

Data collection through design optimization
methods in MABTs-based experiments

Greta Monacelli, M.Sc.

Supervised by Prof. Tomás Ward¹ and Prof. Brendan Murphy²

¹ Dublin City University, ² University College Dublin



A thesis presented for the degree of Doctor of Philosophy

SCHOOL OF COMPUTING
DUBLIN CITY UNIVERSITY

June 2026

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy, is entirely my own work, and that I have exercised reasonable care to ensure that the work is original and have conformed to the regulations on the use and declaration of Generative AI, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Generative AI was used minimally in the development of this thesis, and only in the final year of the project. I utilized *ChatGPT* to suggest relevant papers on various occasions to properly justify a statement. Furthermore, I occasionally utilized *ChatGPT*, *Grammarly* or *Gemini* to identify grammar errors and sections of the text which could be improved in flow and clarity. In the later stages of the work, *Gemini* was also used to improve the readability of the code. All suggested changes were manually reviewed and implemented.

Signed: Greta Monacelli (Candidate) ID No.: 20213106 Date: June 29, 2026

To Mum, Dad, and Giulia

Acknowledgments

This work would not have been possible without the support of many around me. I would like to thank the Insight SFI Research Centre for Data Analytics for funding this work. I would like to thank my supervisors, Prof. Tomás Ward and Prof. Brendan Murphy. During this journey, they were always present to give me advice and encouragement. Their support was crucial for my growth as a researcher and for the success of this work. I wish to thank my colleagues, Lili Zhang and Marie McCarthy, for the inspiring conversations. I would like to thank my examiners Prof. Sofia Villar and Prof. Paolo Guassoni for the time dedicated to reviewing this work and giving me valuable feedback, as well as the independent chair Prof. Marija Bezbradica for managing the defence process.

Many others supported me behind the scenes in my personal life. First of all, my family: Dad, Mum, and Giulia remained my unmovable roots in these years of exploration. Then, all my relatives: Uncle Giorgio, Aunt Patrizia, Grandma Vanda, Uncle Gerry, Sara, Cristina, and many others. Everyone was always there to lend an ear to my troubles and projects. Then, there are the many friends who stayed by my side or whom I met during this Ph.D. My childhood friends, Federica, Luisa, Grazia, and the others. My longtime college friend and travel companion, Chiara. My friends from DCU, Devika and Ambru. And, finally, my online friends Gosu, Akami, Froggy, Alex, and all the others that spent time with me during the breaks.

To everyone mentioned here, thank you for your unwavering support, shared knowledge, and encouragement throughout this journey.

Contents

Notation table	vi
1 Introduction	1
2 Background	8
2.1 Multi-armed bandit tasks	8
2.2 Computational psychology models	12
2.3 Design optimization	14
2.4 Optimization methods	17
3 Adaptive data collection for intraindividual studies affected by adherence	19
3.1 The experiment	20
3.2 Previous work	23
3.3 Method	24
3.3.1 Control chart	25
3.3.2 Selection of the significance level and the triggering starting point	26
3.3.3 Adherence	28
3.4 Results	32
3.5 Discussion	39
3.6 Supplementary material	42
3.6.1 Practitioner guidelines for the selection of the starting point	42
3.6.2 Stopping rule	44

3.6.3	Other computational problems	47
4	Improving data collection through an inhomogeneous MCMC algorithm	49
4.1	The experiment	51
4.2	Learning model	52
4.3	Design optimization framework	60
4.4	Results	64
4.5	Discussion and outlook	78
5	Adaptive data collection through an inhomogeneous MCMC algorithm	82
5.1	The experiment	83
5.2	Selection of the new optimal design	87
5.3	Results	92
5.4	Conclusions and future directions	100
6	Conclusion	102

Notation table

Symbol/Abbr.	Description
<i>Abbreviations</i>	
MABT	Multi-armed bandit task
DO	Design optimization
EMA	Ecological momentary assessment
TD	Temporal difference
RW	Rescorla-Wagner
WL	Extended Rescorla-Wagner (win-lose)
WSLS	Win-stay-lose-shift
<i>General notation</i>	
K	Number of arms
N	Number of total trials for an EMA study
T	Number of total trials for a MABT
c_t, r_t	Choice and reward at trial t
$\mathbf{d} = (p_1, p_2)$	Experimental design (reward probabilities)
$U(d)$	Global utility function for design d
<i>Model parameters</i>	
κ	Significance level
$\alpha, \alpha_w, \alpha_l$	Learning rate (speed of value updates) in RW and WL models
τ	Inverse temperature (exploit/explore balance) in RW and WL models
ϵ	Stochasticity in WSLS model

List of Figures

3.1	Graphical representation of the algorithms	29
3.2	Total observations per subject: real vs. simulated data	33
3.3	eCDF of F1 score and utility in simulated and real data	34
3.4	MSE for varying sample sizes and method of moments requirements in Beta parameter estimation	43
3.5	Example of the optimal set A^* for a fixed experiment's length	44
3.6	Sensitivity analysis: impact of artificial data on parameter estimation	47
4.1	Simulated 2ABT experiment example for Model 1 (RW)	54
4.2	Simulated 2ABT experiment example for Model 2 (WL)	57
4.3	Simulated 2ABT experiment example for Model 3 (WSLS)	59
4.4	Cooling down schedule	64
4.5	Results for Model 1 (RW)	65
4.6	Results for Model 2 (WL)	66
4.7	Results for Model 3 (WSLS)	67
4.8	Regret analysis for Model 1 (RW)	70
4.9	Regret analysis for Model 2 (WL)	71
4.10	Regret analysis for Model 3 (WSLS)	72
5.1	Graphical representation of the experiment timeline	85
5.2	Example data collected through adaptive settings ($b = 24$)	86
5.3	Example data collected through adaptive settings ($b = 80$)	87
5.4	Parameter estimation error comparison	93

5.5	Experimental design trajectories	95
5.6	Boxplots comparing performance measures	97
5.7	Relative regret for cohorts with varying batch lengths	98

List of Tables

3.1	Wilcoxon signed-rank test results	37
3.2	Participant burden and information loss metrics	45

List of Publications and Collaborations

- McCarthy, M., Zhang, L., Monacelli, G., & Ward, T. (2021). Using Methods From Computational Decision-making to Predict Nonadherence to Fitness Goals: Protocol for an Observational Study. *JMIR Research Protocols*, 10(11), e29758. <https://doi.org/10.2196/29758>
- Monacelli, G., Zhang, L., Schlee, W., Langguth, B., Ward, T. E., & Murphy, T. B. (2023). Adaptive data collection for intraindividual studies affected by adherence. *Biometrical Journal*, e2200203. <https://doi.org/10.1002/bimj.202200203>
- Zhang, L., Monacelli, G., Vashisht, H., Schlee, W., Langguth, B., & Ward, T. (2022). The Effects of Tinnitus in Probabilistic Learning Tasks: Protocol for an Ecological Momentary Assessment Study. *JMIR Research Protocols*, 11(11), e36583. <https://doi.org/10.2196/36583>

Dissemination

Monacelli, G., Murphy, T. B., & Ward, T. (2023). Design optimization methods for decision-making tasks using Inhomogeneous Markov Chain Simulation [Poster, Computational Psychology Conference, Dublin].

Monacelli, G., Zhang, L., Schlee, W., Langguth, B., Ward, T., & Murphy, T. B. (2022). Adaptive Data Collection for intraindividual studies affected by adherence [Poster, Royal Statistical Society Conference, Aberdeen].

Monacelli, G., Zhang, L., Ward, T., & Murphy, T. B. (2021). Bayesian design-optimization for the multi-armed bandit tasks [Poster, Conference on Applied Statistics in Ireland, Online].

Data collection through design optimization methods in MABTs-based experiments

Greta Monacelli

Abstract

In computational psychology, multi-armed bandit tasks (MABTs) are often used as experiments to study human decision-making. Due to the increased computational power available to researchers and the possibility of implementing online experiments with mobile applications, there is a wide selection of experimental designs for these tasks. Design optimization (DO) methods can help researchers select appropriate experimental designs and guide decisions through an objective analysis of the experiment structure. In this work, we apply DO methods to three different contexts of MABT-based experiments and evaluate their effects on the data collection process and subsequent analysis. In particular, we utilise DO methods to select the optimal time to administer an MABT-based experiment to a subject based on their adherence history and self-reported intensity of symptoms. For this purpose, we employ control chart approaches to identify unusual values of the symptoms and then combine DO methods to include information on the subject's adherence. Our approach results in better data quality than a naive approach for both simulated and real subjects, especially for cohorts substantially impacted by adherence. Moreover, we investigate the optimal reward probabilities of a two-armed bandit task in both a static and an online experiment. We utilise an inhomogeneous MCMC algorithm to analyse the design variable space and select an optimal design: our results show that this method identifies designs that improve parameter recovery in a simulated setting, up to twice that of a random experimental design selection. We conclude that DO methods are effective in guiding experimental design choices. This work can be used to inform decisions on experimental designs for decision-making studies and contributes to the investigation of the effectiveness of DO methods in this area. These methods have the potential to reduce the patient burden in decision-making experiments, which is critical from both an ethical and data quality perspective.

Chapter 1

Introduction

In recent years, new technologies have enhanced research in the computational psychology field. A greater computational power allows for the application of more complex statistical methods and models such as machine learning techniques or MCMC algorithms. Moreover, new tools such as mobile applications or smartwatches greatly simplify the data collection of large data sets through the use of self-reports and sensors. These are efficient tools for the collection of longitudinal data in prolonged experiments.

Within the computational psychology field, these new methods can be applied, in particular, to decision-making experiments based on multi-armed bandit tasks (MABTs). These experiments are studied to better understand the human decision-making process as promising results link this type of data to the inner biological functions of the brain. Furthermore, as other computational psychology experiments, MABTs are also currently being investigated as potential new biomarkers for psychiatric and psychological conditions (Kwon et al., [2023](#)). This line of research could simplify the diagnosis of these conditions, aiding the practitioners in both identifying the problem and tracking improvements after an intervention. Progress in this area can deliver additional and always appreciated objective measures for the psychiatric field (Kwon et al., [2023](#)).

In this new context, many implementation challenges and efficiency problems remain open. Mobile technologies and outside-the-lab experiments allow for more

complex algorithms to be developed in order to collect data more efficiently. This often results in experiments that are much more burdensome for the subjects or costly for the researchers. This is true for the computational psychology field in general, but applies in particular to MABT experiments. Indeed, these tasks may require a high number of trials to guarantee good estimates of the parameters. This can increase the burden placed on the subjects, potentially decreasing data quality due to dropout or distractions. Furthermore, multiple administrations of such tasks may be necessary over time, further increasing the burden on the subjects. Thus, reducing the number of trials during a decision-making experiment, while obtaining as much information as possible during the data collection process, is of interest.

Design optimization (DO) methods are a possible solution to tackle these problems. These approaches have been applied in many different areas to optimize experimental designs. Their framework allows researchers to analyse an experiment through simulations and suggest an optimal strategy for the data collection. The criteria of optimality are set by the researcher through the selection of a utility function. Thus, this framework is highly versatile. Furthermore, DO methods can be used adaptively allowing the researcher not only to identify a good generic design, but also to tailor it to the individual subjects throughout the experiment.

There is a consistent corpus of research that suggests DO methods are a promising tool to improve the data collection process for psychology and psychiatric experiments (Kwon et al., 2023) and some work has been done in the specific context of MABT experiments (Melinscak & Bach, 2020; Valentin et al., 2024; S. Zhang & Lee, 2010). The success of these methods could positively impact research conducted in the field of computational psychology for both researchers and their subjects. Indeed, on one hand, researchers would have access to a valuable tool to improve the data collection process and focus their resources on targeted experiments. On the other hand, participating in experiments — especially when tedious tasks are involved — could be a less stressful experience for the subjects because of the tailoring of the experiments with DO approaches.

Nevertheless, more work is necessary to investigate the advantages and the bias introduced by this method. The application of these methods in computational psychology problems is relatively recent compared to other research fields such as clinical trials. Therefore, many questions remain open. There are instances where DO methods have yet to be applied or tested in both simulated and real settings. Moreover, more studies on the effectiveness and limitations of these methods are needed to guarantee an appropriate use of these tools in the field of computational psychology. Finally, the heterogeneity of the experimental designs available to the researcher in the computational psychology and psychiatry fields makes this goal even more challenging.

Of particular interest due to the new tools enabling longitudinal experiments in behavioural research is how DO methods perform in the context of MABT experiments. The tasks that are included in these experiments are often tedious and lengthy for the subjects. Thus, DO methods could be helpful to decrease the burden on subjects by optimizing the data collection process and, consequently, reducing the length of the experiments. These considerations lead us to the core hypothesis of this work: **DO methods improve data collection in MABT experiments**. Specifying a class of psychology experiments consistently restricts the area of research. Nevertheless, this hypothesis is still very broad to address as the number of possible factors — that is, design variables — which influence an MABT experiment is massive. Therefore, we further restrict our attention to more specific experiments.

For simplicity, in this work, we differentiate between internal and external factors. We define *internal factors* all those design variables which are defined by the mathematical framework of MABTs, that is, the reward probabilities, the length of the experiment, the magnitude of the rewards, and the number of arms. Every design variable not included in this list is considered an *external factor* in this work. Examples are “when” or “where” the experiment is delivered, a stimulus added to the environment during the experiment, etc. External factors are usually more difficult

to control than internal factors. This is especially true in an outside-the-lab setting where many variables of the experiments are outside of the researchers’ control. Therefore, tracking external factors can be challenging both because there are many to consider and because they are difficult to monitor. On the other hand, internal factors are easier to track as they are usually set up by the researcher at the start of the experiment. Nevertheless, the number of internal factors to include in a DO problem increases exponentially as the complexity of the MABT increases. This implies that as researchers aim to tackle more complex experimental designs, the application of DO methods becomes increasingly challenging.

Therefore, we tackle our research question from different perspectives, based on the type of design variable we are considering — internal or external. We obtain the three sub-hypotheses, which we investigate through the analysis of different experimental design challenges in MABT experiments. The first sub-hypothesis is that *DO methods improve data collection in MABT experiments when applied on internal design factors*. In particular, we focused on the reward probabilities of a two-armed bandit task, see Chapter 4. The second sub-hypothesis is that *DO methods can be applied adaptively/sequentially on internal factors to improve data collection in MABT experiments*. For this hypothesis, we utilised a DO method to adaptively change the reward probabilities of a two-armed bandit task over time, see Chapter 5. Finally, the third hypothesis takes a step back from the MABT experiments and focuses on the external factors which might affect the data collection process in these cases. Thus, we investigate if *DO methods improve data collection in MABT experiments also when applied to external factors*. In particular, we consider the best delivery times for the task and select them through an adaptive algorithm based on the DO framework, see Chapter 3.

To rigorously test these hypotheses and quantify what constitutes an “improvement” in data collection, it is essential to establish clear evaluation criteria at the outset. Throughout this thesis, the effectiveness of Design Optimization (DO) methods is evaluated through three primary lenses:

- **Parameter recovery:** The accuracy and precision with which the latent cognitive parameters of a given computational model (e.g., learning rates or inverse temperatures) can be estimated from the resulting behavioural data.
- **Regret:** The estimation penalty incurred by selecting a sub-optimal experimental design compared to an optimal one. In the context of DO, this measures the cost of poor design choices on the final data quality.
- **Burden-utility trade-off:** The balance between maximising the informational value of the data collected (utility) and minimizing participant fatigue, stress, or non-adherence (burden). This trade-off is considered for both external factors (e.g., the interruption cost and adherence drops associated with task delivery timing) and internal factors (e.g., the cognitive fatigue induced by highly informative but difficult experimental designs).

These evaluation criteria map directly to the main contributions of this work, allowing us to assess both the internal and external factors of MABT experiments:

- **Evaluating burden–utility trade-offs (Chapter 3):** Addressing external factors, we suggest a novel adaptive algorithm for task delivery timing in psychological experiments. By accounting for subject adherence and self-reported symptoms (e.g., in a tinnitus study), this contribution explicitly tackles the burden-utility trade-off, demonstrating how adaptive timing maximises useful data collection while mitigating patient non-adherence and burden.
- **Evaluating parameter recovery and regret in static designs (Chapter 4):** Addressing internal factors, we utilise the DO framework to optimize the reward probabilities of a static two-armed bandit task across three different computational models. We quantify the regret associated with standard, non-optimized designs and demonstrate that DO methods yield specific reward choices that significantly improve parameter recovery accuracy.
- **Evaluating parameter recovery and regret in adaptive designs (Chapter 5):** Extending the internal factor analysis, we update the reward probabil-

ities batch-by-batch for the Rescorla-Wagner model. We evaluate this adaptive approach against the baseline static designs to show a consistent, dynamic improvement in parameter recovery for shorter batch sizes.

Ultimately, these contributions provide practical insights and guidelines for interested practitioners (e.g., identifying when adaptive methods are most effective and how they impact patient burden), alongside a future roadmap to further extend these results and validate them in real-world settings.

In conclusion, this work focuses on the application and evaluation of DO methods in MABT experiments. The final goal is to show that these are valid tools to improve the effectiveness of the data collection process in these experiments, both for internal (the reward probabilities of the MABT) and external variables (when to administer the tasks). In the experiments we considered, DO methods improved the data collection process by allowing a better follow-up analysis or a larger amount of relevant data. We hope that this work will be a useful contribution to the application of DO methods to MABT experiments, which are of increasing interest to researchers in the psychological field.

The remainder of this thesis is laid out as follows. In Chapter 2, we review the multi-armed bandit task framework, computational psychology models such as the Rescorla-Wagner model, and the design optimization framework. This serves both as a review of the state-of-the-art approaches and establishes our motivation for applying DO methods to psychological experiments. In Chapter 3, we propose an algorithm that uses control charts and subject adherence histories to optimally select the time of task administration. We test the algorithm on real data from a tinnitus study. We show that adaptive timing improves data collection, especially in cohorts affected by adherence. In Chapter 4, we focus on internal factors of MABT experiments — in particular, the reward probabilities. We apply a DO method based on Inhomogeneous MCMC to optimise static designs for models such as the Rescorla-Wagner model and the stochastic win-stay-lose-shift model. We demonstrate that certain reward choices yield better parameter recovery. In Chapter

5, we extend the work of Chapter 4 into an adaptive framework, where the reward probabilities are updated batch-by-batch. The method shows a small, consistent improvement in parameter estimation for shorter batch sizes. Finally, in Chapter 6, we summarise our findings and highlight that DO methods improve data quality in MABT experiments when applied to both external timing factors and internal reward schemes. We also outline two key directions for future work. A first technical line of research aimed at expanding the work to include different models, integrating Bayesian approaches, and reducing the computational costs. A second applicative line based on the validation of these tools with real-world data and the development of practitioner-friendly tools.

