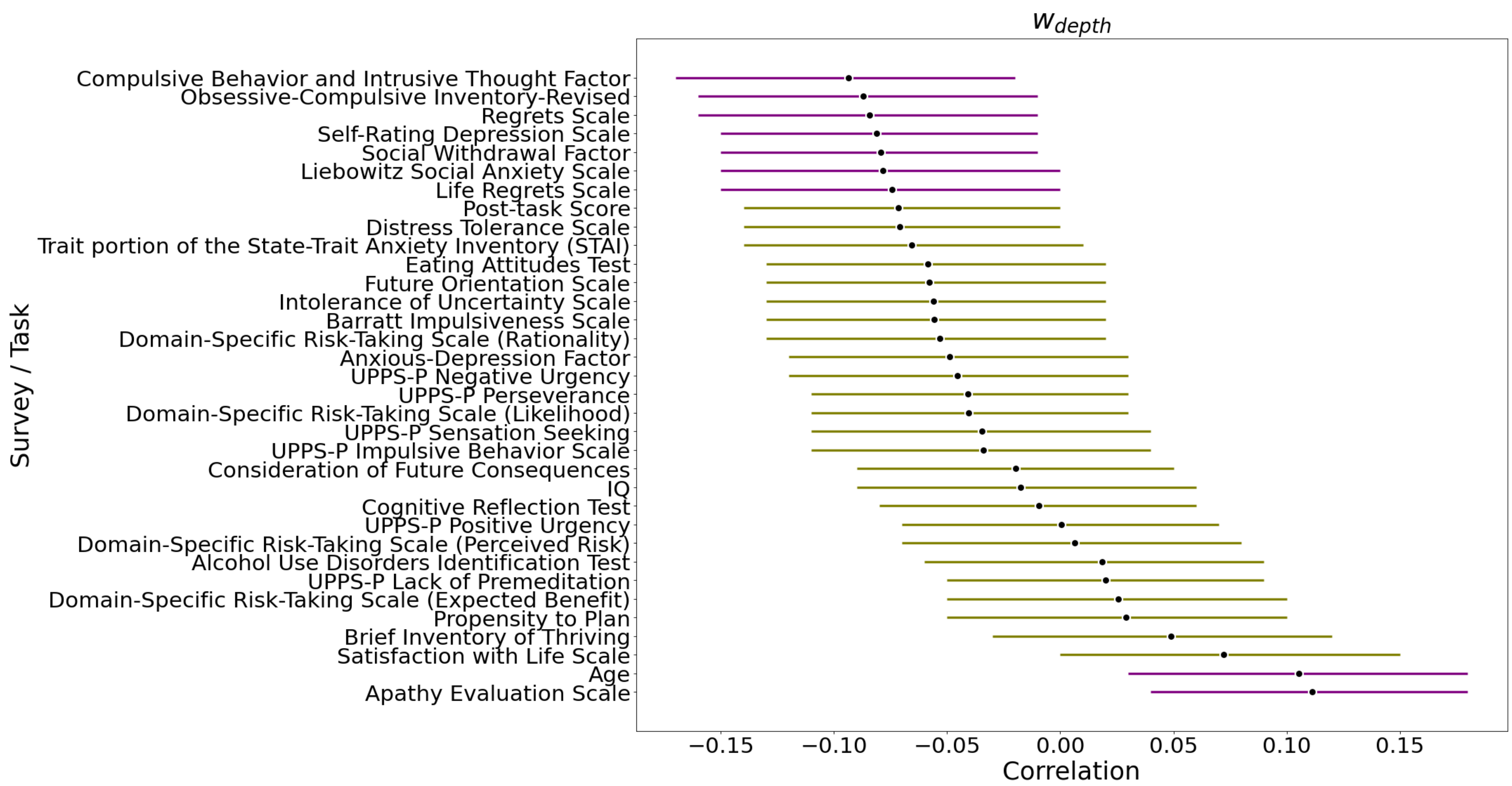


Figure A.17: Spearman's rank correlation coefficient between questionnaires and inferred depth cost weight. Bars denote 95% confidence interval. Purple bars denote p-values below 0.05.

Symbols

B	Set of belief-states in a meta-level MDP
C	Set of actions, or cognitive operations, in a meta-level MDP
R	Set of rewards (or, interchangeably, costs) in a meta-level MDP
T	Transition function in a Meta-level Markov decision process

Abbreviations

MDP	Markov decision process
BIC	Bayesian information criterion
BMS	Bayesian model selection
HDI	highest posterior density interval
MAP	maximum a posteriori probability estimate/estimation
MDP	Markov decision process
MLE	maximum-likelihood estimate/estimation
AUDIT	Alcohol Use Disorders Identification Test
AES	Apathy Evaluation Scale
BIS-10	Barratt Impulsiveness Scale
EAT-26	Eating Attitudes Test
LSAS	Liebowitz Social Anxiety Scale
OCI-R	Obsessive-Compulsive Inventory-Revised
SDS	Self-Rating Depression Scale
STAI	Trait portion of the State-Trait Anxiety Inventory
CFC	Consideration of Future Consequences
FOS	Future Orientation Scale
PTP	Propensity to Plan
BIT	Brief Inventory of Thriving
SWLS	Satisfaction with Life Scale
UPPS-P	Urgency, Premeditation (lack of), Perseverance (lack of), Sensation Seeking, Positive Urgency, Impulsive Behavior Scale
DOSPERT	Domain-Specific Risk-Taking
IUS-12	Intolerance of Uncertainty Scale
DTS	Distress Tolerance Scale
CRT	Cognitive Reflection Test
ICAR-16	International Cognitive Ability Resource

List of Tables

3.1 Bayesian Model Selection Results for Experiment 1. 34

4.1 Variance explained in true parameter value, for reported mixed effects analysis. 39

5.1 Mann-Whitney tests for parameters MLEs between Experiment 1 and Experiment 3. All tests are two-sided. 66
5.2 Mann-Whitney tests for parameters HDI spreads between Experiment 1 and Experiment 3. All tests are two-sided. 67
5.3 Average Model Parameters for Experiment 3. 67
5.4 Variance explained in dependent variable by the model parameters. 68
5.5 Variance explained in dependent variable by behavioral measures. 71
5.6 Variance explained in dependent variable by most-frequent strategy. . . . 72
5.7 Variance Inflation Factor for Independent Variables. 75

A.1 Posterior-Predictive Check: Correlations for click type for the top three models by BIC. 91
A.2 Average Planning Operation Likelihood per Trial for each Model. 92
A.3 Average Model Parameters for Experiment 1. 92
A.4 Average Model Parameter for Participants in Decreasing Variance Environment. 92
A.5 Average Model Parameter for Participants in Constant Variance Environment. 92
A.6 Bayesian Model Selection Results for Participants in Decreasing Variance Environment. 93
A.7 Bayesian Model Selection Results for Participants in Constant Variance Environment. 94
A.8 Experiment 2: Effect on baseline block ordering on MLEs, per block. . . 101
A.9 Experiment 2: Effect on baseline block ordering on MLEs, per block. . . 101
A.10 Simulations: HDI Statistics for Each Cost Parameter. 102
A.11 Simulations: Correlation Between HDIs and True Parameter Values. . . . 102

A.12 Simulations: Correlation Between HDIs and Absolute MLE Error. 102
A.13 Experiment 2: Correlation between HDIs and MLE estimates in test block. 102
A.14 Experiment 2: Correlation between HDIs and inferred temperature values in test block. 103

List of Figures

2.1 Example Mouselab-MDP trial. 9

3.1 (a) A diagram of the meta-level MDP. (b) In this example time step, the participant just clicked on the node highlighted in yellow ($C_{t=2}$) with a cost of -1 ($R_{t=2}$) and ends up in a new belief state ($B_{t=3}$). The numbers outside the nodes denote each node's index in the belief state. 14

3.2 (a) The Bayesian Information Criteria for each model. The lower the value, the better the fit. The vertical line denotes the Bayesian Information Criterion of simulated participants with costs matched to those of the winning model. (b) The average trial log likelihood for each trial, for the top three models. 23

3.3 Bayesian information criterion when inferring MLE parameter estimates from all 40 trials. 24

3.4 The fixed-effect model comparison for the "decreasing variance" environment. 29

3.5 The fixed-effect model comparison for the "constant variance" environment. 30

4.1 Spearman correlations between simulated and inferred parameter values. . 40

4.2 (a) The posterior model probabilities P(inferred model | simulated model), which show the probabilty of a given model being the best for the simulated datasets. (b) The inverted posterior model probabilities P(simulated model | inferred model), which show the probability of a simulated model having generated the data, given that it is the best model. 45