Chapter 2

Background

This section reviews the current literature on applying DO methods to MABT experiments in psychological research. We begin by outlining the context of decision-making experiments based on MABTs, the central focus of this work. Next, we review the most common modelling approaches for the data collected during these experiments. Finally, we introduce the DO framework, illustrating how this methodology enhances data collection in computational psychology and MABT experiments specifically.

2.1 Multi-armed bandit tasks

Multi-armed bandit (MAB) problems provide a mathematical framework for modelling sequential decision-making. An MAB consists of multiple “arms”, each yielding rewards based on underlying probability distributions. Players sequentially choose arms to maximise their cumulative reward over time. Optimal strategies for various MAB problems include analytical solutions and computationally efficient reinforcement learning approaches (Sutton & Barto, 2018). Clinical trials frequently appear in MAB literature as a key application area. In this context, the researcher is the player, and the treatments are the arms. As participants enter a study, the researcher assigns them to treatments based on an optimal strategy. The MAB framework offers alternatives to traditional randomization, balancing the need for statistical power with the ethical imperative of prioritising life-saving treatments (Berry, 2006).

While mathematical methods can solve MAB problems optimally, human players rarely achieve this optimality. The patterns of these suboptimal behaviours have drawn the attention of computational psychologists, who use MAB problems to study human decision-making. In this field, the MAB transitions from an abstract decision-making framework into an experimental behavioural task. In a typical psychological setup, the researcher fixes the task length and reward probabilities, administers the MAB task to a cohort of subjects, and records their choices. The behavioural data is then analysed using specialised statistical models to infer the underlying cognitive processes (see Section 2.2).

The fundamental difference between these applications lies in the knowledge of the reward probabilities of the arms. In standard MAB problems and clinical trials, the reward probabilities are unknown to the decision-maker (the researcher or algorithm), who must optimize outcomes despite this. In computational psychology, however, the reward probabilities are unknown only to the human subject. The subject must infer these probabilities internally through trial and error. The researcher, conversely, knows the true probabilities because they designed the experiment. Furthermore, in adaptive experimental settings, the researcher can dynamically alter these probabilities to maximise the information gained about the subject's decision-making strategies.

Translating this theoretical framework into practice, several well-known decision-making experiments in computational psychology can be mathematically described as multi-armed bandit tasks (MABTs). One example is the Iowa Gambling Task, a stationary four-armed bandit task (Haines et al., 2018; Steingroever et al., 2014, 2015, 2018). Another example is the (stationary) two-armed bandit task used by Hertwig et al. (2004) to study human reactions to risky choices. More recently, the reversal learning task, a non-stationary two-armed bandit task, has been used to highlight subjects' difficulties adapting to a volatile environment in studies about anxiety disorders (Aylward et al., 2019; Browning et al., 2015).

While MABTs specifically focus on how subjects maximise rewards in artificial

game settings and the cognitive techniques they employ to achieve this, they belong to a much broader field of decision-making research. Other experimental settings tackle different aspects of decision processes. For example, perceptual decision-making tasks investigate the subject’s reaction to external sensory stimuli (Shadlen & Kiani, 2013). Some experiments, such as the trolley problem or the ultimatum game, explore the moral and social impact on decision-making. The delay discounting tasks and similar tasks study the trade-off between immediate and delayed rewards. Additionally, decision-making problems are not only a research domain of psychology but also of economics (Padoa-Schioppa, 2011). Game theory, which studies reward maximisation in the presence of multiple agents, is a popular example of decision-making studies developed in this area. Note that the borders between these types of experiments are fluid; for example, MABTs can be modified to include other factors in the study, such as delayed rewards (Cheng et al., 2022).

An important area of research in this setting is the link between the behavioural patterns that emerge from MABT analysis and the neurobiological function of the brain. Parameters inferred from reinforcement learning models in bandit tasks have been shown to map onto the integrity of specific neural circuits underlying human decision-making. The activation of dopaminergic brain regions has been observed during the completion of MABTs and similar decision-making experiments (Loosen et al., 2025). In particular, exploration–exploitation parameters are linked to the activation of specific brain regions (dorso-lateral prefrontal cortex, right dorsal anterior cingulate cortex, anterior insula and superior temporal gyrus) (Sazhin et al., 2025). Moreover, Blanchard and Gershman (2018) found that the insula and the dorsal anterior cingulate cortex are more active for exploration behaviour rather than in the exploitative counterpart. These computational parameters, therefore, provide a bridge between observed behaviour and underlying neural mechanisms (Kwon et al., 2023). For example, trial-by-trial prediction errors appear to modulate both BOLD responses in striatal areas (MRI) (Pessiglione et al., 2006) and mid-frontal theta power (EEG) (Cavanagh et al., 2010).

This cross-methodological alignment strengthens the argument that model-derived latent variables are tracking meaningful neurobiological processes, although test-retest reliability is still an ongoing challenge (Jami et al., 2025). Importantly, this integration has practical implications for computational psychiatry. Parameters such as diminished learning rates or exaggerated exploration tendencies can be mapped onto disruptions in dopaminergic and frontostriatal circuitry, providing mechanistically interpretable biomarkers for disorders ranging from substance dependence to obsessive-compulsive disorder (Jami et al., 2025). Thus, computational models serve as a methodology that translates observed behaviour into measurable indices of neural integrity.

Furthermore, groups affected by different psychological or psychiatric conditions have been shown to perform differently in these tasks, encouraging their study as biomarkers (Jami et al., 2025; Yousefi et al., 2019). Some initial results, for example, show that patients affected by addiction or anxiety disorders perform differently from healthy subjects (Ahn et al., 2016; Aylward et al., 2019; Browning et al., 2015). Moreover, parameters of models used in the computational psychology field to analyse this data have been linked to brain wave activity (Kwon et al., 2023). This further encourages the analysis of this data in order to better understand human behaviour at a neurobiological level as well.

Despite these promising application areas, MABTs and similar decision-making experiments are not without their limitations. In particular, *reliability* and *construct validity* have been identified as key weaknesses of these tasks. Reliability is the ability to replicate results when repeating the experiment in analogous settings and has been shown to be poor across a range of decision-making tasks (Karvelis et al., 2023). While some techniques that help strengthen reliability have been found — for example, multiple-session experiments and hierarchical Bayesian methods — more work in this area is needed before these tasks can be employed as phenotypes or screening methods in mental health disorders (Karvelis et al., 2023). Another key limitation is construct validity, which refers to a task’s ability to accurately measure

the specific psychological phenomenon intended by the researcher. It has been shown that task results are heavily dependent on the task design, challenging the generalization of these experimental measures to intrinsic decision-making paradigms (Karvelis et al., 2023). For example, weak links were highlighted between measures of risk-taking behaviour from decision-making experiments and support/compliance with regulations during the COVID-19 pandemic (Buelow et al., 2024). *Ecological validity* — that is, the ability to translate experimental insights into real-world knowledge about subject behaviour — is an instance of construct validity and is just as important. In general, MABTs have a simple or toy-like structure, so that many aspects of real-world decision processes such as social constraints or delayed rewards are overlooked. Static rewards are also rarely representative of human perception of gains/losses, yet they are frequently used (Gigli & Stella, 2025; Leqi et al., 2023).

2.2 Computational psychology models

The type of behavioural data that MABT experiments generate is as follows. Assume the MABT has $K \in \mathbb{N}^+$ arms. The rewards delivered by each arm S_i with $i = 1, \dots, K$ are selected respectively from the probability distributions f_i with $i = 1, \dots, K$. These often belong to the same family of distributions, such as the Bernoulli or the Normal distribution. The experiment is conducted through $T \in \mathbb{N}^+$ trials, and, for each trial, participants are asked to select one of the arms. After each choice, participants receive a reward sampled from the distribution relative to the selected arm. The goal of the participants is to maximise the final amount of rewards by the end of the experiment. Therefore, the data collected is both the sequence of choices made by the subject — $\mathbf{c} \in \{1, \dots, K\}^T$ — and the sequence of rewards received by the subject — $\mathbf{r} \in \mathbb{R}^T$.

One of the most well-known models to analyse such data is the Rescorla-Wagner model. This approach was initially introduced to study animal learning in Pavlovian settings and produced relevant results in that field (Miller et al., 1995). It was later adapted to MABT experiments, i.e. to non-Pavlovian settings where the subjects are

allowed to manipulate the game through their choices. This was achieved by adding a decision rule — such as the softmax rule (Dayan & Daw, 2008; Kruschke, 2008).

Stemming from this idea, many new models have been developed (Dai et al., 2015; Gershman, 2015; Haines et al., 2018; Lin et al., 2020; Sutton & Barto, 2018). For example, in Lin et al. (2020) the authors implement a double-stream model to highlight subject’s responses to positive and negative rewards separately. Aylward et al. (2019) utilise a non-static two-armed bandit task and a statistical model adequate for its analysis in order to highlight subjects’ responses to changing environments. These are usually referred to as temporal differences (TD) models or model-free approaches (Sutton & Barto, 2018). This class of methods is particularly popular in the computational psychology field because links between biological functions and the prediction error utilised in these models have been proven in many instances (Dayan & Daw, 2008). Very popular methods for the estimation of these models are (Hierarchical) Bayesian estimation approaches and computational packages in this direction have been implemented (Ahn et al., 2017).

There are other possible approaches to model this type of data such as model-based approaches and unsupervised machine learning. Model-based approaches seem to better describe sharp changes in decision-making compared to model-free approaches (Dayan & Daw, 2008). On the other hand, unsupervised machine learning has the potential to achieve better results compared to both the previous options if a large amount of data is available, but is less suited to study the inner biological workings of the decision-making process (Dezfouli et al., 2019).

The main idea behind TD models is to utilise a latent variable to track the subjects’ estimation of the rewards/losses of a given arm. Thus, the mathematical framework is similar to that of state space model approaches. However, the translation of the models utilised in the computational psychology field to the language of state space models is not always immediate. Slight modifications to the original models may be necessary to adapt them to this new framework. Nevertheless, the framework of state space models could provide a significant advantage for the computational psychology

field (Yousefi et al., 2019). Indeed, thanks to the adaptability of its framework, it could be used to develop broadly applicable packages in the field.

Since the MABTs framework involves a very large number of possible problems, the number of possible experimental designs which can be implemented based on them is very large as well. This poses a generalization challenge as it is non-trivial to choose an appropriate experimental design for the research question posed by the researcher among so many possibilities. Subsequently, the choice of a suitable statistical model for the data collected through this experimental design is just as difficult, considering the numerous options proposed by the literature for such purposes. As described above, the state space model approach proposed by Yousefi et al. (2019) is a possible solution to adapt to the still open question of new models in the computational psychology area. On the other hand, DO methods, see Section 2.3, can aid in selecting the experimental design.

2.3 Design optimization

Design optimization (DO) methods provide a mathematical framework to aid researchers in designing more efficient experiments. This approach relies on the application of optimization methods to screen the design variable space of an experiment. The objective is to identify the best candidates as experimental designs to achieve the goals of the study. Such goals are set by the researcher through the choice of a (local) utility function.

Here, we briefly describe the notation of DO methods. A more detailed explanation of the DO framework can be found in Chaloner and Verdinelli (1995). Let $\mathbf{x}$ denote the samples of the problem, i.e. the sequence of choices and rewards $\mathbf{c}, \mathbf{r}$ collected throughout the experiment. Select a suitable family of models $P_{\boldsymbol{\theta}}$ where $\boldsymbol{\theta} \in \Theta$ are the model parameters. Let $\mathbf{d} \in D$ be the design variables of the experiment. Define a *local utility function* $u(\mathbf{d}, \mathbf{x}, \boldsymbol{\theta})$ relative to specific designs $\mathbf{d} \in D$. The global utility for unknown data and parameters under the probability model selected is obtained by weighting the local utility function with the joint probability $p(\mathbf{x}, \boldsymbol{\theta})$ (Pitt &

Myung, 2019). Then, we obtain

$$U(\mathbf{d}) = \int u(\mathbf{d}, \mathbf{x}, \boldsymbol{\theta}) p(\mathbf{x}, \boldsymbol{\theta}) d\mathbf{x} d\boldsymbol{\theta}. \quad (2.1)$$

As is common in the literature, we assume that the joint probability is equal to the product of the marginals, i.e. $p(\mathbf{x}, \boldsymbol{\theta}) = p(\mathbf{x})p(\boldsymbol{\theta})$, which simplifies the computation of the integral in Equation (2.1).

Design optimization methods were originally motivated by their application to clinical trials and a large amount of literature about them covers this topic. For example, DO methods have been suggested to balance ethical issues and research-focused objectives in the submission of treatments to subjects during new-drug testing. This problem is usually formulated as a multi-armed bandit task, where the researcher has to sequentially select which treatments to assign to each participant — which arm to pull — to simultaneously measure the effectiveness of the treatment and provide ethical medical assistance to the subjects. In this context, Bayesian DO methods have been increasingly used. Indeed, they effectively account for the cumulative information gained throughout an experiment by continuously updating prior knowledge using Bayes’ theorem (Berry, 2006).

On the other hand, as mentioned before, MABTs can also describe experiments frequently used in the computational psychology field. In this case, the goal is not to solve the MABTs optimally but rather to analyse how the subjects approach it sub-optimally. The experiments are designed so that subjects sequentially select an arm in the MABTs in order to maximise their overall gains by the end of the experiment. There are two ways to utilise DO methods in this context. First, it is possible to screen the design variable space of the MABTs — for example, the reward probabilities of the arms. Second, DO methods can be applied to decide external factors of the MABT experiment — for example, when or where to submit the task.

Design optimization (DO) methods can be applied before or during an experiment, and they can also be adapted midstream to ongoing studies. This has become especially relevant as new technological tools become increasingly available, which

has also made complex algorithms for data collection in longitudinal studies much more feasible. An example is decentralised clinical trials where the data collection process follows the study participants to their home through the use of wearable sensors or smartphone apps. Similarly, researchers are using the longer time scale available thanks to these tools to deliver ecological momentary assessments — that is longitudinal experiments studying human health in participants’ daily lives — to better evaluate the stability and validity of their hypotheses in real-world settings.

A number of papers tackle the first problem. An analysis of the benefits of DO methods applied to this problem has been conducted for a static two-armed bandit task; the goal of the authors was to collect data efficiently in order to discern the best of two models (S. Zhang & Lee, 2010). Similarly, Valentin et al. (2024) focused on parameter recovery and model discrimination while providing a good workflow for the implementation of DO methods in this area. On the other hand, Melinscak and Bach (2020) utilised a similar approach to analyse a semi-stationary two-armed bandit task. The extension to a semi-stationary problem allows for the inclusion of more recently developed MABT experiments in the computational psychology field such as the reversal-learning task. In all cases, simulations suggest that applying DO methods improves data quality. Furthermore, Valentin et al. (2024) applied DO in a real-world experiment with encouraging results. Nevertheless, the vast design space associated with MABT experiments presents multiple directions for further improvement, such as conducting more real-world experiments or extending the investigation into larger design variable spaces.

For the second problem, the literature is much wider. Indeed, the same DO approaches which can be applied or developed for external factors in MABT experiments are also applicable to other kinds of experiments or tasks. DO methods were applied to a variety of other contexts such as for new material discovery (Lookman et al., 2019) or for neurophysiology experiments (Pillow & Park, 2016). In the context of computational psychology, there is growing interest in design optimization (DO) methods and their sequential counterpart, adaptive design optimization (ADO), for

various applications (Kwon et al., 2023). For example, ADO methods have been proven effective in decreasing the number of trials necessary to estimate parameters from a delay discounting task (Cavagnaro et al., 2010). In this context, it is worth mentioning that DO methods are not the only possible approach that can potentially improve the data collection process. Examples are control chart approaches or rule-based algorithms which have been proven effective for this purpose as well (Hekler et al., 2018; Hulme et al., 2021; Koizumi et al., 2012; Mohan, 2021; Thomas & Bond, 2015).

DO and ADO methods have thus proved to be effective in various contexts to aid the experimental design process; however, as every statistical tool, these methods have some limitations. Firstly, a debated downside of DO methods is their sensitivity to model misspecification. Possible solutions have been investigated in the context of mixture models (Chaloner & Verdinelli, 1995). In the specific context of computational psychology, this topic is even more relevant as the identification of a reliable and general approach to represent behavioural data is still under investigation (Dezfouli et al., 2019; Kwon et al., 2023). Secondly, adaptive data collection methods can potentially introduce bias into the data, skewing the follow-up analysis. Two approaches have been used to tackle this challenge: develop methods which consider the bias during the inference process or change the data collection algorithm to mitigate the bias (Nogas et al., 2021). Lastly, DO and ADO methods require statistical and computational knowledge. Some work has been done to develop tools for practitioners, but more progress in this direction is still needed to allow for a wide and effective use of these tools (Yang et al., 2021; Yousefi et al., 2019).

2.4 Optimization methods

Implementing DO methods requires the application of optimization techniques. While it is sometimes possible to solve these problems analytically, such solutions are usually restricted to very simple cases. The models utilised in computational psychology are often complex, and the design variable space of MABTs is large. Consequently,

these scenarios necessitate computational tools to solve the underlying optimization problems. A literature review of DO methods used in Bayesian DO can be found here (Ryan et al., 2016). We only mention three approaches often used in the computational psychology field: the most basic grid-based approach, Gaussian-based methods, and MCMC algorithms. Each of these approaches has been used effectively in the field of computational psychology (Cavagnaro et al., 2010; Melinscak & Bach, 2020; Valentin et al., 2024; S. Zhang & Lee, 2010).

In particular, in S. Zhang and Lee (2010) an MCMC algorithm based on augmented probabilities was applied. This algorithm is of particular interest as it merges the evaluation and exploration steps of the MCMC algorithm in an elegant theoretical framework and allows for a sensitivity analysis at no additional computational costs. Furthermore, as the utility of the DO problem is equal — up to a multiplying constant — to the selection probability of the MCMC algorithm due to the augmented probability method, such a sensitivity analysis is also a regret analysis for the selection of the best experimental design over its competitors.