4.3 (a) The posterior model probabilities P(inferred model | simulated model), which show the probabilty of a given model being the best for the simulated datasets. (b) The inverted posterior model probabilities P(simulated model | inferred model), which show the probability of a simulated model having generated the data, given that it is the best model. 46

A.1 The fixed-effect model comparison with the expanded range of models (not all models pictures here). 89

A.2 Absolute fit versus post-test questionnaire. 90
A.3 Average number of clicks per level, simulated for different combinations of the depth cost weight and the non-forward search cost weight. 95
A.4 Average number of clicks per level, simulated for different combinations of the depth cost weight and the effort cost weight. 96
A.5 Average number of clicks per level, simulated for different combinations of the depth cost weight and the distance cost weight. 97
A.6 Average number of clicks per level, simulated for different combinations of the depth cost weight and the non-backward search cost weight. . . . 98
A.7 Parameter recovery under all three methods, for datasets simulated under all three methods. 100
A.8 Correlation between questionnaire measures. 106
A.9 Spearman's rank correlation coefficient between questionnaires and total number of clicks. Bars denote 95% confidence interval. Purple bars denote p-values below 0.05. 107
A.10 Spearman's rank correlation coefficient between questionnaires and number of early clicks. Bars denote 95% confidence interval. Purple bars denote p-values below 0.05. 108
A.11 Spearman's rank correlation coefficient between questionnaires and number of middle clicks. Bars denote 95% confidence interval. Purple bars denote p-values below 0.05. 109
A.12 Spearman's rank correlation coefficient between questionnaires and number of late clicks. Bars denote 95% confidence interval. Purple bars denote p-values below 0.05. 110
A.13 Spearman's rank correlation coefficient between questionnaires and inferred effort cost weight. Bars denote 95% confidence interval. Purple bars denote p-values below 0.05. 111
A.14 Spearman's rank correlation coefficient between questionnaires and inferred non-forward search cost weight. Bars denote 95% confidence interval. Purple bars denote p-values below 0.05. 112
A.15 Spearman's rank correlation coefficient between questionnaires and inferred non-backward search cost weight. Bars denote 95% confidence interval. Purple bars denote p-values below 0.05. 113
A.16 Spearman's rank correlation coefficient between questionnaires and inferred distance cost weight. Bars denote 95% confidence interval. Purple bars denote p-values below 0.05. 114

A.17 Spearman's rank correlation coefficient between questionnaires and inferred depth cost weight. Bars denote 95% confidence interval. Purple bars denote p-values below 0.05. 115

Bibliography

Alslaity, A., Chan, G., and Orji, R. (2023). A panoramic view of personalization based on individual differences in persuasive and behavior change interventions. *Frontiers in Artificial Intelligence*, **6**.

Anderson, J. R. (1989). A rational analysis of human memory. In H. L. R. III and F. I. M. Craik, editors, *Varieties of Memory and Consciousness: Essays in Honour of Endel Tulving*, pages 195–210. Lawrence Erlbaum Associates.

Anderson, J. R. (2013). *The Adaptive Character of Thought*. Psychology Press.

Arora, S. and Doshi, P. (2021). A survey of inverse reinforcement learning: Challenges, methods and progress. *Artificial Intelligence*, page 103500.

Becker, F., Wirzberger, M., Pammer-Schindler, V., Srinivas, S., and Lieder, F. (2023). Systematic metacognitive reflection helps people discover far-sighted decision strategies: A process-tracing experiment. *Judgment and Decision Making*, **18**, e15.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, **57**(1), 289–300.

Blais, A.-R. and Weber, E. U. (2006). A domain-specific risk-taking (DOSPERT) scale for adult populations. *Judgment and Decision Making*, **1**(1).

Bolenz, F., Kool, W., Reiter, A. M., and Eppinger, B. (2019). Metacontrol of decision-making strategies in human aging. *eLife*, **8**, e49154.

Brodersen, K. H., Ong, C. S., Stephan, K. E., and Buhmann, J. M. (2010). The balanced accuracy and its posterior distribution. In *2010 20th International Conference on Pattern Recognition*, pages 3121–3124. IEEE.