In this chapter, we have introduced and reviewed the literature on the three main tools needed for this work. In particular, we introduced MABT experiments within the computational field and highlighted their application to the study of human decision-making. We considered the vast landscape of models introduced to analyse these experiments. Finally, we introduced the DO framework and its possible application to improve data collection in MABT experiments. In the following chapters, we will build on these concepts by focusing on their practical implementation and evaluation.

Chapter 3

Adaptive data collection for intraindividual studies affected by adherence

Recently, the use of mobile technologies in Ecological Momentary Assessments (EMA) and Interventions (EMI) has made it easier to collect data suitable for intraindividual variability studies in the medical field. Nevertheless, especially when self-reports are used during the data collection process, there are difficulties in balancing data quality and the burden placed on the subject. In this work, we address this problem for a specific EMA setting which aims to administer a demanding task to subjects at high/low values of a self-reported variable. We adopt a dynamic approach inspired by control chart methods and design optimization techniques to obtain an EMA triggering mechanism for data collection which takes into account both the individual variability of the self-reported variable and of the adherence. We test the algorithm in both a simulation setting and with real, large-scale data from a tinnitus longitudinal study. A Wilcoxon signed-rank test shows that the algorithm tends to have both a higher F_1 score and utility than a random schedule and a rule-based algorithm with static thresholds, which are the current state-of-the-art approaches. In conclusion, the algorithm is proven effective in balancing data quality and the burden placed on

the participants, especially in studies where data collection is impacted by adherence.

The code, the simulated data, and supplementary material about the implementation of the algorithms are available from the author or in the GitHub repository (<https://github.com/gretamonacelli/adc-tinnitus/releases/tag/v1.0.0>).

Due to privacy and ethical reasons, the data from the TrackYourTinnitus application is not publicly available. It is available on request (winfried.schlee@ieee.org).

3.1 The experiment

In recent years, mobile technologies have made longitudinal data in the medical field more easily available and, at the same time, have allowed the implementation of more complex and challenging experimental designs. Mobile applications or activity trackers can be used to collect multiple sensor data or self-reports for long periods of time. These new technologies have been effectively applied both in the behavioural field, through Ecological Momentary Assessments (EMA) and Interventions (EMI) (De Vries et al., 2021), and in clinical trials, through electronic patient reported outcomes or ePRO systems. These outside-the-lab experiments allow for more complex research goals and designs. For example, it is possible to send notifications through a mobile application to the participants based on the recognition of a relevant context such as symptom severity or physical location (De Vries et al., 2021; Hekler et al., 2018; Hulme et al., 2021). However, as the complexity of the experiments increases, the logistic challenges to correctly deliver them also increase and efficiency in data collection becomes necessary to reduce the costs of a lengthy experiment. The term costs may refer to different downsides of the experiment based on the context; examples are the energy consumption of the mobile application (Hekler et al., 2018) or the burden placed on the subjects during the data collection process.

In this context, intraindividual studies have gathered increasing attention (Ji et al., 2020). Intraindividual variability is an important aspect in medical research. For example, it has been used to detect neurological and psychological conditions (MacDonald et al., 2006). Moreover, it can negatively impact medical studies as

differences between groups may become less evident, especially for small sample-sized studies (MacDonald et al., 2006). The type of data collected in the outside-the-lab experiment mentioned above can be used for intraindividual variability studies. Indeed, for each subject, multiple observations of the same variable are collected over time. Moreover, the more complex designs made available by mobile technologies can be used. This allows us to optimize the data collection process and focus on the intraindividual differences of a variable when specific conditions are satisfied. In this work, we focus on an EMA which aims to present a demanding task to subjects at high/low values of a self-reported variable using a mobile application. The demanding task is assumed to be burdensome or costly for either the participants or the researcher, so that triggering it for every interaction of the subject with the application is not feasible.

An example of such an EMA setting for tinnitus patients is what motivated this work (L. Zhang et al., 2022). Tinnitus has great repercussions on the quality of life of chronically affected patients (Probst et al., 2016). Thus, evaluating the impact of this condition on cognitive processes and decision-making is of interest. Hence, the EMA study aims to collect both tinnitus self-reports and decision-making data from a group of tinnitus patients and compare their performance at an intraindividual level. In particular, from a neurological investigation perspective, the clinical question is the following: *is there a difference in decision-making behaviour during severe tinnitus episodes compared to less distressing periods?* To answer this question, we aim to compare the performance of a study participant in decision-making tasks collected during high and low self-reported states of tinnitus severity.

Multi-armed bandit tasks are used to collect the decision-making data. These tasks have been used effectively in the emerging field of computational psychiatry to evaluate human decision-making and are considered potential new biomarkers for psychiatric and psychological conditions (Ahn et al., 2016). Nevertheless, they are relatively lengthy and the collection of such data can be burdensome on the subjects. Therefore, there is a necessity to adaptively track the subjective high and

low states of tinnitus severity and trigger these lengthy tasks only in such cases. This decreases the number of multi-armed bandit tasks which need to be administered to each subject while still allowing an intraindividual comparison at meaningful values of tinnitus severity.

Notice that the experiment described here can be easily generalized: the multi-armed bandit task can be replaced by a generic *additional task* which has great cost either for the researcher (expensive equipment) or for the subject (lengthy or invasive); similarly, researchers can choose the *self-reported variable* based on their research question. An example of another possible application is clinical trials which rely on ePRO systems. These research methods often require an extensive questionnaire to be administered, such as the *EuroQoL 5d-5L*, to patients on their mobile phone when symptoms tracked through a wearable device show abnormal values (Kenning et al., 2024; McClaine et al., 2024). A careful selection of the triggering time for the questionnaire thanks to an adaptive tracking of the symptoms could limit the burden placed on the subjects during the data collection process.

There are two possible immediate, but sometimes ineffective, solutions to this sampling problem. The first option is a *random algorithm* which randomly selects when to trigger the additional tasks. The second option is a *static algorithm* which triggers the additional tasks for values of the self-reported variable above or below predetermined thresholds. Several studies in the behavioural field have implemented either the first or the second option in similar contexts to the one tackled in this work (De Vries et al., 2021). These sampling strategies are relatively easy to implement, but they do not consider three major problems which, instead, we ought to tackle. First, the expected value and variance of the self-reported variable may vary across subjects. For example, as we know from prior studies, tinnitus patients present a diversity of self-reported symptom ranges (Probst et al., 2016). Second, it is necessary to balance the necessity of the researcher to collect enough data with the risk of overburdening the subjects with a overly intensive data collection process. Third, we need to consider low adherence to the study. This is important, especially when

self-reports are considered, as they place additional burden on the subjects and can lead to both incomplete data and early drop-out from the experiment. Hence, the former triggering strategies may not trigger at all for some patients and excessively trigger for others. Moreover, there could be cases in which only additional tasks for high values of the self-reported variables are triggered and none for low values (or vice versa). In these cases, we expect to require much more data to increase the likelihood of capturing data at the extremes targeted. This reason necessitated the development of the adaptive triggering algorithm described here.

In this work, we develop an algorithm which tackles these three factors: the variation of the self-reported variable over time, the burden of lengthy experiments and the adherence of the subjects. The algorithm is heavily inspired by control chart approaches and adapts to the individual subject based on their past history of self-reports (Fricker et al., 2008). Moreover, it uses a design optimization method to change the definition of *high* and *low* values for the self-reported variable in an adaptive way, based on the expected adherence of the subject and on the balance between the burden and the efficacy of the experiment. Thus, the algorithm is designed to collect data as soon as possible for non-adherent subjects and, in contrast, to burden the adherent ones only when strictly necessary. The algorithm outperforms, according to our metrics, the state-of-the-art approaches — the random schedule and the static algorithm — for both simulated and real tinnitus data; moreover, our results suggest that the algorithm is particularly effective when the adherence of the subjects is highly variable — a very common situation in EMA studies which rely on self-reports.

3.2 Previous work

Several adaptive methods have been developed to improve data collection in EMA and EMI, but none focuses on the delivery of an additional and demanding task for a predetermined number of times at extreme self-reported values or on the explicit use of the adherence history to improve the algorithm. For example, Mohan (2021)

and Hekler et al. (2018) increase the efficacy of EMI through adaptive rule-based algorithms. Hekler et al. (2018) propose a statistical approach to decrease the energy consumption of the mobile application used in the study while maintaining the effective monitoring of a specific heart-condition. Additionally, Thomas and Bond (2015) and Hulme et al. (2021) have developed solutions which deliver just-in-time notifications to the participants through approaches based, respectively, on predetermined rules and hidden Monte Carlo methods.

Previous work from two different fields has been combined to obtain the results in this work: anomaly — or outlier — detection and design optimization. A similar approach has been used by Koizumi et al. (2012) to search for an optimal significance level in an anomaly detection problem. First, we looked at the anomaly/outlier detection literature to identify extreme values for the quantity of interest. Identifying unusual patterns in the data concerns many different applications and there is a vast literature both in the computational field - anomaly detection (Chandola et al., 2009) - and in the statistical field - outlier detection (Zimek et al., 2012). Secondly, we used a design optimization perspective to obtain a more balanced definition of extreme values which takes into account the adherence of the subjects and, thus, the sample size of the data they provide during the experiment. Design optimization methods have been applied in various contexts to balance data quality and experimental costs: examples include clinical trials for drug testing (Sylvester, 1988) and experiments in computational psychiatry (Cavagnaro et al., 2010).

3.3 Method

Let us consider an N_d day long experiment. During this time, a participant in the study is asked to self-report a quantity of interest N_h times per day through notifications pushed to their mobile phone. The time at which these notifications are sent can be either prefixed or randomized. At the end of the experiment, the participant will have received a total of $N = N_d N_h$ notifications. The data is collected as a time series $\mathbf{x} = \{x_t\}_{t=1}^N$, where $x_t \in [0, 1] \cup \{\text{NaN}\}$, $t \in \{1, \dots, N\}$ is the discrete

time and NaN denotes missing values. We wish to trigger an additional task for *high* and *low* values of the quantity of interest x_t .

3.3.1 Control chart

Let us first consider the simplified case of a completely adherent participant, so that the time series $\mathbf{x}$ does not contain any missing values (NaN). Assume that $\mathbf{x}$ is an i.i.d. sample from a random variable $X \sim \text{Beta}(\delta, \xi)$, i.e. a Beta distribution with parameters δ and ξ . Let the *triggering starting point* $S \in \{2, \dots, N\}$ be the first time point after which the additional task can be triggered. For every $s \in \{S-1, \dots, N\}$, consider the unbiased estimators of the expected value and the variance of the sub-sample $\mathbf{x}_{1:s} = \{x_t\}_{t=1}^s$, i.e.

$$\hat{\mu}_s = \frac{1}{s} \sum_{r=1}^s x_r \quad \text{and} \quad \hat{\sigma}_s^2 = \frac{1}{(s-1)} \sum_{r=1}^s (x_r - \hat{\mu}_s)^2.$$

Assume that, for all $s \in \{S-1, \dots, N\}$,

$$0 < \hat{\sigma}_s^2 < \hat{\mu}_s(1 - \hat{\mu}_s). \quad (3.1)$$

Then, the estimates of the parameters δ and ξ based on the sub-sample $\mathbf{x}_{1:s}$ obtained through the method of moments are

$$\hat{\delta}_s = \hat{\mu}_s \nu_s \quad \text{and} \quad \hat{\xi}_s = (1 - \hat{\mu}_s) \nu_s, \quad (3.2)$$

where $\nu_s = \hat{\mu}_s(1 - \hat{\mu}_s) / \hat{\sigma}_s^2 - 1$. The method of moments was preferred to the maximum likelihood estimation due to computational reasons since there is no closed-form solution for the latter option.

Let $\kappa \in [0, 1]$ be the *significance level*. Then, we say that the observation x_t is *extreme* if

$$x_t < z_{\frac{\kappa}{2}}(\hat{\delta}_{t-1}, \hat{\xi}_{t-1}) \quad \text{or} \quad x_t > z_{1-\frac{\kappa}{2}}(\hat{\delta}_{t-1}, \hat{\xi}_{t-1}), \quad (3.3)$$

where $t \geq S$ and $z_\beta(\delta, \xi)$ is the β -quantile of the Beta distribution with parameters δ

and ξ . If Condition (3.3) is satisfied at time t , then the additional task is triggered.

In the anomaly detection and the outlier detection literature, the control chart approach is usually used to identify observations which are *unlikely* under a probability model, i.e. *outliers*. Nevertheless, it can be applied in the context of this work by assuming that no outliers or anomalies are present in the data. Indeed, under this assumption, the values identified through Equation (3.3) are considered valid even though unlikely. Hence, they can be considered *extremes* because of their position 'close' to the boundaries of $[0, 1]$, outside of the interval $[z_{\frac{\kappa}{2}}(\hat{\delta}_{t-1}, \hat{\xi}_{t-1}), z_{1-\frac{\kappa}{2}}(\hat{\delta}_{t-1}, \hat{\xi}_{t-1})]$.

3.3.2 Selection of the significance level and the triggering starting point

In Section 3.3.1, the criteria to select the triggering starting point S and the significance level κ are not specified. In the outlier and anomaly detection literature, both of these design variables are usually chosen by the researcher based on domain knowledge or previous studies (Chandola et al., 2009). In this section, we investigate the choice in the formal framework of design optimization in order to aid the researcher in the final selection as Chaloner and Verdinelli (1995) and following a similar approach to Koizumi et al. (2012).

First, we define a utility function that represents the goal of the researcher. We assume that, in order to maintain a balance between the burden placed on the subjects and the necessity of the researcher to collect enough data, the primary goal of the study is to trigger the additional task on average $v \in \{1, \dots, N\}$ times for each participant.

To define the utility, let us consider the time series $\mathbf{w} = \{w_t\}_{t=S}^N$ defined as

$$w_t = \begin{cases} 1 & \text{if the control chart triggers the additional task at time } t \\ 0 & \text{otherwise.} \end{cases}$$

Assume that $\mathbf{w}$ is an i.i.d. sample from a Bernoulli random variable $W \sim \text{Ber}(\kappa)$. Let the random variable V represent the total number of additional tasks triggered during the experiment. Then, $V \sim \text{Bin}(\kappa, N - S + 1)$ follows a Binomial distribution with parameters κ over $N - S + 1$ independent experiments. Thus, the expected value of V is equal to $E(V) = (N - S + 1)\kappa$. We choose as utility function $U_1 : \{2, \dots, N\} \times [0, 1] \rightarrow \mathbb{R}$ defined as

$$U_1(S, \kappa) = -(E(V) - v)^2.$$

Thus, $U_1(S, \kappa)$ associates higher value to experimental designs (S, κ) for which the expected number of triggers $E(V)$ is as close as possible to the desired one v .

Let $D = \{(S, \kappa) \in \{2, \dots, N\} \times [0, 1]\}$. Then, we aim to find an optimal design (S^*, κ^*) such that

$$(S^*, \kappa^*) \in \arg \max_{(S, \kappa) \in D} U_1(S, \kappa), \quad (3.4)$$

The design optimization problem in Equation (3.4) has as solutions

$$A^* = \left\{ (S^*, \kappa^*) \in D \mid \kappa^* = \max \left(0, \min \left(1, \frac{v}{(N - S^* + 1)} \right) \right) \right\}. \quad (3.5)$$

Therefore, experimental designs chosen in the set A^* are all optimal for collecting, on average, v additional tasks.

Smaller values of the significance level κ should be preferred since they result in more significant intraindividual comparisons at the end of the experiment. Indeed, a smaller value of κ results in more extreme values of the self-reported variable to be flagged as high or low. The additional tasks are then submitted to the subjects at those extreme states. The bigger the difference between a high state and a low state, the better we expect the difference in the subject's performance in the additional task to be highlighted. If κ^* is chosen to be small, then, by the constraint in Equation (3.5), the starting point S^* is also small. However, while smaller values of S^* can be useful in case of low adherence, they will also likely result in worse data quality since control charts rely on past observations to improve over time.

We suggest to first fix S^* based on the domain knowledge of adherence in the field and, in any case, to not choose too small a value as it can negatively impact the estimation of the parameters $\hat{\delta}_s$ and $\hat{\xi}_s$. Once S^* has been fixed, Equation (3.5) gives the optimal significance level κ^* , i.e.

$$\kappa^* = \begin{cases} 0 & \text{if } N - S^* + 1 < 0 \\ 1 & \text{if } 0 \leq N - S^* + 1 \leq v \\ \frac{v}{(N - S^* + 1)} & \text{otherwise.} \end{cases} \quad (3.6)$$

Combining Equation (3.5) and the control chart in Section 3.3.1, we obtain Algorithm 1. Notice that, since the thresholds in the control chart are symmetric with respect to the probability, Algorithm 1 triggers on average $v/2$ additional tasks for both high and low values of the quantity of interest.

Algorithm 1

```

1: Set  $N \in \mathbb{N}^+$  and  $S^*$ ,  $v \in \{2, \dots, N\}$ . Set  $\kappa = \max\left(0, \min\left(1, \frac{v}{(N - S^* + 1)}\right)\right)$ .
2: for  $t = 1, \dots, N$  do
3:   collect  $x_t$ 
4:   if  $t \geq S^*$  and  $x_t \neq \text{NaN}$  then
5:     compute  $\hat{\delta}_{t-1}$  and  $\hat{\xi}_{t-1}$  as defined in Equation (3.2)
6:     if Condition (3.3) holds then
7:       trigger the additional task
8:     end if
9:   end if
10: end for
```

3.3.3 Adherence

The selection of the significance level in Section 3.3.2 relies on complete adherence from the subjects to the EMA. Indeed, the definition of κ^* in Equation (3.6) assumes prior knowledge of the final number of samples N . Equivalently, missing data points are not considered. Nevertheless, this assumption is almost never satisfied in clinical longitudinal studies which rely on self-reports (De Vries et al., 2021).

Assume the dataset has missing data. Then, the final total number of samples

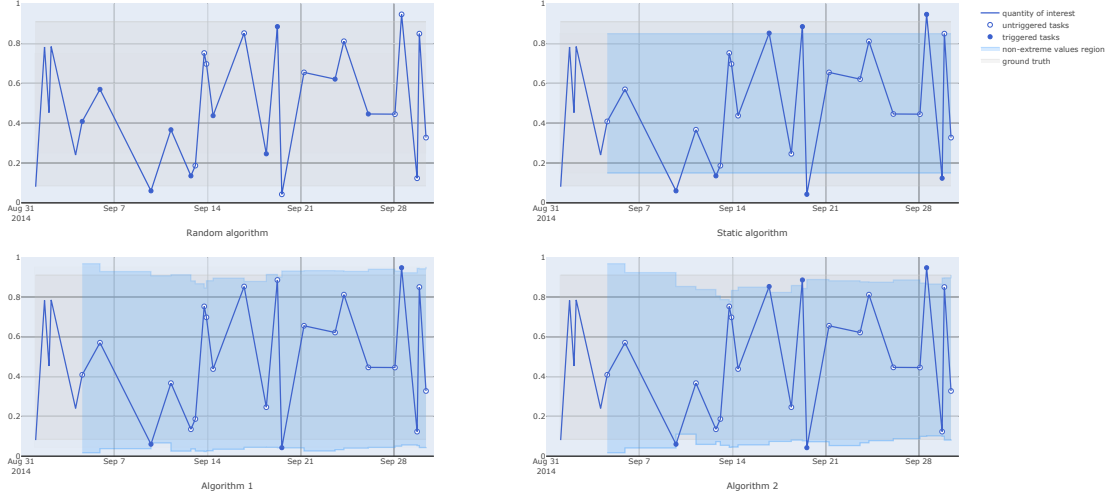


Figure 3.1: Graphical representation of the random algorithm, the static algorithm, Algorithms 1 and 2. The blue line represents the time series of the self-reported tinnitus severity for a synthetic patient. We set $S^* = 6$. Dots colored in white represent not-triggered tasks, while those colored in blue represent triggered ones. The blue area is the region of non-extreme values. In the case of the static algorithm, this consists in all observations between 0.15 and 0.85. On the other hand, for Algorithms 1 and 2, it represents the confidence intervals defined adaptively by the algorithms during the data collection process. The gray area represents the ground truth. Its thresholds are computed using Equation (3.3) and selecting the optimal significance level κ^* based on the true adherence of the subjects — known a posteriori at the end of the experiment. In particular, κ^* is chosen replacing in Equation (3.6) the number of notifications N with the number of samples collected for the individual subject N' . Therefore, triggered tasks outside this area are defined as true positives, triggers inside are false positives. Non-triggered tasks inside the area are true negatives, non-triggered tasks outside the area are false negatives. Notice that the confidence intervals of Algorithm 2 are smaller than those of Algorithm 1.

for the quantity of interest is $N' < N$. In this case, Algorithm 1 triggers on average a lower number of additional tasks than v . Indeed, if $N' > v + S - 1$, using Equation (3.6) we obtain that

$$\kappa^* = \frac{v}{N - S + 1} < \frac{v}{N' - S + 1}.$$

Thus, the optimal significance level for the missing data scenario should be larger than the one selected assuming complete adherence to trigger a similar number of additional tasks.

In order to tackle this problem, we can replace N in Algorithm 1 with an estimator of N' . If data from a pilot study are available, it is possible to use a simple estimate

of the total final number of samples. For example, let B be the set of all the subjects in the pilot study and N'_j , $j = 1, \dots, |B|$ be the total number of samples collected for each subject j . Then, N can be estimated by the average number of samples per subject, i.e. $\overline{N'} = \frac{1}{|B|} \sum_{j \in B} N'_j$. In this work, we used the tinnitus dataset to obtain the estimate $\overline{N'}$ and, in the following, we will always assume that Algorithm 1 uses it.

Nevertheless, pilot studies are not always available. In those cases, it may be difficult to predict accurately the average adherence of the subjects. Moreover, even if this data is available, small sample size, external factors, interventions and/or individual differences may decrease the accuracy of the algorithm.

Another possible approach is to infer a simple adaptive estimate of N' . The idea is, at each time step of the experiment, to only consider the history of adherence of a participant. In particular, we track the frequency of the subject's successful interactions with the applications, that is the number of successfully collected self-reports and the number of ignored prompts by the subject. This allows us to better understand the compliance of the subject to the study and approximate the expected final number of self-reports N' registered by each specific subject. Finally, we replace N' in Equation (3.6) with this estimation to obtain a new algorithm. The advantage of this method is that the estimation of N' does not require prior external data.

Let

$$a_t = \begin{cases} 0 & \text{if } x_t = \text{NaN} \\ 1 & \text{if } x_t \neq \text{NaN}, \end{cases}$$

so that the time series $\mathbf{a} = \{a_t\}_{t=1}^N$ represents the adherence of the subjects.

Assume that a participant decides to report the quantity of interest randomly and let the r.v. $A \sim \text{Ber}(\chi)$ represent the participant's behaviour. A success represents an interaction with the app, while a loss represents a missing data point. Then, the number of successes N' , i.e. the total number of samples we collect from that subject at the end of the experiment, follows a Binomial distribution with parameters χ and

N . In particular,

$$E(N') = \chi N. \quad (3.7)$$

We can estimate the probability of success χ adaptively through the method of moments, thus obtaining

$$\hat{\chi}(s) = \frac{\sum_{i=1}^s a_i}{s} \quad S^* - 1 \leq s \leq N'. \quad (3.8)$$

Then, replacing χ in Equation (3.7) provides the following estimator for the expected value of final samples N' :

$$\hat{N}'(t) = \hat{\chi}(t-1)N. \quad (3.9)$$

Thus, we obtain the following adaptive definition for the significance level, where κ is now a function of the time t :

$$\kappa^*(t) = \begin{cases} 0 & \text{if } \hat{N}'(t) - S^* + 1 < 0 \\ 1 & \text{if } 0 \leq \hat{N}'(t) - S^* + 1 \leq v \\ \frac{v}{(\hat{N}'(t) - S^* + 1)} & \text{otherwise.} \end{cases} \quad (3.10)$$

In conclusion, we obtain Algorithm 2.

Algorithm 2

```

1: Set  $N \in \mathbb{N}^+$  and  $S^*, v, R \in \{2, \dots, N\}$ .
2: for  $t = 1, \dots, N$  do
3:   collect  $x_t$ 
4:   if  $t \geq S^*$  and  $x_t \neq \text{NaN}$  then
5:     compute  $\hat{\delta}_{t-1}$  and  $\hat{\xi}_{t-1}$  as defined in Equation (3.2)
6:     compute  $\hat{N}'(t)$  and  $\kappa^*(t)$  as defined in Equations (3.9) and (3.10)
7:     if Condition (3.3) holds and  $\sum_{i=1}^{t-1} w_i \leq R$  then
8:       trigger the additional task
9:     end if
10:   end if
11: end for

```

Notice that the definition of extreme values in Algorithm 2 depends on the adherence of the participant. More adherent participants are expected to have lower

values of the adaptive significance level $\kappa^*(t)$ than less adherent ones. Thus, the additional tasks will be triggered for more *significantly* unlikely/extreme values of x_t if the participant is adherent. In contrast, the algorithm tends to collect data even if it is not highly significant/extreme from a statistical viewpoint for non-adherent subjects. This characteristic can be both an advantage and a disadvantage. It is an advantage as Algorithm 2 adapts to the adherence of the subjects as well as to the history of the quantity of interest. This allows the researcher to collect a good amount of data for their experiment while preserving data quality and reducing the burden placed on the subjects by the experiment. On the other hand, it is a disadvantage because, after the collection of the data, this inconsistency should not be overlooked in the subsequent analysis of intraindividual or intra-group differences.

3.4 Results

In this section we compare a random and a static schedule with Algorithms 1 and 2 in both a simulation setting and on real data from a tinnitus longitudinal study. The random schedule randomly selects 10 interactions of the user with the app to trigger the additional task. If the user does not respond to the self-report at the selected times, the additional task trigger shifts to the next interaction of the user with the application until a separate trigger is set. The static algorithm is a rule-based approach with pre-fixed thresholds at 0.15 and 0.85. All results in this work were produced using R (Team, 2023). The figures were obtained using the `plotly` package (Sievert, 2020).

In both the simulation and the real data analysis, $v = 4$ and $N = 180$ are fixed. Similarly, the triggering starting point is set to $S^* = 6$. The latter decision was taken by considering the adherence in the real data and trying to balance two contrasting necessities. On one hand, it is important to start collecting the additional task early as the adherence of the subjects decreases exponentially, see Figure 3.2. On the other hand, both Algorithms 1 and 2 depend on past observations and are not reliable without some prior data. In particular, the method of moments used for

the parameter estimation may encounter computational errors if S^* is too small. For more details on the choice of the starting point, we refer to the supplementary material.

Simulated data We simulated two data sets. In both cases, the data was generated from the statistical model used to develop Algorithm 2 so that the adherence follows a Binomial distribution and the quantity of interest follows a Beta distribution. The Beta distribution parameters for the simulated participants are sampled from a uniform distribution in the interval $[0.5, 10]$. However, in the *first simulated data* the adherence is considered the same for all subjects. In particular, the total number of samples for each subject was sampled from a Binomial distribution with parameters $N = 180$ and $\chi = 0.19$. On the other hand, in the *second simulated data* more variability is added to the adherence by sampling, for each subject, a probability of success χ from a Beta distribution $\text{Beta}(1, 4)$. As shown in Figure 3.2, this results in a data set more similar to the real data.

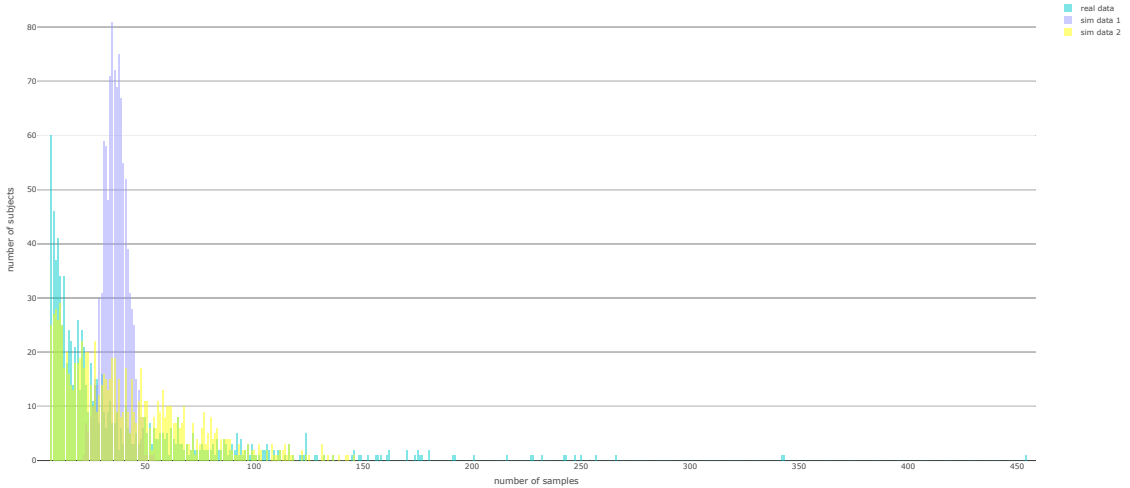


Figure 3.2: Comparison between the total number of observations per subject in the real data and the simulated data. Notice that some participants in the real data set have a higher number of observations than expected as they interacted with the application more than 6 times per day. In those cases, all samples are considered to estimate the parameters of the Beta distribution and an additional task can be triggered at those time-points. However, only the first 6 samples per day are considered for the estimation of $\hat{\chi}(t - 1)$ in Equation (3.8).

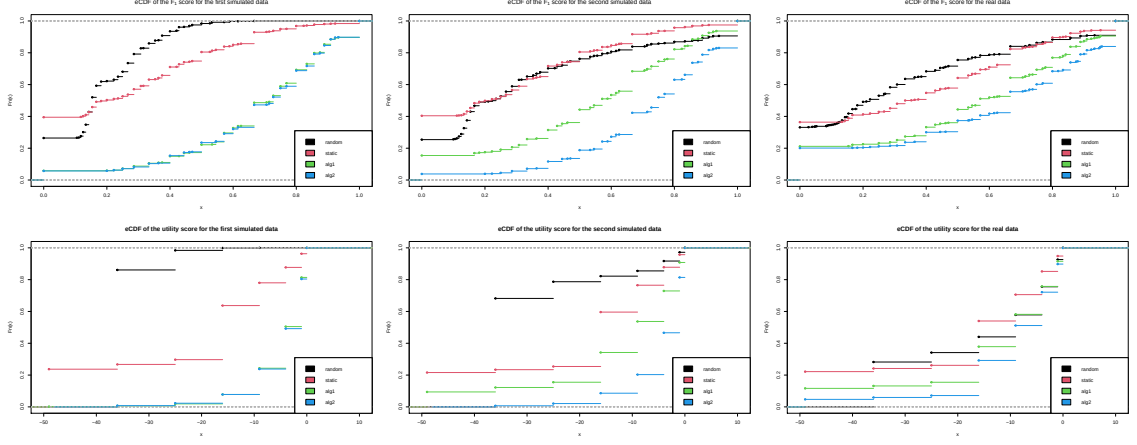


Figure 3.3: Empirical cumulative distribution functions (eCDF) of the F_1 score and the utility u_1 in both the simulated — first and second columns — and real data — third column.

Real data In order to investigate the performance of the algorithms in a real scenario, we consider a dataset collected during a tinnitus-focused study where moment-to-moment tinnitus symptoms were reported using a mobile application called TrackYourTinnitus (Probst et al., 2016). Since intraindividual fluctuations may be an important characteristic feature of an individual’s tinnitus and it is of great importance to investigate the behaviour changes while the individual is suffering from tinnitus, this dataset is of significant practical value to test our triggering algorithm.

From the original data, we considered only the *user_id*, *save*, *save_date* and *question_2* columns. These columns collect, respectively, the users identification number, the time and/or date on which the user interacted with the TrackYourTinnitus application and the self-reported tinnitus severity. The tinnitus severity is represented by a value between 0 and 1. From this subset of the original data, we delete all rows which have at least a missing value and all duplicates. For some subjects, different tinnitus severity values were reported by the same user at the same identical time; in those cases, we only consider the first-listed of such observations. Since both Algorithm 1 and 2 require a pre-fixed length of the experiment, we further restrict our attention to only the 30 days after the first interaction of each participant with the application. For most of the participants (77%, or 2130/2752) there is no information loss as they interacted with the application only within this time-frame.

Finally, as the starting point has been fixed to $S^* = 6$, we do not consider users who interacted less than 6 times with the application. Notice that, after cleaning the data, only around 33%, (911/2752) of the initial users remain.

The substantial data loss observed during the cleaning of the tinnitus dataset is a reflection of the well-documented challenge of participant adherence in EMA studies. While this reduction in sample size is a limitation, it is important to clarify that the scope of this work was specifically to improve the experimental experience and data quality for subjects demonstrating at least a baseline level of adherence. In this study, “slight adherence” is operationally defined as a minimum of six interactions ($S^* = 6$). This threshold was selected to balance the competing goals of maximizing participant inclusion against the mathematical necessity of obtaining stable parameter estimates via the method of moments. Nevertheless, it must be acknowledged that this approach introduces a potential selection bias, as the final results represent a cohort of more engaged users. This bias should be considered in any follow-up conclusions and remains a relevant factor for future applications of these methods, which will not be immune to the persistent problem of participant dropout.

Statistical analysis In order to compare the performances of the four algorithms, we consider two complementary measures: the F_1 score (Sokolova & Lapalme, 2009) and the estimated utility $u_1 := (\sum_{i=1}^N w_i - v)^2$. Both metrics serve to evaluate the burden placed on the subject relative to the information gained, but they provide different viewpoints on this trade-off. Moreover, they indirectly provide information on the usefulness of the data collected for the follow-up analysis of behavioural performance across states of high versus low symptom severity.

The utility function u_1 measures how effectively the algorithm achieves the goals set by the design optimization problem. Specifically, it assesses whether the number of successful triggers closely aligns with the desired target v (i.e. to trigger on average v additional tasks at extreme values of the self-reported variable). This not only informs the researcher whether they have collected sufficient relevant data for

subsequent analyses, but also controls for overburdening the subjects with excessive prompts or, conversely, failing to engage them sufficiently.

In contrast, the F_1 score summarizes the efficiency of the data collection process by explicitly accounting for false positives and false negatives. An improved F_1 score directly enhances the validity of the follow-up analysis by ensuring that the triggers effectively happened at high/low values of the self-reported variable. Rather than focusing solely on the absolute number of correct hits, the F_1 score evaluates the ratio of successful triggers against unnecessary or missed prompts, providing a clearer picture of the overall efficiency and the true burden experienced by the participant.

To compute the F_1 score, we measure the accuracy of the algorithm against a ground truth. This ground truth — the true high and low values of the self-reported variable for each subject — is defined using Equation (3.6). For this computation, the significance level κ^* is defined as in Equation (3.3), but the total number of data points N is known *a posteriori* and does not require an approximation. Consequently, the ground truth is tailored to the individual subject and varies across individuals (see Figure 3.1). We computed these measures for each individual in the first simulated dataset across all considered algorithms. The empirical cumulative distribution functions (eCDF) of these quantities are shown in Figure 3.3 (first column). The same analysis was performed for the second simulated dataset and the real data, as shown in the second and third columns of Figure 3.3.

As for each individual we have measures of F_1 score and utility for each of the four algorithms, we can utilise the Wilcoxon signed-rank test for paired data in order to compare the performances of the algorithms using the `wilcoxon.test` function in R (Brunner et al., 2018). Notice that the computation of the F_1 score produced some NaN, which were omitted in the statistical test. The results are shown in Table 3.1.