Bustamante, L. A., Oshinowo, T., Lee, J. R., Tong, E., Burton, A. R., Shenhav, A., Cohen, J. D., and Daw, N. D. (2023). Effort foraging task reveals positive correlation between individual differences in the cost of cognitive and physical effort in humans. *Proceedings of the National Academy of Sciences*, **120**(50), e2221510120.

Callaway, F., Lieder, F., Krueger, P., and Griffiths, T. (2017). Mouselab-MDP: A new paradigm for tracing how people plan. In *The 3rd Multidisciplinary Conference on Reinforcement Learning and Decision Making (RLDM 2017)*. RLDM.

Callaway, F., Lieder, F., Das, P., Gul, S., Krueger, P. M., and Griffiths, T. (2018). A resource-rational analysis of human planning. In C. Kalish, M. A. Rau, X. J. Zhu, and T. T. Rogers, editors, *Proceedings of the 40th Annual Meeting of the Cognitive Science Society, CogSci 2018, Madison, WI, USA, July 25-28, 2018*.

Callaway, F., van Opheusden, B., Gul, S., Das, P., Krueger, P. M., Griffiths, T. L., and Lieder, F. (2022). Rational use of cognitive resources in human planning. *Nature Human Behaviour*, pages 1–14.

Campbell, Z., Zakzanis, K. K., Jovanovski, D., Joordens, S., Mraz, R., and Graham, S. J. (2009). Utilizing virtual reality to improve the ecological validity of clinical neuropsychology: an FMRI case study elucidating the neural basis of planning by comparing the Tower of London with a three-dimensional navigation task. *Applied Neuropsychology*, **16**(4), 295–306.

Carleton, R. N., Norton, M. P. J., and Asmundson, G. J. (2007). Fearing the unknown: A short version of the Intolerance of Uncertainty Scale. *Journal of Anxiety Disorders*, **21**(1), 105–117.

Claeskens, G. and Hjort, N. L. (2008). *The Bayesian information criterion*, page 70–98. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Condon, D. M. and Revelle, W. (2014). The International Cognitive Ability Resource: Development and initial validation of a public-domain measure. *Intelligence*, **43**, 52–64.

Dang, J., King, K. M., and Inzlicht, M. (2020). Why are self-report and behavioral measures weakly correlated? *Trends in Cognitive Sciences*, **24**(4), 267–269.

Das, P., Callaway, F., Griffiths, T., and Lieder, F. (2019). Remediating cognitive decline with cognitive tutors. In *The 4th Multidisciplinary Conference on Reinforcement Learning and Decision Making (RLDM 2019)*, pages 98–102. RLDM.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., and Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, **69**(6), 1204–1215.

Diener, E., Emmons, R. A., Larsen, R. J., and Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, **49**(1), 71–75.

East-Richard, C., R-Mercier, A., Nadeau, D., and Cellard, C. (2020). Transdiagnostic neurocognitive deficits in psychiatry: A review of meta-analyses. *Canadian Psychology/Psychologie canadienne*, **61**(3), 190.

Eccles, J. S. (1983). Expectancies, values, and academic behaviors. In *Achievement and achievement motives*, pages 75–146. Freeman.

Eisenberg, I. W., Bissett, P. G., Enkavi, A. Z., Li, J., MacKinnon, D. P., Marsch, L. A., and Poldrack, R. A. (2019). Uncovering the structure of self-regulation through data-driven ontology discovery. *Nature Communications*, **10**(1).

Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., and Poldrack, R. A. (2019). Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proceedings of the National Academy of Sciences*, **116**(12), 5472–5477.

Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). Statistical power analyses using G$*$Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, **41**(4), 1149–1160.

Felso, V. and Lieder, F. (2023). How do individual differences in a planning task relate to psychometric measures and real-world planning? *PsyArXiv*.

Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., and Salkovskis, P. M. (2002). The Obsessive-Compulsive Inventory: Development and validation of a short version. *Psychological Assessment*, **14**(4), 485–496.

Fradkin, I., Adams, R. A., Parr, T., Roiser, J. P., and Huppert, J. D. (2020). Searching for an anchor in an unpredictable world: A computational model of obsessive compulsive disorder. *Psychological review*, **127**(5), 672.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, **19**(4), 25–42.

Funke, J. and Krüger, T. (1993). „Plan-a-Day" (pad). *Bonn: Psychologisches Institut der Universität Bonn*.

Gabaix, X. and Laibson, D. (2005). Bounded rationality and directed cognition.

Gabaix, X. and Laibson, D. (2017). Myopia and discounting. Technical report, National Bureau of Economic Research.

Garner, D. M., Olmsted, M. P., Bohr, Y., and Garfinkel, P. E. (1982). The eating attitudes test: psychometric features and clinical correlates. *Psychological Medicine*, **12**(4), 871–878.

Gelman, A., Meng, X.-L., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, pages 733–760.

Gigerenzer, G. (2008). Why heuristics work. *Perspectives on psychological science*, **3**(1), 20–29.

Gillan, C. M. and Daw, N. D. (2016). Taking psychiatry research online. *Neuron*, **91**(1), 19–23.

Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A., and Daw, N. D. (2016). Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *eLife*, **5**.

Grant, F., Guille, C., and Sen, S. (2013). Well-being and the risk of depression under stress. *PLoS one*, **8**(7), e67395.

Griffiths, T. L., Callaway, F., Chang, M. B., Grant, E., Krueger, P. M., and Lieder, F. (2019). Doing more with less: meta-reasoning and meta-learning in humans and machines. *Current Opinion in Behavioral Sciences*, **29**, 24–30.

Groetsch, C. (2015). Linear inverse problems. In O. Scherzer, editor, *Handbook of Mathematical Methods in Imaging*, pages 3–46. Springer New York, New York, NY.

Gul, S., Krueger, P. M., Callaway, F., Griffiths, T. L., and Lieder, F. (2018). Discovering rational heuristics for risky choice. *KogWis 2018*.

Hamilton, J. L., Connolly, S. L., Liu, R. T., Stange, J. P., Abramson, L. Y., and Alloy, L. B. (2015). It gets better: Future orientation buffers the development of hopelessness and depressive symptoms following emotional victimization during early adolescence. *Journal of Abnormal Child Psychology*, **43**, 465–474.

Harackiewicz, J. M., Rozek, C. S., Hulleman, C. S., and Hyde, J. S. (2012). Helping parents to motivate adolescents in mathematics and science: An experimental test of a utility-value intervention. *Psychological Science*, **23**(8), 899–906.

Hay, N., Russell, S., Tolpin, D., and Shimony, S. E. (2014). Selecting computations: Theory and applications. *arXiv preprint arXiv:1408.2048*.

He, R., Jain, Y. R., and Lieder, F. (2021). Measuring and modelling how people learn how to plan and how people adapt their planning strategies the to structure of the environment. In *Proceedings of ICCM 2021, 19th International Conference on Cognitive Modelling*.

Hedge, C., Powell, G., and Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, **50**(3), 1166–1186.

Held, L. and Bové, D. S. (2020). Likelihood and bayesian inference. *Statistics for Biology and Health. Springer, Berlin, Heidelberg,*.

Heller, A. S., Ezie, C. C., Otto, A. R., and Timpano, K. R. (2018). Model-based learning and individual differences in depression: The moderating role of stress. *Behaviour Research and Therapy*, **111**, 19–26.

Holt, D. V., Rodewald, K., Rentrop, M., Funke, J., Weisbrod, M., and Kaiser, S. (2011). The Plan-a-Day approach to measuring planning ability in patients with schizophrenia. *Journal of the International Neuropsychological Society*, **17**(2), 327–335.

Hopkins, A. K., Gillan, C., Roiser, J., Wise, T., and Sidarus, N. (2022). Optimising the measurement of anxious-depressive, compulsivity and intrusive thought and social withdrawal transdiagnostic symptom dimensions. *PsyArXiv*.

Hulleman, C. S., Barron, K. E., Kosovich, J. J., and Lazowski, R. A. (2016). Student motivation: Current theories, constructs, and interventions within an expectancy-value framework. *Psychosocial skills and school systems in the 21st century: Theory, research, and practice*, pages 241–278.

Huys, Q. J., Lally, N., Faulkner, P., Eshel, N., Seifritz, E., Gershman, S. J., Dayan, P., and Roiser, J. P. (2015). Interplay of approximate planning strategies. *Proceedings of the National Academy of Sciences*, **112**(10), 3098–3103.

Jain, Y. R., Gupta, S., Rakesh, V., Dayan, P., Callaway, F., and Lieder, F. (2019). How do people learn how to plan? In *Conference on Cognitive Computational Neuroscience (CCN 2019)*, pages 826–829.