We find that, for both the simulated and the real data, Algorithm 2 tends to have larger F_1 scores and utility values u_1 than all the other options. In particular, Algorithm 2 performs as well or better than Algorithm 1 but, unlike the other option, does not rely on any prior knowledge of the population adherence. Moreover, in the

Table 3.1: Results of the Wilcoxon signed-rank test. In the first row, we find the p-values relative to the distribution of the F_1 score for the first simulated data. The notation $algorithm > algorithm'$ denotes that the distribution of the F_1 score of $algorithm$ is stochastically greater than the one of $algorithm'$. Equivalently, $algorithm$ tends to have higher F_1 scores than $algorithm'$. The notation $algorithm > algorithm'$ represents the alternative hypothesis of the test, which is accepted with the p-value listed in the Table. The same explanation applies to the second row of the table, which is relative to the distribution of the global utility u_1 for the first simulated data. The following rows are relative, respectively, to the second simulated data and the real data.

		static > random	alg1 > random	alg1 > static	alg2 > random	alg2 > static	alg2 > alg1
sim data 1	F_1	$4.2 \cdot 10^{-16}$	$2.3 \cdot 10^{-152}$	$1.2 \cdot 10^{-127}$	$6.7 \cdot 10^{-154}$	$8.8 \cdot 10^{-130}$	0.14
	u_1	$1.2 \cdot 10^{-91}$	$3.6 \cdot 10^{-166}$	$5.8 \cdot 10^{-120}$	$4.2 \cdot 10^{-165}$	$2.0 \cdot 10^{-119}$	0.22
sim data 2	F_1	1	$1.9 \cdot 10^{-34}$	$2.3 \cdot 10^{-68}$	$3.9 \cdot 10^{-123}$	$1.9 \cdot 10^{-133}$	$3.6 \cdot 10^{-62}$
	u_1	$4.8 \cdot 10^{-54}$	$1.2 \cdot 10^{-102}$	$3.6 \cdot 10^{-28}$	$1 \cdot 10^{-148}$	$1.6 \cdot 10^{-117}$	$1.2 \cdot 10^{-73}$
real data	F_1	$3.4 \cdot 10^{-4}$	$7.6 \cdot 10^{-29}$	$1.4 \cdot 10^{-30}$	$4.1 \cdot 10^{-46}$	$5.6 \cdot 10^{-42}$	$1.2 \cdot 10^{-9}$
	u_1	1	$2.4 \cdot 10^{-2}$	$2.1 \cdot 10^{-21}$	$4.5 \cdot 10^{-16}$	$3.3 \cdot 10^{-46}$	$1.9 \cdot 10^{-13}$

comparison between Algorithms 2 and 1 we obtain smaller p-values in the second simulated case compared to the first one, see the last column of Table 3.1. This suggests that the improvement — both for the F_1 score and the utility u_1 — brought by Algorithm 2 compared to Algorithm 1 is more evident in the second simulated data and less remarkable in the first simulated data. This discrepancy may be caused by the greater variability of adherence for the second simulated data compared to the first, see Figure 3.2. The p-values obtained for the real data are also consistent with this hypothesis. Indeed, the real data has higher variability than the first simulated data but, unlike the second simulated data, it was collected in a non-ideal context. This suggests that the additional complexity of Algorithm 2 compared to Algorithm 1 is more “useful” for studies where there is a high variability in the adherence of the subjects.

Practical considerations Both Algorithms 1 and 2 assume the self-reports follow a Beta distribution and consist of independent and identically distributed (i.i.d.) samples. This implies that the data must be bounded between 0 and 1. *A posteriori* checks using histograms can verify this behaviour. Additionally, autocorrelation function (ACF) plots can help ensure no time dependency is present, satisfying the i.i.d. assumption. Similarly, subject adherence — meaning the subject’s interactions

Diagnostics & practical deployment checklist

1. Assumptions

- **I.I.D. Beta distribution:** self-reports are independent and bounded.
Diagnostic: histogram fit for distribution; autocorrelation (ACF) plots to rule out temporal dependence.
- **I.I.D. adherence:** adherence rate remains constant.
Diagnostic: Track exponential drop-offs.
- **Estimator validity:** method of moments Condition 3.1 holds.
Diagnostic: flag sequences of identical reports (zero-variance) and select appropriately S^* .

2. Practical deployment checklist

- **Targeted to intensity:** the algorithms are targeted to collect data at symptoms intensity extremes, and might miss insights related to frequency or other factors.
- **Known *a priori* (by practitioner):** target triggers v , experiment duration N and triggering starting point S^* .
- **Unknown *a priori* (to practitioner and algorithm):** subject-specific distributions' parameters (δ, ξ) and the adherence rate χ .
- **Real-world impact:** unknown parameters are estimated on-the-fly in Algorithms 1 and 2. Known parameters require careful experimenter calibration to balance subject burden, data sufficiency, and estimator stability.

with the app — is assumed to follow an i.i.d. binomial distribution with a constant adherence rate over time, which can be verified using similar diagnostics. Note that these checks can only be performed *a posteriori*, after data collection. Consequently, pilot studies may be necessary to ensure the algorithms are suitable for specific settings. While integrating online diagnostic checks during the data collection process would greatly benefit long experiments or pilot studies to ensure the protocol is proceeding as planned, developing this feature is outside the scope of this work.

Additionally, the algorithms assume that behavioural changes are driven by symptom severity, defined as the deviation of self-reports from a subject's mean. Consequently, the triggering mechanism targets extreme severity values. If this

assumption proves incorrect, we anticipate a null result in subsequent task performance comparisons. While the algorithms primarily monitor symptom intensity, they also adapt to frequency by tracking whether a subject’s behaviour becomes more erratic or stable over time. Nevertheless, the algorithms are not designed to identify good triggering points if behavioural shifts are driven purely by variance or factors unrelated to intensity, thus potentially missing some good data spots in these cases. This introduces a common trade-off in non-random data collection studies whereby targeting a specific setting we preclude a potential more general analysis. Ultimately, the algorithm works as intended as long as severity has even a small impact on behaviour. Moreover, further studies might extend the algorithms to use variance - or other factors - if that aligns better with the purposes of the researcher’s main research question.

Unlike the static algorithm, the adaptive algorithms introduced in this work determine parameters dynamically. The only parameters the experimenter must set for the runs are S^* , v , and N . The specific adherence rate χ and the parameters (δ, ξ) of the underlying distribution are assumed unknown *a priori* and are estimated dynamically as new data points are collected. This adaptability is designed to better tailor the experiments to real-world variability during data collection. Consequently, the selection of the predefined parameters S^* , v , and N becomes more relevant. The experimenter must balance the burden a high number of triggers v might cause subjects against ensuring they have enough data at the end of the experiment N for follow-up analysis. Furthermore, selecting the starting point S^* requires a critical trade-off: an S^* that is too small risks estimator crashes in the method of moments, whereas an S^* that is too large delays the triggering schedule, risking the loss of the subject’s active adherence window — see also Section [3.6.1](#).

3.5 Discussion

In conclusion, we have presented a valid alternative to the current state-of-the-art sampling methods for EMA studies which aim to identify high and low values of a

self-reported variable to trigger a lengthier and more burdensome task. We developed Algorithms 1 and 2, which have been proven effective to collect data in this context. Indeed, they both outperform static and random triggering approaches in a simulated scenario. The same pattern is confirmed in the real tinnitus severity data. While Algorithm 1 is easier to implement than Algorithm 2, the latter does not rely on prior knowledge of the population adherence and achieves similar or better results. Moreover, the analysis suggests that Algorithm 2 is slightly more effective than Algorithm 1 in increasing data quality (obtaining a desired number of additional tasks for a higher value of u_1) while preserving precision (similar F_1 scores), especially for studies with high variability in adherence. This is especially important as many fields, such as the tinnitus research area, rely on self-reports for which adherence is a challenging problem.

Algorithms 1 and 2 can be easily adapted to various contexts, but still rely on strong statistical assumptions which may not hold in real scenarios. It is possible to generalize these algorithms to other distributions: arguments in this work remain valid, with adequate changes to the notation, if the quantity of interest X follows an absolutely continuous distribution. Nevertheless, it is still necessary to have easy access to the quantile of the distribution in order to effectively deploy the algorithm. On the other hand, we assume that the distribution of the quantity of interest X is known *a priori*. In addition, the quantity of interest and the adherence of the subjects are assumed independent and identically distributed even though time is often an important factor in longitudinal studies. Despite these limitations, it can be easily adapted to the necessity of the researcher and we have demonstrated its effectiveness through simulations and large-scale real-world analysis.

Future work There are many possible ways to further develop the arguments in this work. Using statistical models which consider time for both the quantity of interest and the adherence such as autoregressive models could improve the performance of the algorithm in real case scenarios (Fricker et al., 2008). Moreover, more than one quantity of interest is often collected during a longitudinal study

(Probst et al., 2016) and there may be an interest in developing a multivariate version of the algorithms (Chandola et al., 2009). In particular, including objective measures, such as measurements from wearable sensors, may increase the effectiveness of the algorithms (MacDonald et al., 2006). In this direction, hierarchical methods could inspire future work as they have been successfully applied in the field of anomaly detection (Sottas et al., 2011). Finally, the algorithm developed in this work provides an adaptive sampling schedule for deployment of a secondary burden-heavy task contingent on data from a lighter, more frequent data collection based on self-reports. Nevertheless, it does not address the sampling scheme for the self-reports. An adaptive sampling scheme which predicts the unlikely states and sends a notification to the participants only when needed could further increase the balance between data quality and the burden placed on the subjects (Hulme et al., 2021).

3.6 Supplementary material

This section contains supplementary material on the implementation of the algorithms presented in this chapter.

3.6.1 Practitioner guidelines for the selection of the starting point

There are two competing goals that require the attention of the practitioner in the selection of the starting point S^* . On one hand, S^* has to be large enough so that the method of moments estimation provides a “good enough” estimate of the parameters and does not encounter computational errors. A sufficiently large number of observations is necessary to obtain a “good” estimate through the method of moments. Indeed, as the sample size increases, the estimation of the mean squared error decreases, see Figure 3.4. Notice that if we start collecting the additional task data at time S^* , then the smallest sample size considered by Algorithms 1 and 2 is $S^* - 1$. Hence, the preference for a large S^* .

Moreover, in order to avoid computational errors, the condition

$$0 < \hat{\sigma}_s^2 < \hat{\mu}_s(1 - \hat{\mu}_s). \quad (3.11)$$

has to be satisfied for all $s \in \{S^* - 1, \dots, N\}$, where $\hat{\sigma}_s^2$ and $\hat{\mu}_s$ are the method of moments estimates of the variance and expected value of a Beta distributed sub-sample $\mathbf{x}_{1:s}$. This condition is always satisfied if the estimated expected value and variance are replaced by their true values. Assuming that the estimates $\hat{\sigma}_s^2$ and $\hat{\mu}_s$ tend to the true values σ^2 and μ^2 as s tends to infinity — as happens for good estimates, then we expect Condition (3.11) to hold for large data samples — i.e. large values of S^* , see also Figure 3.4. Nevertheless, because of the structure of

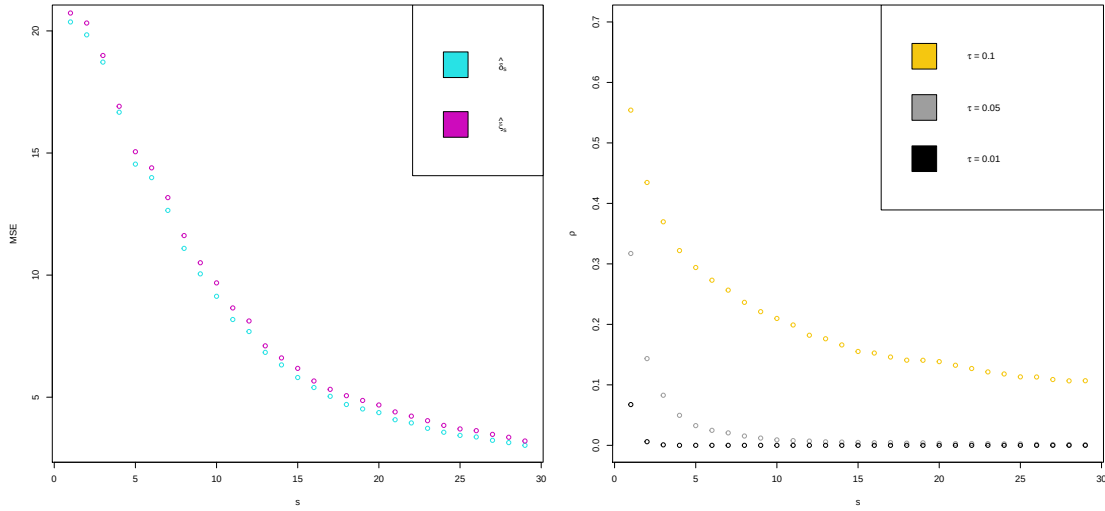


Figure 3.4: On the left, a simulation of the mean squared error (MSE) for different data sample sizes in the method of moment estimation of the Beta distribution parameters. For each sample size s , we simulated 5000 pair of Beta distribution parameters from a uniform distribution in $[0.5, 10]$. Then, for each pair, s observations were sampled and the method of moments estimates computed. As expected, the MSE of these estimates decreases as the sample size increases. On the right, an analysis of the prevalence of data which violates the left hand side of Condition (3.11) in the same simulation setting. In particular, ρ is the approximated probability that the variance for a sample of size s from a Beta distribution is smaller than τ . We notice that ρ decreases as the sample size increases.

both Algorithms 1 and 2, it is necessary to estimate the parameters for small sample sizes. This may result in computational problems both in a simulated and real setting as the likelihood of obtaining a constant — or almost constant — sequence of observations is increased. In those cases, the sample variance is zero — or almost zero, violating the left hand side of Condition (3.11). This causes the estimates to either be undefined or explode. Alternatively, the right hand side of Condition (3.11) may not be satisfied, producing negative estimates for the parameters. This problem is non-trivial as the experimental design considered and the real data analysed can contain sequences of all 0 and 1 — or values too close to them. Then, both $\hat{\sigma}_s^2$ and $\hat{\mu}_s$ tend to zero, sometimes resulting in a violation of Condition (3.11), see also Figure 3.6 and Section 3.6.3. We may obtain too small estimates of the variance and of the right hand expression of Condition (3.11). All these suggest that a greater S^* is less likely to produce errors in the code.

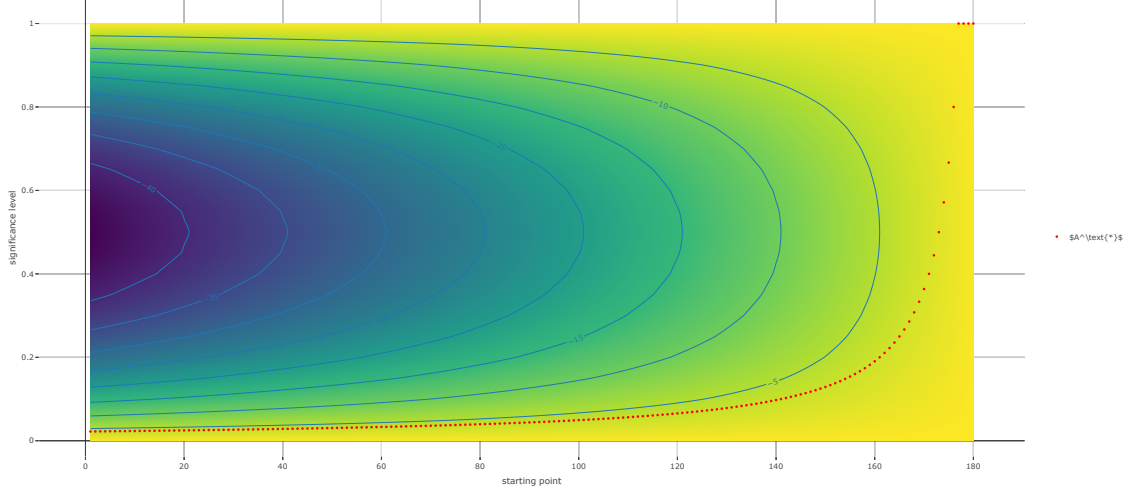


Figure 3.5: The red dots represent the set A^* computed for an experiment of length $N = 180$ and assumed $v = 4$. Notice that, as a function of the optimal starting point S^* , the optimal significance level κ^* increases slowly in the first half of the experiment. This suggests that any starting point S in the first half of the experiment should not impact significantly the final performance of the algorithm for completely adherent subjects. The extension of the function U_2 in Equation (3.12) into the domain $D = [2, N] \times [0, 1]$ is represented by the contour plot.

On the other hand, S^* needs to be small enough so that the risk of collecting too little data for non-adherent subjects is low. As the adherence of the subjects decreases exponentially (as it is shown in Figure 3.2), an early starting point increases the chances of obtaining enough data for a larger number of participants. The final selection of S^* needs to achieve a balance between these two goals. In this work, we set $S^* = 6$. Plotting the optimal region A^* can further aid the researcher in the selection of the experimental design and highlight how the choice of the starting point impacts the significance of the following triggers, see Figure 3.5.

3.6.2 Stopping rule

Reducing the variance of the final number of triggers V is a second meaningful objective when computing the optimal significance level κ^* : indeed, once the mean has been fixed, this concentrates the values of V in proximity of $E(V)$. Recall that $\text{Var}(V) = (N - S + 1)\kappa(1 - \kappa)$ and define a second expected utility function

Table 3.2: In the burden side of the Table we report the proportion of subjects for which at least one additional task has been prevented by a stopping rule set to 10. In the information loss side of the Table we report the proportion of subjects for which at least a false positive or a false negative was introduced by the same stopping rule.

	Burden			Information loss		
	static	alg1	alg2	static	alg1	alg2
Sim data 1	0.218	0.002	0	0.219	0.001	0
Sim data 2	0.201	0.073	0	0.201	0.064	0

$U_2 : \{2, \dots, N\} \times [0, 1] \longrightarrow \mathbb{R}$ as

$$U_2(S, \kappa) = -\text{Var}(V). \quad (3.12)$$

A contour plot of the continuous extension of the function U_2 into the domain $D = [2, N] \times [0, 1]$ is given in Figure 3.5.

As the variance is only a secondary goal, we consider the design optimization problem for U_2 constrained to the domain A^* , i.e.

$$(S', \kappa') \in \arg \max_{(S^*, \kappa^*) \in A^*} U_2(S^*, \kappa^*). \quad (3.13)$$

Then,

$$\begin{aligned} U_2(S^*, \kappa^*) &= -(N - S^* + 1)\kappa^*(1 - \kappa^*) \\ &= -v(1 - \kappa^*) \end{aligned}$$

where the second-last step is justified by the constraint $(S^*, \kappa^*) \in A^*$. Thus, the optimal solution for the optimization problem in Equation (3.13) is $(N - v, 1)$. This design corresponds to triggering the additional task in the last v interactions of the subject with the app. There are two main problems with this approach: first, in real applications, there is no guarantee that the subject will adhere to the study until the end; second, in order to trigger the additional task on significantly high or low values x_t we aim to implement κ as small as possible.

In conclusion, reducing the variability of the total number of triggers produced by the algorithm results in a trivial solution for the design optimization problem. Nevertheless, the argument above shows that there is a trade-off between the need for precision — decreasing the variance of the total number of observations — and the need for unlikeliness — keeping the significance level κ small. In this chapter we assume that both S^* and, subsequently, κ^* are chosen to be relatively small. This can result in high variance for the final number of triggers. In other words, the algorithm may trigger either too many or too few tasks for certain subjects.

Hence, we suggest to add some simple deterministic rules to either force an additional task if there were not enough or stop their triggering if too many have already been triggered. In this chapter, we added a stopping rule which, for each subject, prevented the algorithm from triggering the additional task more than 10 times. In Table 3.2, we show that, at least in a simulated context, this choice has almost no impact on Algorithms 1 and 2.

This approach aligns with several practical considerations. The cap of $v = 10$ triggers serves as a "burden budget" to ensure the experimental design remains approachable for subjects. Without this cap, all schedules would risk “spamming” participants with an excessive number of triggers during outlier events. Thus, this constraint reflects the real-world necessity of minimizing participant fatigue to maintain adherence. The stopping rule is particularly important for the static case, since it lacks the capacity of Algorithms 1 and 2 to adapt to unusually frequent high or low self-reports. Thus, this limit ensures the static algorithm collects at least as much data as the adaptive cases (which target 4 triggers) while not overburdening the subjects. This provides a conservative and robust basis for comparison of the static algorithm with the Algorithms 1 and 2 in the settings of this work and a reasonable experimental design for the static algorithm in real-world settings as well.

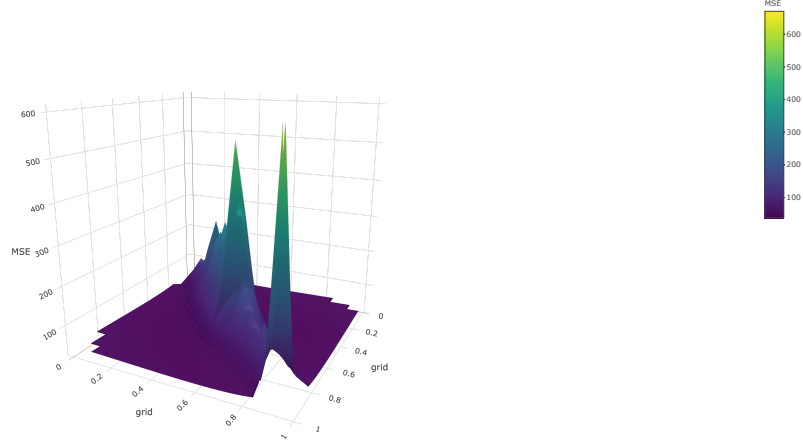


Figure 3.6: Sensitivity analysis for the impact of the artificial data on the Beta parameter estimation. The figure was obtained by considering 1624 simulated samples of length 5 for which Condition (3.11) is not satisfied. We only considered samples of length 5 as it is the smallest sample size considered in this work, i.e. $S^* - 1 = 5$, and Condition (3.11) is more likely to not be satisfied for small sample sizes. The mean squared error of the Beta parameter estimation obtained by adding artificial — taken from a discretization of the set $[0, 1]^2$ — was computed. The figure shows the sum of the MSE for both the first and second parameters.

3.6.3 Other computational problems

A careful selection of a starting point S^* following the instructions in Section 3.6.1. partially solves the computational problems of the estimation of the Beta parameters. Nevertheless, we could still occasionally encounter computational errors due to a violation of Condition (3.11). The probability of this happening converges to zero, but it is always non-zero for a finite sample. Therefore, only when Condition (3.11) is not satisfied, we added in our code some artificial data to regularize the problem. Since one artificial data point may not be enough to increase the variance of the sample — a subject may report a constant sequence of exactly that value — we decided to add two artificial observations from the set $[0, 1]^2$.

To show how this choice impacts the estimation of the parameters, we conducted a sensitivity analysis. We show the results in Figure 3.6. We conclude that points placed on the corners of $[0, 1]^2$ and on the diagonal $\{(h, h) \in [0, 1]^2\}$ should be avoided. Points in the corners do not guarantee that the right hand side of Condition

(3.11) is satisfied since we increase the likelihood of obtaining a sequence of only 0 or 1 — or a sequence of values too similar to it. On the other hand, points close to the diagonal result in a higher mean squared error. In this work, we used the artificial data 0.4 and 0.6.

Chapter 4

Improving data collection through an inhomogeneous MCMC algorithm

In Chapter 3, we analysed how DO methods can improve data collection by carefully selecting external experimental design settings, such as the time of task administration. In this chapter, we focus on the application of DO methods to internal factors in MABT experiments, in particular, the reward probabilities.

We investigate the selection of an optimal design for a stationary two-armed bandit task using a statistical method proposed by Müller et al. (2004) based on inhomogeneous Markov chain simulation. This algorithm has been used in various studies. For example, simulations show its effectiveness in reducing the number of trials necessary to assess memory performance (Cavagnaro et al., 2010). In Stroud et al. (2001), the authors rely on this same approach to compute optimal sampling times in a pharmacological context. Finally, this approach has been used to solve sequential games (Ekin et al., 2023).

This algorithm is of particular interest as it merges the evaluation and exploration steps of the MCMC optimization in an elegant theoretical framework and allows for a sensitivity analysis at no additional computational cost. The algorithm simultaneously searches the design variable space through an MCMC algorithm and solves the optimization problem. Unlike a grid-based approach, this method focuses only on the

most relevant regions of the design variable space. Furthermore, as the utility of the DO problem is equal — up to a multiplying constant — to the selection probability of the MCMC algorithm due to the augmented probability method, a sensitivity analysis is also a regret analysis for the selection of the best experimental design over its competitors. This enables an investigation of the regret and relative uncertainty in choosing near-optimal solutions and allows quantification of the loss of choosing a suboptimal solution over an optimal one.

A high number of trials is usually required to guarantee adequate parameter estimation in these settings. However, DO methods have the potential to decrease these requirements through a more careful data collection schedule. Subsequently, this also decreases the burden placed on the subjects due to shorter experiments and ensures better data quality as participants' attention remains more stable when experiments remain engaging.

We find that DO methods effectively identify and avoid under-performing designs in MABTs for TD learning models, where suboptimal experimental designs can more than double parameter estimation errors. Although several near-optimal designs exist — making the selection of an adequate design likely — there are also severely inefficient designs that have a detrimental impact on the results. Consequently, DO methods can safeguard experimental design selection and ensure stable results. This is especially true for shorter experiments, where the impact of experimental design selection is most noticeable and data quantity cannot compensate for a poor design choice.

The work is composed of three main steps. First, we introduce the experiment in Section 4.1. Secondly, we select a probability model and set up a parameter estimation procedure in Section 4.2. Then, we describe the design optimization framework and the optimization approach, see Section 4.3. The results and some considerations on the usefulness of this approach are reported in Section 4.4. Lastly, in Section 4.5, we discuss limitations and future research directions for this work.

The code and the simulated data used in this chapter are available in the GitHub

repository (<https://github.com/gretamonacelli/docmab/releases/tag/v1.0.0>).

4.1 The experiment

MABT experiments provide a flexible mathematical framework for researchers to develop new experiments for the study of human decision-making. Firstly, they are intuitive to deliver. Secondly, while some technical knowledge is needed for their implementation, the use of mobile applications for the submissions of such experiments has the potential to deliver them to a wide audience. Several studies on the digital delivery of these experiments have already been conducted (McCarthy et al., 2021).

While already covered in Chapter 2, we briefly remind the reader of some information about MABT experiments. In their initial description, an MABT required a subject to choose sequentially between multiple slot machines and pull their "arm" to receive the respective rewards. The reward was then sampled from a probability distribution relative to the selected slot machine. In computational psychology, it is uncommon to present the options to the subjects as slot-machines, but the probabilistic nature of the rewards remains unchanged and the terminology of "arms" persisted.

The notation for this experiment is as follows. During the task, a subject is asked to choose between two arms for a fixed $T \in \mathbb{N}^+$ number of trials. Each arm gives a reward according to a probability distribution. In particular, in this work, we model the first and second arm rewards as two Bernoulli random variables $S_1 \sim \text{Ber}(p_1)$ and $S_2 \sim \text{Ber}(p_2)$, with $p_1, p_2 \in [0, 1]$. If the subject chooses the first arm, we sample from S_1 , and vice versa. The subject obtains either a positive reward — equal to 1 — or no reward — equal to 0. The goal of the subject is to maximise the final amount of rewards by the end of the experiment.

The sequence of choices made by a subject is represented as a random vector $\mathbf{C} = (C_1, \dots, C_T)$. Each component C_t is a Bernoulli random variable which takes value in $\{1, 2\}$ and represents the arm chosen by the subject at time $t \in \{1, \dots, T\}$.

The reward obtained by the subject at time t depends on their choice C_t and can be described as

$$R_t = S_1 \mathbb{1}_{\{C_t=1\}} + S_2 \mathbb{1}_{\{C_t=2\}}. \quad (4.1)$$

Then, the random vector $\mathbf{R} = (R_1, \dots, R_T)$ represents the sequence of rewards obtained by the subject during the whole experiment.

By the end of the experiment, the researcher has collected the vectors of observations $\mathbf{c}$ and $\mathbf{r}$. The number of trials T and the probabilities p_1 and p_2 are assumed known *a priori* by the experimenter. We assume that subjects are completely adherent. This dataset forms the input for the subsequent modelling process.

4.2 Learning model

Because subjects are asked to sequentially make choices during these tasks, MABT experiments are suited to study decision-making over time and the exploration-exploitation balance in subjects' learning strategies. Consequently, the statistical models used to describe data collected through these tasks are frequently referred to as *learning models*. These models reduce the subjects' behaviour to a lower-dimensional representation, summarising it in just a few parameters. These parameter estimates can be interpreted as a phenotypic signature. Encouraging results indicate that these parameters can be used to differentiate between groups, linking different performances in MABT experiments to different psychiatric conditions — for example, healthy subjects versus subjects affected by anxiety disorder (Aylward et al., 2019). This motivates the use of these learning models in the search for new candidate biomarkers in psychiatry.

In the design optimization framework, the learning model which describes MABT experiment data is assumed known *a priori*. This allows the generation of simulated data and the evaluation of the local utility to identify the optimal designs. The model is usually described as $p(\mathbf{y}|\boldsymbol{\theta})$ where $\mathbf{y}$ is the data and $\boldsymbol{\theta}$ is the vector of the model parameters. In the case of MABT experiments, $p(\mathbf{y}|\boldsymbol{\theta}) = p(\mathbf{c}, \mathbf{r}|\boldsymbol{\theta})$.

As the MABT framework allows the generation of a huge number of experimental designs depending on the specific goals of the researcher, multiple computational models have been developed to analyse the specific data produced and highlight particular aspects of decision-making behaviour, see Chapter 2. Due to the variety of possible experimental designs and models possibly implemented, research to develop both these aspects of decision-making experiments based on MABTs is still ongoing.

In this work, we consider three different models. The first model is a simple TD model, very similar to the model applied in Melinscak and Bach (2020). The second model is an extension of the first, as it additionally assumes that subjects react differently to wins and losses (Lin et al., 2020). Both these models assume that the knowledge accumulated during the experiment up to the current trial influences the next choice of a subject. On the other hand, the last model — often called the stochastic win-stay lose-shift (S. Zhang & Lee, 2010) — assumes that only the immediate precedent choice and reward impact the next choice of the subject. In this work, we will refer to these models, respectively, as Model 1, Model 2 and Model 3.

Model 1 Similar to current research in the field (Melinscak & Bach, 2020; Valentin et al., 2024), we utilise a TD algorithm with a softmax rule inspired by the Rescorla-Wagner model. The model utilises two parameters: the *learning rate* $\alpha \in [0, 1]$ and the *inverse temperature* $\tau \in [0, +\infty)$. The first assumption of this model is that the choices of the subjects only depend on the model parameters and the past history of the subject in the experiment. Therefore, the probability that a subject at time t makes a choice c_t and obtains a reward r_t is described by the following equation:

$$p(r_t, c_t | \mathbf{r}, \mathbf{c}, \alpha, \tau) = p(r_t, c_t | \mathbf{r}_{1:t-1}, \mathbf{c}_{1:t-1}, \alpha, \tau).$$

We further assume that the dependency on past experience is expressed through

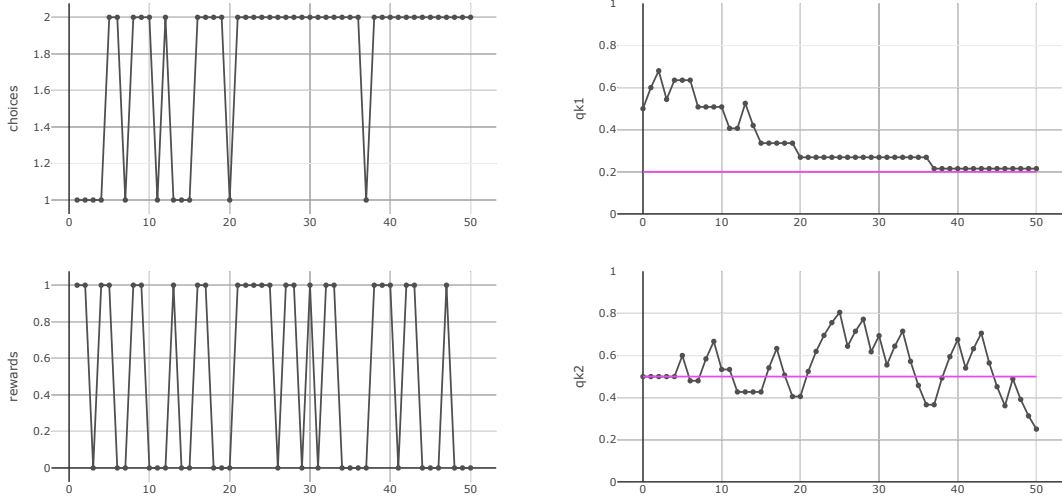


Figure 4.1: Example of a simulated 2ABT experiment for Model 1. The learning rate is fixed to 0.2, and the inverse temperature parameter is equal to 15. The reward probabilities for the arms are fixed to 0.2 and 0.5, respectively. We plot the choices, the rewards, and the empirical estimates $Q_{1,\cdot}$ and $Q_{2,\cdot}$ of the reward probabilities for the simulated subject.

cumulative values $Q_{1,t-1}$ and $Q_{2,t-1}$ defined as follows:

$$Q_{1,j+1} = Q_{1,j} + \alpha(r_j - Q_{1,j})\mathbb{1}_{\{c_j=1\}}$$

$$Q_{2,j+1} = Q_{2,j} + \alpha(r_j - Q_{2,j})\mathbb{1}_{\{c_j=2\}}$$

$$Q_{1,0} = 0.5$$

$$Q_{2,0} = 0.5.$$

These values represent the subjects' rough estimation of the reward probability of each arm — p_1 and p_2 — at time $j + 1$. The learning rate α is then the weight the subject gives to the most recent trial while updating their estimation. We store these values in a $2 \times (T + 1)$ -dimensional matrix Q . Notice that Q depends on the history of both the choices $\mathbf{c}$ and the rewards $\mathbf{r}$ of a subject, see Figure 4.1.

Let $\mathbf{c}_{1:t}$ and $\mathbf{r}_{1:t}$ be the sub-vectors of $\mathbf{c}$ and $\mathbf{r}$ with only the first t entries. Then, the learning model assumes

$$\begin{aligned} p(r_t, c_t | \mathbf{r}, \mathbf{c}, \alpha, \tau) &= p(r_t, c_t | \mathbf{r}_{1:t-1}, \mathbf{c}_{1:t-1}, \alpha, \tau) \\ &= p(r_t, c_t | Q_{1,t-1}, Q_{2,t-1}, \alpha, \tau). \end{aligned}$$

Since the rewards are sampled independently from the past history of the experiment due to our definition of the experimental design, we obtain that

$$\begin{aligned}
p(r_t, c_t | \mathbf{r}, \mathbf{c}, \alpha, \tau) &= p(r_t, c_t | \mathbf{r}_{1:t-1}, \mathbf{c}_{1:t-1}, \alpha, \tau) \\
&= p(r_t, c_t | Q_{1,t-1}, Q_{2,t-1}, \alpha, \tau) \\
&= p(r_t | c_t) p(c_t | Q_{1,t-1}, Q_{2,t-1}, \alpha, \tau).
\end{aligned}$$

The model is completed by a softmax decision rule, which is used to model the selection of the next arm and introduces the inverse temperature parameter τ . We write

$$p(c_t | Q_{1,t-1}, Q_{2,t-1}, \alpha, \tau) = \begin{cases} \frac{\exp(\tau Q_{1,t-1})}{\exp(\tau Q_{1,t-1}) + \exp(\tau Q_{2,t-1})} & \text{if } c_t = 1 \\ \frac{\exp(\tau Q_{2,t-1})}{\exp(\tau Q_{1,t-1}) + \exp(\tau Q_{2,t-1})} & \text{if } c_t = 2. \end{cases} \quad (4.2)$$

The inverse temperature parameter measures the exploration/exploitation behaviour of a subject. In particular, large values of τ indicate that the subject tends to exploit prior knowledge, while for small values the subject tends to explore the design variable space more.