Jain, Y. R., Callaway, F., Griffiths, T. L., Dayan, P., He, R., Krueger, P. M., and Lieder, F. (2022). A computational process-tracing method for measuring people's planning strategies and how they change over time. *Behavior Research Methods*, pages 1–43.

Johnson, E. J., Payne, J. W., Bettman, J. R., and Schkade, D. A. (1989). Monitoring information processing and decisions: The mouselab system. Technical report, Center for Decision Studies, Fuqua School of Business, Duke University.

Joireman, J., Sprott, D. E., and Spangenberg, E. R. (2005). Fiscal responsibility and the consideration of future consequences. *Personality and individual differences*, **39**(6), 1159–1168.

Joireman, J. A. (1999). Additional evidence for validity of the consideration of future consequences scale in an academic setting. *Psychological Reports*, **84**(3, Pt 2), 1171–1172.

Kafer, K. and Hunter, M. (1997). On testing the face validity of planning/problem-solving tasks in a normal population. *Journal of the International Neuropsychological Society*, **3**(2), 108–119.

Karvelis, P., Paulus, M. P., and Diaconescu, A. O. (2023). Individual differences in computational psychiatry: a review of current challenges. *Neuroscience & Biobehavioral Reviews*, page 105137.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, **90**(430), 773–795.

Kerby, D. S. (2014). The simple difference formula: An approach to teaching nonparametric correlation. *Comprehensive Psychology*, **3**, 11–IT.

Klašnja-Milićević, A., Vesin, B., Ivanović, M., and Budimac, Z. (2011). E-learning personalization based on hybrid recommendation strategy and learning style identification. *Computers & Education*, **56**(3), 885–899.

Kliegel, M., Martin, M., McDaniel, M. A., and Phillips, L. H. (2007). Adult age differences in errand planning: The role of task familiarity and cognitive resources. *Experimental Aging Research*, **33**(2), 145–161.

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., *et al.* (2017). The hierarchical taxonomy of psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, **126**(4), 454.

Kubala, V., Konidaris, G., and Greenwald, A. (2019). Inverse reinforcement learning from a learning agent. In *The 4th Multidisciplinary Conference on Reinforcement Learning and Decision Making (RLDM 2019)*. RLDM.

Le, D. Q. and Jang, D. (2023). Individual differences and situational constraint predict information search in negotiation planning. *Group Decision and Negotiation*, **32**(3), 667–699.

Liebowitz, M. R. (1987). Social phobia. *Modern Problems of Pharmacopsychiatry*.

Lieder, F. and Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, **43**.

Lieder, F., Chen, O. X., Krueger, P. M., and Griffiths, T. L. (2019a). Cognitive prostheses for goal achievement. *Nature Human Behaviour*, **3**(10), 1096–1106.

Lieder, F., Callaway, F., Jain, Y. R., Krueger, P. M., Das, P., Gul, S., and Griffiths, T. (2019b). A cognitive tutor for helping people overcome present bias. In *The 4th Multidisciplinary Conference on Reinforcement Learning and Decision Making (RLDM 2019)*. RLDM.

Litman, L., Robinson, J., and Abberbock, T. (2017). Turkprime. com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, **49**(2), 433–442.

Lynam, D. R., Smith, G. T., Whiteside, S. P., and Cyders, M. A. (2006). The UPPS-P: Assessing five personality pathways to impulsive behavior. *West Lafayette, IN: Purdue University*, **10**.

Lynch, J. G., Netemeyer, R. G., Spiller, S. A., and Zammit, A. (2010). A generalizable scale of propensity to plan: The long and the short of planning for time and for money. *Journal of Consumer Research*, **37**(1), 108–128.

MacLeod, A. K. and Conway, C. (2005). Well-being and the anticipation of future positive experiences: The role of income, social networks, and planning ability. *Cognition & Emotion*, **19**(3), 357–374.

MacLeod, A. K., Coates, E., and Hetherton, J. (2008). Increasing well-being through teaching goal-setting and planning skills: Results of a brief intervention. *Journal of Happiness Studies*, **9**(2), 185–196.

Marin, R. S., Biedrzycki, R. C., and Firinciogullari, S. (1991). Reliability and validity of the Apathy Evaluation Scale. *Psychiatry Research*, **38**(2), 143–162.