In order to apply the design optimization method, it is necessary to estimate the parameters α and τ . The likelihood function is of the form

$$\begin{aligned}
L(\alpha, \tau) &= p(\mathbf{r}, \mathbf{c} | \alpha, \tau) \\
&= \prod_{t=1}^T p(r_t | c_t) p(c_t | Q_{1,t-1}, Q_{2,t-1}, \alpha, \tau) \\
&= \prod_{t=1}^T p(r_t | c_t) \left[\frac{\exp(\tau [Q_{1,t-1} \mathbb{1}_{\{c_t=1\}} + Q_{2,t-1} \mathbb{1}_{\{c_t=2\}}])}{\exp(\tau Q_{1,t-1}) + \exp(\tau Q_{2,t-1})} \right].
\end{aligned}$$

Taking the logarithm, we obtain that

$$\begin{aligned}
l(\alpha, \tau) &= \sum_{t=1}^T \log(p(r_t | c_t)) + \sum_{t=1}^T \log \left(\frac{\exp(\tau [Q_{1,t-1} \mathbb{1}_{\{c_t=1\}} + Q_{2,t-1} \mathbb{1}_{\{c_t=2\}}])}{\exp(\tau Q_{1,t-1}) + \exp(\tau Q_{2,t-1})} \right) \\
&= \sum_{t=1}^T \log(p(r_t | c_t)) + \tau \sum_{t=1}^T [Q_{1,t-1} \mathbb{1}_{\{c_t=1\}} + Q_{2,t-1} \mathbb{1}_{\{c_t=2\}}] - \\
&\quad \sum_{t=1}^T \log(\exp(\tau Q_{1,t-1}) + \exp(\tau Q_{2,t-1})).
\end{aligned}$$

The term $\sum_{t=1}^T \log(p(r_t|c_t))$ can be ignored in the MLE as it does not depend on α or τ . We denote the MLE estimates as

$$(\hat{\alpha}, \hat{\tau}) = \arg \max_{(\alpha, \tau) \in [0, 1] \times [0, +\infty)} l(\alpha, \tau). \quad (4.3)$$

To avoid computational problems, we restrict the domain of the inverse temperature parameter to $[0, 100]$ and use a bounded version of the optimization function `optim` for maximum likelihood estimation in R (Team, 2023). Restricting the parameter space to a compact set ensures the existence of at least one parameter pair that maximises the likelihood. Furthermore, this constraint prevents extreme estimates of the inverse temperature parameter, which are common in softmax reinforcement learning models, and facilitates reliable numerical optimization.

Model 2 In Model 1, subjects react to losses and wins equally. This assumption is not often true for real subjects. Indeed, people tend to give more weight to losses than to wins; to tackle this issue, Lin et al. (2020) proposed a model, always inspired by the Rescorla-Wagner model, which considers wins and losses separately in the learning process.

This model — called the double-stream model — uses three parameters: the learning rate relative to wins $\alpha_w \in [0, 1]$, the learning rate relative to losses $\alpha_l \in [0, 1]$ and the inverse temperature parameter $\tau \in [0, +\infty)$. The probability that a subject at time t makes a choice c_t and obtains a reward r_t is described by the following equation:

$$p(r_t, c_t | \mathbf{r}, \mathbf{c}, \alpha_w, \alpha_l, \tau) = p(r_t | c_t) p(c_t | Q'_{1,t-1}, Q'_{2,t-1}, \alpha_w, \alpha_l, \tau),$$

where $Q'_{1,t-1}, Q'_{2,t-1}$ are defined as

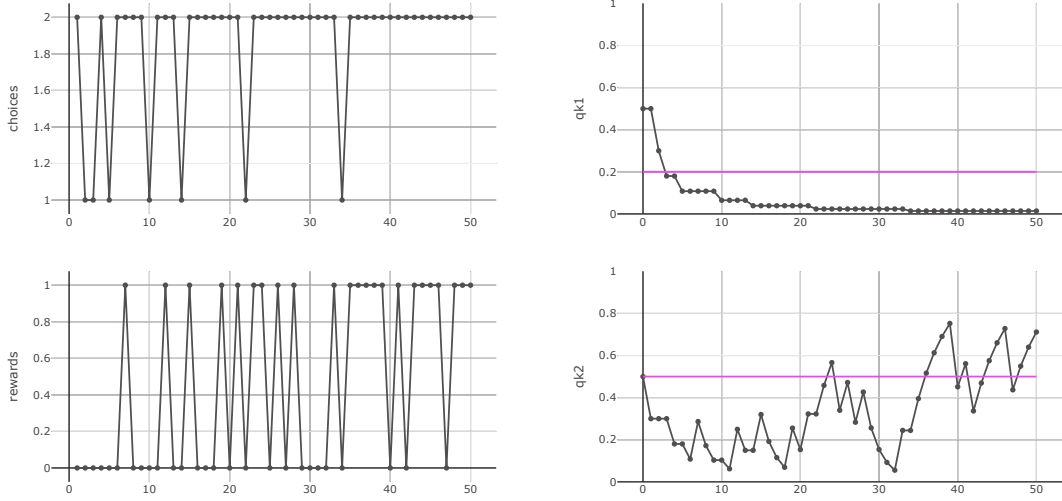


Figure 4.2: Example of a simulated 2ABT experiment for Model 2. The learning rate relative to wins is fixed to 0.2, while the learning rate for losses is equal to 0.4. The inverse temperature parameter is equal to 15. The reward probabilities for the arms are fixed to 0.2 and 0.5 respectively. We plot the choices, the rewards and the empirical estimates $Q'_{1,\cdot}$ and $Q'_{2,\cdot}$ of the reward probabilities for the simulated subject.

$$Q'_{1,j+1} = Q'_{1,j} + \alpha_w(r_j - Q'_{1,j})\mathbb{1}_{\{c_j=1, r_j=1\}} + \alpha_l(r_j - Q'_{1,j})\mathbb{1}_{\{c_j=1, r_j=0\}}$$

$$Q'_{2,j+1} = Q'_{2,j} + \alpha_w(r_j - Q'_{2,j})\mathbb{1}_{\{c_j=2, r_j=1\}} + \alpha_l(r_j - Q'_{2,j})\mathbb{1}_{\{c_j=2, r_j=0\}}$$

$$Q'_{1,0} = 0.5$$

$$Q'_{2,0} = 0.5,$$

with $j = 1, \dots, T - 1$. See also Figure 4.2 for a representation of Q' .

This softmax rule is

$$p(c_t | Q'_{1,t-1}, Q'_{2,t-1}, \alpha_w, \alpha_l, \tau) = \begin{cases} \frac{\exp(\tau Q'_{1,t-1})}{\exp(\tau Q'_{1,t-1}) + \exp(\tau Q'_{2,t-1})} & \text{if } c_t = 1 \\ \frac{\exp(\tau Q'_{2,t-1})}{\exp(\tau Q'_{1,t-1}) + \exp(\tau Q'_{2,t-1})} & \text{if } c_t = 2. \end{cases}$$

Following similar steps to those executed for Model 1, we obtain the following log-likelihood

$$l(\alpha_w, \alpha_l, \tau) = \text{constant} + \tau \sum_{t=1}^T [Q'_{1,t-1} \mathbb{1}_{\{c_t=1\}} + Q'_{2,t-1} \mathbb{1}_{\{c_t=2\}}] - \sum_{t=1}^T \log(\exp(\tau Q'_{1,t-1}) + \exp(\tau Q'_{2,t-1})).$$

We denote the MLE estimators as $\hat{\alpha}_w, \hat{\alpha}_l$ and $\hat{\tau}$. Similar to Model 1, the domain of the inverse temperature parameter is restricted to $\tau \in [0, 100]$ for computational reasons.

Model 3 There are simpler models which derive from the reinforcement learning literature, such as the ones used in Valentin et al. (2024) and S. Zhang and Lee (2010). The last model we analyse in this work belongs to this category and is called stochastic win-stay lose-shift. As the name suggests, the model assumes that a subject will select the same arm as they chose on the previous trial if they just received a reward and switch to the other arm if they just suffered a loss. An example of data simulated through this model is shown in Figure 4.3. Some additional white noise is added to the subject's behaviour, which is described by the only parameter of the model. Unlike the two previous models, Model 3 does not encode the whole past experience of a subject but only the previous step. This makes the model much more computationally efficient, but likely less accurate at describing real human behaviour.

Let $\epsilon \in [0, 1]$ be the noise parameter. The probability that a subject at time t makes a choice c_t and obtains a reward r_t is described by the following equation:

$$p(r_t, c_t | \mathbf{r}, \mathbf{c}, \epsilon) = p(r_t | c_t) p(c_t | c_{t-1}, r_{t-1}, \epsilon).$$

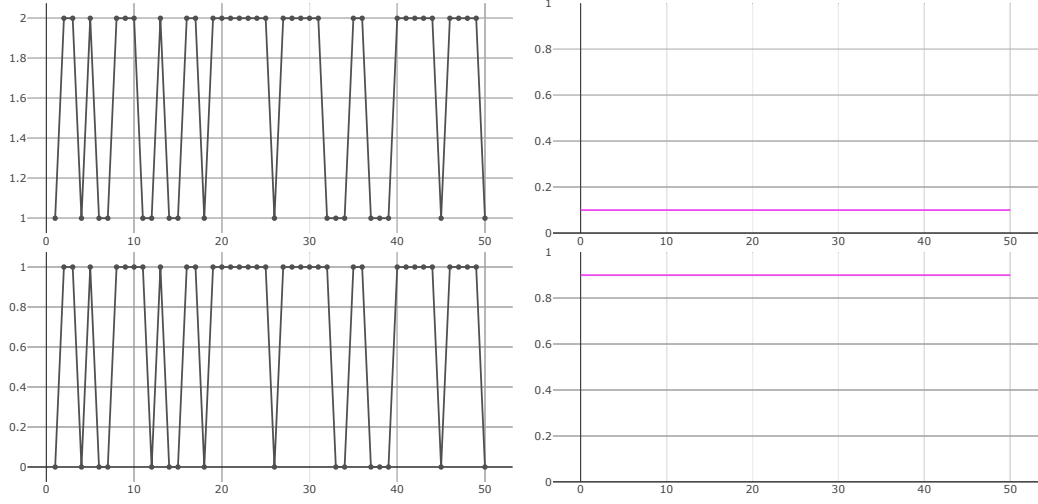


Figure 4.3: Example of a simulated 2ABT experiment for the stochastic win-stay lose-shift model. The noise parameter ϵ is fixed to 0.3. The reward probabilities for the arms are fixed to 0.1 and 0.9 respectively. We plot the choices and the rewards of the simulated subject.

In particular,

$$p(c_t = i | c_{t-1}, r_{t-1}, \epsilon) = \begin{cases} 0.5 & \text{if } t = 1 \\ 1 - \epsilon & \text{if } t > 1, r_{t-1} = 1 \text{ and } c_{t-1} = i \\ \epsilon & \text{if } t > 1, r_{t-1} = 0 \text{ and } c_{t-1} = i \\ 1 - \epsilon & \text{if } t > 1, r_{t-1} = 0 \text{ and } c_{t-1} \neq i \\ \epsilon & \text{if } t > 1, r_{t-1} = 1 \text{ and } c_{t-1} \neq i \end{cases}$$

with $i = 1, 2$.

Consider

$$a_t := (1 - \epsilon) \mathbb{1}_{\{r_{t-1}=1, c_{t-1}=1\}} + \epsilon \mathbb{1}_{\{r_{t-1}=0, c_{t-1}=1\}} + (1 - \epsilon) \mathbb{1}_{\{r_{t-1}=0, c_{t-1}=2\}} + \epsilon \mathbb{1}_{\{r_{t-1}=1, c_{t-1}=2\}}$$

for $t = 2, \dots, T$. Then, the log-likelihood is equal to

$$l(\epsilon) = \text{constant} + \log(0.5) + \sum_{t=2}^T \log \left[a_t \mathbb{1}_{\{c_t=1\}} + (1 - a_t) \mathbb{1}_{\{c_t=2\}} \right].$$

The MLE estimator is denoted by $\hat{\epsilon}$.

4.3 Design optimization framework

In order to investigate the design variables, we utilise a design optimization framework. In particular, we assign a utility value to each experimental design — averaged across the behaviour of many different subjects. Then, we analyse the results to identify optimal designs which improve the data collection process according to the criteria described by the utility.

Similar to other works in this area, we consider a static two-armed bandit task with a fixed number of trials. Thus, the design variable space $D = [0, 1]^2$ consists of the reward probabilities of the two arms. It is worth highlighting that the dimension of the design variable space increases rapidly as the complexity of the MABTs increases. For a generic static MABT, the design variable space is $D = \{[0, 1]^{T \times K}, K \in \mathbb{N}^+, T \in \mathbb{N}^+\}$ where K is the number of arms and T the number of trials. This explains the necessity to select a simple experiment, as for example a two-armed bandit task, to make the problem tractable (Melinscak & Bach, 2020; Valentin et al., 2024; S. Zhang & Lee, 2010).

The local utility and the learning models analysed in this work are symmetric with respect to the MABT arms. Consequently, the points (p_1, p_2) and (p_2, p_1) are mathematically equivalent in the static analyses. Unless otherwise stated, however, the simulations and figures in this thesis use the full design space $D = [0, 1]^2$ without imposing the ordering constraint $p_1 \leq p_2$. Therefore, both $p_1 < p_2$ and $p_1 > p_2$ may appear in reported designs, and symmetric counterparts should be interpreted as equivalent. In the adaptive setting of Chapter 5, no ordering constraint is imposed because designs are updated sequentially and the symmetry need not be preserved over time.

Let us now formally define the DO problem. First, we define a local utility function which assigns a utility value to each design $\mathbf{d} \in D$ assuming that the data — the choices $\mathbf{c}$, the rewards $\mathbf{r}$ and the model parameters $\boldsymbol{\theta}$ — are known. We denote the *local utility* as $u(\mathbf{d}, \mathbf{c}, \mathbf{r}, \boldsymbol{\theta})$. Secondly, we average the local utility across many different outcome to obtain a *global utility* for each design $\mathbf{d} \in D$. Formally this is

done by weighting the local utility function with the joint probabilities $p(\mathbf{c}, \mathbf{r}, \boldsymbol{\theta})$, see (Pitt & Myung, 2019), so that

$$U(\mathbf{d}) = \int u(\mathbf{d}, \mathbf{c}, \mathbf{r}, \boldsymbol{\theta}) p(\mathbf{c}, \mathbf{r}, \alpha, \tau) d\mathbf{c} d\mathbf{r} d\boldsymbol{\theta}. \quad (4.4)$$

In conclusion, we analyse the behaviour of the global utility across the design variable space D to select an optimal design $\hat{\mathbf{d}}$. Thus,

$$\hat{\mathbf{d}} \in \arg \max_{\mathbf{d} \in D} U(\mathbf{d}) \quad (4.5)$$

is the optimization problem we aim to solve.

The results provided by Problem (4.5) depend on the choice of the local utility $u(\mathbf{d}, \mathbf{c}, \mathbf{r}, \boldsymbol{\theta})$ which is defined by the researcher. The local utility represents what the researcher is optimizing for, and is chosen to achieve specific goals. There are many different forms of local utility which have been applied in the DO literature, each with its own purpose. Examples are utility based on the point estimation of the model parameters (Melinscak & Bach, 2020), on information theory or model comparison measures (S. Zhang & Lee, 2010).

The local utility function has to satisfy some conditions. First, $u(\mathbf{d}, \mathbf{c}, \mathbf{r}, \boldsymbol{\theta})$ should be regular enough for Equation (4.4) to be meaningful. Second, the optimization algorithms used to solve Problem (4.5) may require additional conditions. In this work, we utilise an algorithm developed by Müller et al. (2004) which further requires the local utility to be bounded and positive.

To demonstrate how DO methods perform in the context of MABT experiments and, in particular, for TD learning models, we choose a local utility function based on the maximum likelihood estimation of the model parameters. In particular,

$$u(\mathbf{d}, \mathbf{c}, \mathbf{r}, \boldsymbol{\theta}) = 1 - \frac{\|\mathbf{w}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\|^2}{|\boldsymbol{\theta}|}, \quad (4.6)$$

where $|\boldsymbol{\theta}|$ is the number of parameters and $\mathbf{w}$ is a vector of weights. The weights

in Equation (4.6) can be changed to give more importance to one parameter if that aligns with the goals of the researcher. The function $u(\mathbf{d}, \mathbf{c}, \mathbf{r}, \boldsymbol{\theta})$ assigns greater utility to designs which result in smaller squared errors $\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2$. The optimal designs are those which guarantee the smallest squared error for the MLE estimates and we are optimizing for parameter estimation quality. Notice that the function in Equation (4.6) is bounded, positive and differentiable.

The local utility used in this work is only one of the many possible choices and relies on point estimation. As mentioned before, information theory or model comparison based utility can be used instead. The error in Equation (4.6) could be accounted for with absolute values rather than squared errors. In this case, however, the local utility would no longer be differentiable. Additionally, terms not strictly related to the parameter estimation can be added to the utility function in order to track other factors of interest in the experiment such as the frustration of the subjects or the time when the task is administrated.

In order to solve Problem (4.5), we utilise the algorithm developed by Müller et al. (2004), an augmented probability based MCMC algorithm. This method has the advantage of executing the exploration of the design variable space together with the evaluation of the utility thanks to a statistical trick. In particular, an augmented probability — proportional up to a constant to the global utility — is utilised as the target probability of the MCMC chain. This method provides data for a sensitivity analysis without additional computational costs. The pseudo-code of the algorithm can be found in Algorithm 3; see Müller et al. (2004) for a more detailed description.

This algorithm was applied in S. Zhang and Lee (2010) for MABT experiments with good results. The authors analysed the experimental design selection during MABT experiments using the algorithm developed by Müller et al. (2004). They focus on statistical models similar to Model 3 considered in this work. While their method uses Bayesian estimation, we utilised MLE estimation due to the increased implementation complexity of the Bayesian approach for the TD models considered. We added two other models based on TD learning to the analysis. TD learning

Algorithm 3

- 1: Assume that the current state of the chain is $(\mathbf{d}, v)$, where

$$v = \frac{1}{J} \sum_{j=1, \dots, J} \log u(\mathbf{d}, \boldsymbol{\theta}_j, \mathbf{c}_j, \mathbf{r}_j)$$

is the empirical approximation of $\log U(\mathbf{d})$ through sampled data.

- 2: Propose a new design $\tilde{\mathbf{d}} \sim N_{[0,1]}(\tilde{\mathbf{d}}|\mathbf{d})$, where $N_{[0,1]}$ is a truncated normal distribution between 0 and 1.
3: Sample J simulated experiments

$$(\boldsymbol{\theta}_j, \mathbf{c}_j, \mathbf{r}_j) \sim p_{\tilde{\mathbf{d}}}(\boldsymbol{\theta}, \mathbf{c}, \mathbf{r}) = p_{\tilde{\mathbf{d}}}(\mathbf{c}, \mathbf{r}|\boldsymbol{\theta})p(\boldsymbol{\theta}).$$

Then evaluate the estimated logarithm of the global utility $\tilde{v}$.