Newell, A. and Simon, H. A. (1972). *Human Problem Solving*, volume 104. Prentice-hall Englewood Cliffs, NJ.

Orne, M. T. (1996). Demand characteristics. In *Introducing Psychological Research*, pages 395–401. Springer.

Palan, S. and Schitter, C. (2018). Prolific. ac–a subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, **17**, 22–27.

Patton, J. H., Stanford, M. S., and Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology*, **51**(6), 768–774.

Patzelt, E. H., Kool, W., Millner, A. J., and Gershman, S. J. (2019). The transdiagnostic structure of mental effort avoidance. *Scientific Reports*, **9**(1), 1–10.

Payne, J. W., Bettman, J. R., and Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **14**(3), 534.

Peretz, C., Korczyn, A. D., Shatil, E., Aharonson, V., Birnboim, S., and Giladi, N. (2011). Computer-based, personalized cognitive training versus classical computer games: a randomized double-blind prospective trial of cognitive stimulation. *Neuroepidemiology*, **36**(2), 91–99.

Peters, B. R., Joireman, J., and Ridgway, R. L. (2005). Individual differences in the consideration of future consequences scale correlate with sleep habits, sleep quality, and GPA in university students. *Psychological Reports*, **96**(3), 817–824.

Pethtel, O. and Chen, Y. (2014). Life regrets and life satisfaction in mature adults. *The Journal of Happiness & Well-Being*, pages 106–118.

Phillips, L. H., Kliegel, M., and Martin, M. (2006). Age and planning tasks: The influence of ecological validity. *The International Journal of Aging and Human Development*, **62**(2), 175–184.

Ramachandran, D. and Amir, E. (2007). Bayesian inverse reinforcement learning. In *IJCAI*, volume 7, pages 2586–2591.

Rigoux, L., Stephan, K. E., Friston, K. J., and Daunizeau, J. (2014). Bayesian model selection for group studies–revisited. *NeuroImage*, **84**, 971–985.

Roecker, E. B. (1991). Prediction error and its estimation for subset-selected models. *Technometrics*, **33**(4), 459–468.

Rothkopf, C. A. and Dimitrakakis, C. (2011). Preference elicitation and inverse reinforcement learning. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 34–48. Springer.

Rouault, M., Seow, T., Gillan, C. M., and Fleming, S. M. (2018). Psychiatric symptom dimensions are associated with dissociable shifts in metacognition but not task performance. *Biological Psychiatry*, **84**(6), 443–451.

Russell, S. and Wefald, E. (1991). Principles of metareasoning. *Artificial intelligence*, **49**(1-3), 361–395.

Saunders, J. B., Aasland, O. G., Babor, T. F., De La Fuente, J. R., and Grant, M. (1993). Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption-II. *Addiction*, **88**(6), 791–804.

Schmitt, M. and Blum, G. S. (2020). State/trait interactions. *Encyclopedia of Personality and Individual Differences*, pages 5206–5209.

Scholnick, E. K. and Friedman, S. L. (1993). Planning in context: Developmental and situational considerations. *International Journal of Behavioral Development*, **16**(2), 145–167.

Schulte-Mecklenbeck, M., Kühberger, A., and Johnson, J. G. (2011). *A handbook of process tracing methods for decision research: A critical review and user's guide*. Psychology Press.

Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., and Lehman, D. R. (2002). Maximizing versus satisficing: Happiness is a matter of choice. *Journal of Personality and Social Psychology*, **83**(5), 1178–1197.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, pages 461–464.

Seow, T. X., Benoit, E., Dempsey, C., Jennings, M., Maxwell, A., O'Connell, R., and Gillan, C. M. (2021). Model-based planning deficits in compulsivity are linked to faulty neural representations of task structure. *Journal of Neuroscience*, **41**(30), 6539–6550.

Shallice, T., Broadbent, D. E., and Weiskrantz, L. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, **298**(1089), 199–209.

Simms, L. J., Wright, A. G., Cicero, D., Kotov, R., Mullins-Sweatt, S. N., Sellbom, M., Watson, D., Widiger, T. A., and Zimmermann, J. (2022). Development of measures for the hierarchical taxonomy of psychopathology (HiTOP): A collaborative scale development project. *Assessment*, **29**(1), 3–16.

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive psychology*, **7**(2), 268–288.

Simons, J. S. and Gaher, R. M. (2005). The Distress Tolerance Scale: Development and validation of a self-report measure. *Motivation and Emotion*, **29**(2), 83–102.

Spielberger, C. D., Gorsuch, R. L., Lushene, P. R., Vagg, P. R., and Jacobs, A. G. (1983). *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press.

Steffen, J., Marković, D., Glöckner, F., Neukam, P. T., Kiebel, S. J., Li, S.-C., and Smolka, M. N. (2023). Shorter planning depth and higher response noise during sequential decision-making in old age. *Scientific Reports*, **13**(1), 7692.

Steinberg, L., Graham, S., O'Brien, L., Woolard, J., Cauffman, E., and Banich, M. (2009). Age differences in future orientation and delay discounting. *Child Development*, **80**(1), 28–44.

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., and Friston, K. J. (2009). Bayesian model selection for group studies. *NeuroImage*, **46**(4), 1004–1017.

Strathman, A., Gleicher, F., Boninger, D. S., and Edwards, C. S. (1994). The consideration of future consequences: Weighing immediate and distant outcomes of behavior. *Journal of Personality and Social Psychology*, **66**(4), 742–752.

Su, R., Tay, L., and Diener, E. (2014). The development and validation of the comprehensive inventory of thriving (cit) and the brief inventory of thriving (bit). *Applied Psychology: Health and Well-Being*, **6**(3), 251–279.

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. MIT press.

Todd, P. M. and Gigerenzer, G. E. (2012). *Ecological rationality: Intelligence in the world*. Oxford University Press.

Toplak, M. E., West, R. F., and Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the cognitive reflection test. *Thinking & reasoning*, **20**(2), 147–168.

Tymula, A., Rosenberg Belmaker, L. A., Ruderman, L., Glimcher, P. W., and Levy, I. (2013). Like cognitive function, decision making across the life span shows profound age-related changes. *Proceedings of the National Academy of Sciences*, **110**(42), 17143–17148.

Unterrainer, J. M., Rahm, B., Kaller, C. P., Leonhart, R., Quiske, K., Hoppe-Seyler, K., Meier, C., Müller, C., and Halsband, U. (2004). Planning abilities and the Tower of London: is this task measuring a discrete cognitive function? *Journal of Clinical and Experimental Neuropsychology*, **26**(6), 846–856.

Unterrainer, J. M., Kaller, C. P., Halsband, U., and Rahm, B. (2006). Planning abilities and chess: A comparison of chess and non-chess players on the Tower of London task. *British Journal of Psychology*, **97**(3), 299–311.

Vallat, R. (2018). Pingouin: statistics in python. *Journal of Open Source Software (JOSS)*, **3**(31), 1026.

Van Beek, J., Antonides, G., and Handgraaf, M. J. (2013). Eat now, exercise later: The relation between consideration of immediate and future consequences and healthy behavior. *Personality and Individual Differences*, **54**(6), 785–791.

Van Opheusden, B., Galbiati, G., Bnaya, Z., Li, Y., and Ma, W. J. (2017). A computational model for decision tree search. In G. Gunzelmann, A. Howes, T. Tenbrink, and E. J. Davelaar, editors, *Proceedings of the 39th Annual Meeting of the Cognitive Science Society, CogSci 2017, London, UK, 16-29 July 2017*.

VandenBos, G. R. (2007). *APA Dictionary of Psychology*. American Psychological Association.

Whiteside, S. P., Lynam, D. R., Miller, J. D., and Reynolds, S. K. (2005). Validation of the UPPS impulsive behaviour scale: a four-factor model of impulsivity. *European Journal of Personality*, **19**(7), 559–574.

Wilson, R. C. and Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *eLife*, **8**, e49547.

Wise, T. and Dolan, R. J. (2020). Associations between aversive learning processes and transdiagnostic psychiatric symptoms in a general population sample. *Nature Communications*, **11**(1), 1–13.

Wolfbauer, I., Pammer-Schindler, V., and Rosé, C. (2020). Rebo junior: analysis of dialogue structure quality for a reflection guidance chatbot. In *EC-TEL Impact Paper Proceedings 2020: 15th European Conference on Technology Enhanced Learning*.

Zorowitz, S. and Niv, Y. (2023). Improving the reliability of cognitive task measures: A narrative review. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*.

Zung, W. W. (1965). A self-rating depression scale. *Archives of General Psychiatry*, **12**(1), 63.