- 4: Evaluate the acceptance probability

$$\alpha_J = \min[1, \exp(J\tilde{v} - Jv)].$$

With probability α_J accept the new proposal $(\tilde{\mathbf{d}}, \tilde{v})$, otherwise leave $(\mathbf{d}, v)$ unchanged.

- 5: Update J based on the cooling down schedule.
6: Repeat steps 1-5 until convergence.
-

models were also analysed by Valentin et al. (2024) and — in the adaptive case — by Melinscak and Bach (2020). While their research generally yields point estimates for the optimal design, this work focuses heavily on mapping the landscape of these design choices. By detailing the utility and regret of different configurations, we provide a graphical and analytical understanding of exactly how much is gained by employing these optimization methods.

The annealing schedule is defined as $J : \mathbb{N}^+ \rightarrow \mathbb{N}^+$ where

$$J(n) = \lfloor \log(n)/0.5 \rfloor \tag{4.7}$$

and $\lfloor \cdot \rfloor$ is the floor function. Notice that the first iteration is excluded from the cooling down schedule, utilizing a deterministic initialization for the first design variables. The application of a cooling down schedule creates a more peaked target distribution as the number of iterations increases. High values of the target distribution correspond to high utility values. Thus, peaked regions of the target distribution identify the more optimal designs. The cooling down schedule for 100000 simulations according

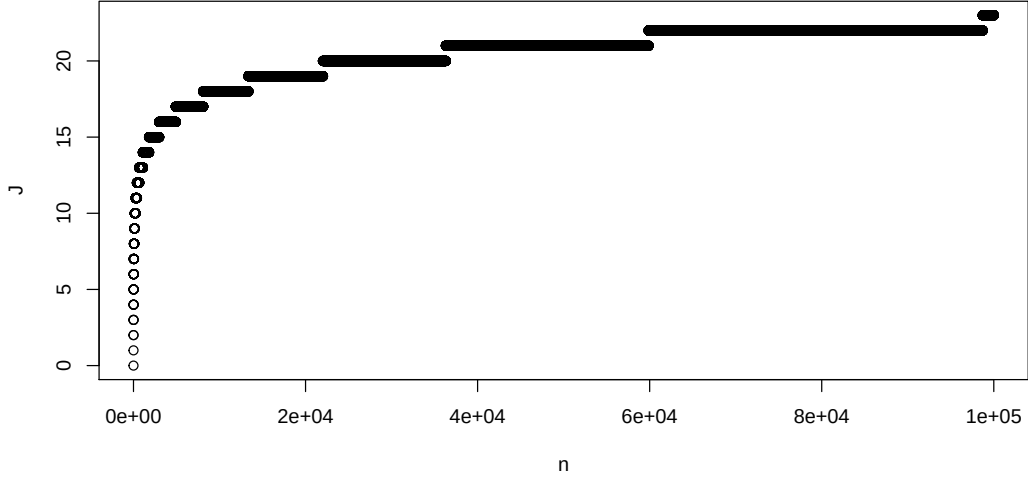


Figure 4.4: Cooling down schedule for 100000 simulations.

to Equation (4.7) is plotted in Figure 4.4.

The simulated experiments in step 2 of Algorithm 3 are sampled based on both the chosen learning model $p_d(\mathbf{c}, \mathbf{r}|\boldsymbol{\theta})$ and the definition of $p(\boldsymbol{\theta})$. In this work, we allow for two different sampling options for the model parameters: we either utilise Beta and Gamma distributions or Uniform distributions. The first case is appropriate when prior knowledge about the expected values of learning rate and inverse temperature parameters is available, allowing for narrower distributions. The second case is preferred when no such prior information exists.

4.4 Results

In this section, we apply the DO method described in Algorithm 3 to the three learning models listed in Section 4.2. Moreover, we investigate the loss or *regret* of not utilizing the optimal designs obtained through Algorithm 3 in all these settings thanks to the information stored in the MCMC chain. The analyses were performed in R. The optimization step required for the MLE estimation, see Equation (4.3), is performed through the `optim` function (Team, 2023). The kernel density estimations were obtained through the `ks` package (Duong, 2007). All plots besides Figure 4.4

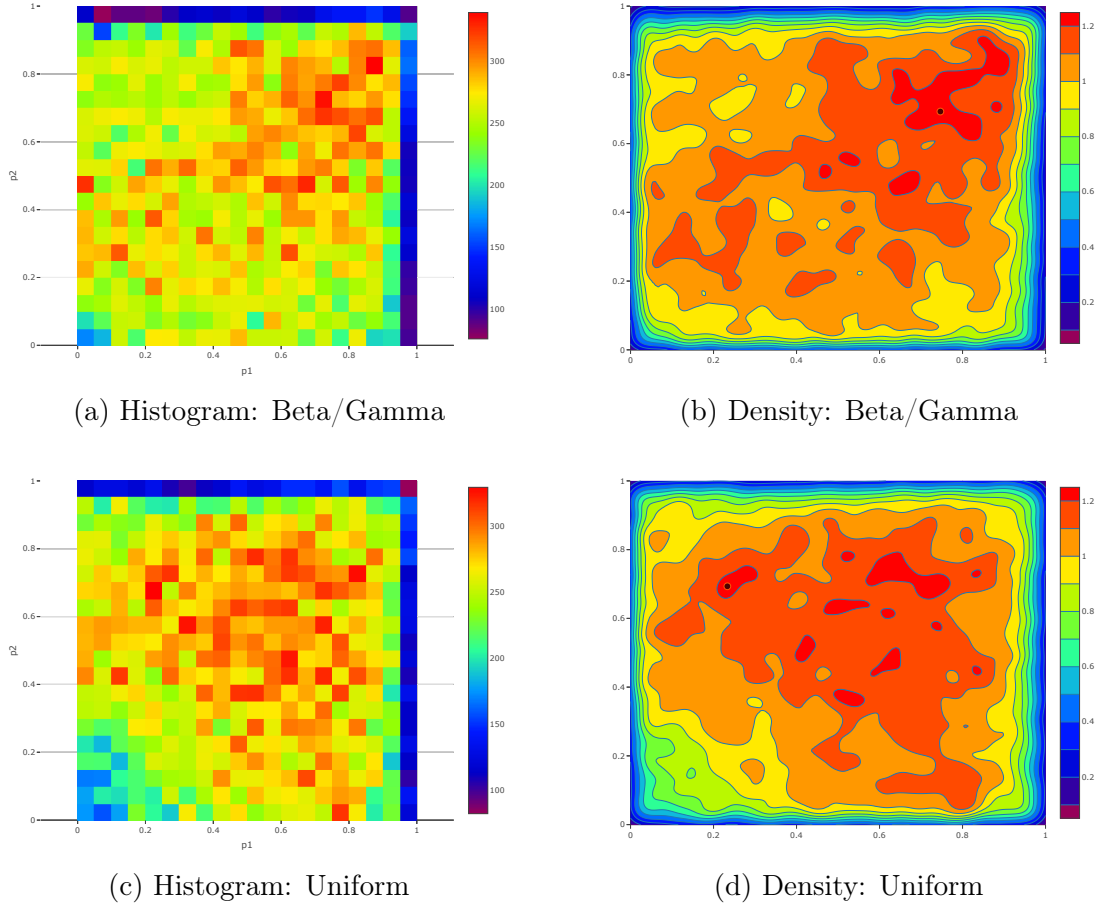


Figure 4.5: Results for Model 1 (two-armed bandit, 50 trials). Figures 4.5a and 4.5c show 2D histograms of simulated design variables for Beta/Gamma and Uniform priors, respectively. Figures 4.5b and 4.5d display the corresponding kernel density estimation. The density $h(p_1, p_2)$ is proportional to the global utility $U(p_1, p_2)$; thus, high-density regions indicate optimal experimental designs. Acceptance rates: 0.86 (Beta/Gamma) and 0.845 (Uniform). The execution time is approximately 8 hours for each case.

were generated through the `plotly` package (Sievert, 2020).

Theoretical convergence is guaranteed under the framework established by Müller et al. (2004), provided the Markov chain is irreducible, the design space is discrete, and the annealing schedule satisfies specific logarithmic decay conditions. Additionally, the utility function is required to be both bounded and strictly positive. To empirically evaluate convergence, we initialized three independent Markov chains with dispersed starting values. Then, we calculated the Gelman-Rubin diagnostic $\hat{R}$ via the `coda` package in R. Although each chain was run for 100 000 iterations (following a 5 000 iterations burn-in), convergence was observed within the first 1 000

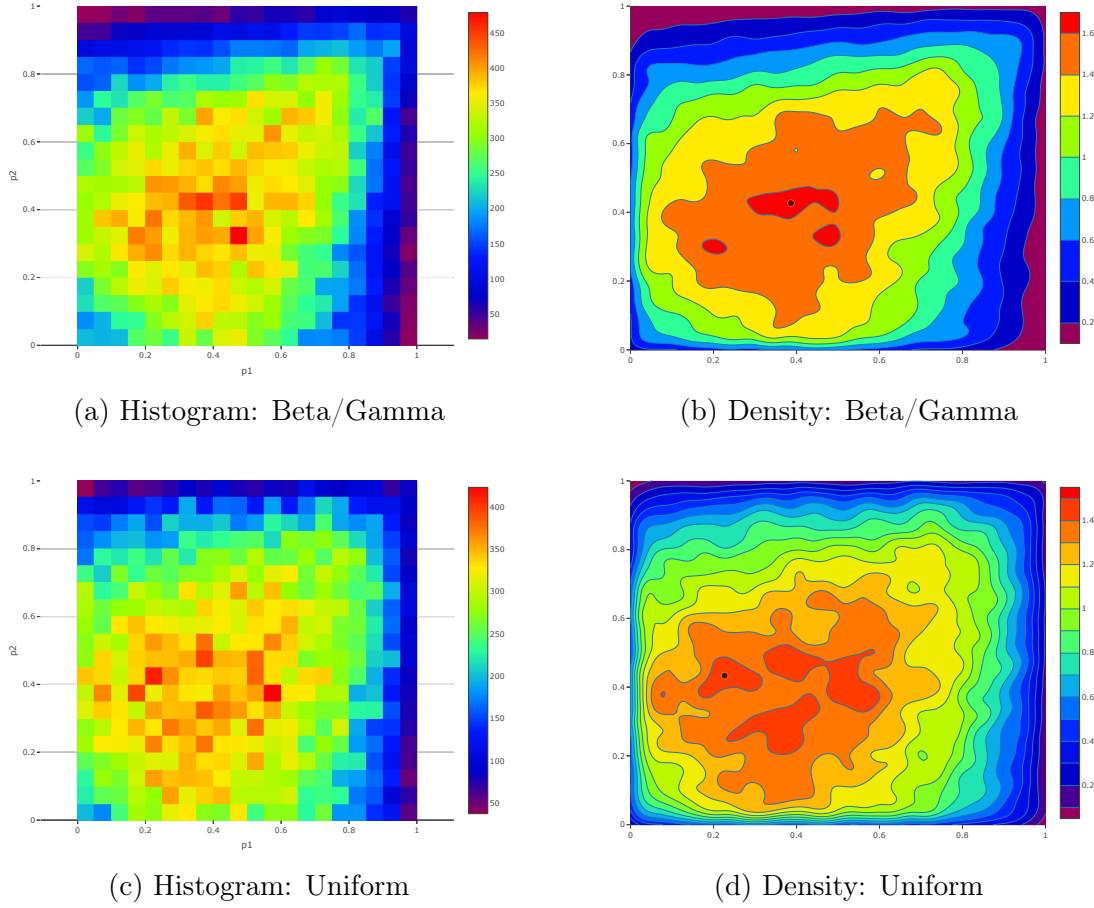


Figure 4.6: Results for Model 2 (two-armed bandit, 50 trials). Figures 4.6a and 4.6c show 2D histograms of simulated design variables for Beta/Gamma and Uniform priors, respectively. Figures 4.6b and 4.6d display the corresponding kernel density estimation. Acceptance rates: 0.78 (Beta/Gamma) and 0.77 (Uniform). The execution time is approximately 22 hours for each case.

iterations, with all $\hat{R}$ values effectively equal to 1. This indicates that the total computational budget provided significant redundancy relative to the time required to reach empirical convergence. The observed acceptance rates were consistently high across all simulations, suggesting a relatively flat utility surface for all models considered. While this confirms the method’s computational efficiency, it implies that the underlying Maximum Likelihood Estimation (MLE) approach may yield a surface that lacks the curvature necessary for highly sensitive design selection. Detailed acceptance rates and execution times for each simulation are documented in the respective figure captions.

First, we applied Algorithm 3 using Model 1 and considering a two-armed bandit task with 50 trials. The model parameters are sampled in two different ways. In

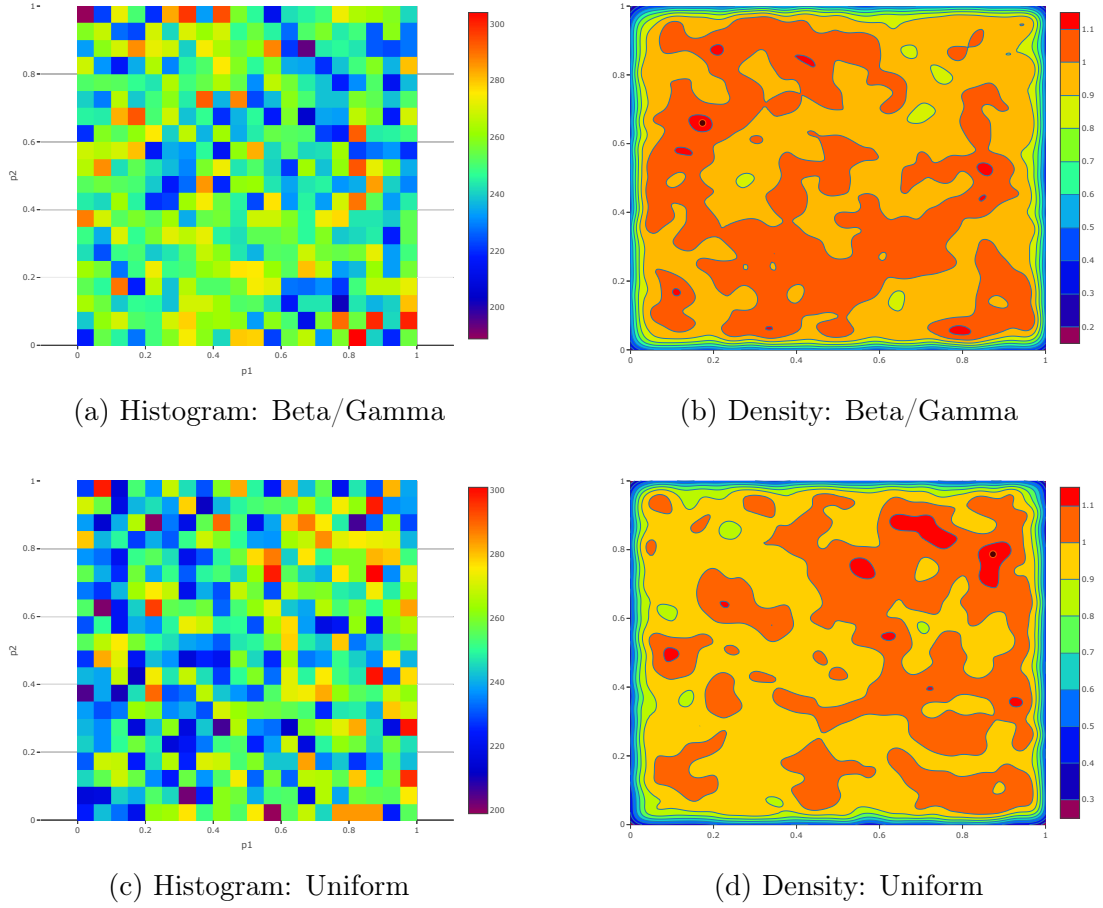


Figure 4.7: Results for Model 3 (two-armed bandit, 50 trials). Figures 4.7a and 4.7c show 2D histograms of simulated design variables for Beta/Gamma and Uniform priors, respectively. Figures 4.7b and 4.7d display the corresponding kernel density estimation. The acceptance rates are 0.89 in both cases. The execution time is slightly less than 25 minutes for each case.

the first case, we utilise a Beta distribution to sample the learning rate and a Gamma distribution to sample the inverse temperature parameter. In particular, $\alpha \sim \text{Beta}(a, b)$ where

$$p(\alpha) = \frac{1}{B(a, b)} \alpha^{a-1} (1 - \alpha)^{b-1},$$

with $a, b > 0$, $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a + b)$ and where Γ is the Gamma function. On the other hand, $\tau \sim \text{Gamma}(v, z)$ where

$$p(\tau) = \frac{z^v}{\Gamma(v)} \tau^{v-1} \exp(-z\tau),$$

where $v, z > 0$ are, respectively, the shape and rate parameters. We assume

$$a = 2, \quad b = 6 \quad \text{and} \quad v = 7.5, \quad z = 0.5. \quad (4.8)$$

The values in Equation (4.8) are chosen so that the expected values of the sampling distributions are 0.25 for the learning rate and 15 for the inverse temperature parameters. These coincide with the theoretical optimal values found for a stationary two-armed bandit task by Gershman (2015). Notice that the probability of sampling an inverse temperature parameter greater than 100 is almost zero for this sampling choice, which decreases the information loss due to the restriction of its domain to $[0, 100]$. In the second case, we utilise uniform distributions so that

$$\alpha \sim \text{Unif}(0, 1) \text{ and } \tau \sim \text{Unif}(0, 100).$$

The local utility takes the form of

$$u(\mathbf{d}, \mathbf{c}, \mathbf{r}, \boldsymbol{\theta}) = 1 - \frac{(\alpha - \hat{\alpha})^2 + \left(\frac{1}{100}(\tau - \hat{\tau})\right)^2}{2}.$$

The results are shown in Figure 4.5. The two-dimensional histograms in Figures 4.5a and 4.5c are an approximation of the augmented density $h(\mathbf{d})$, which is proportional to the global utility function $U(\mathbf{d})$. Hence, the high-density regions in the plot indicate the optimal experimental designs. Similarly, the simulated probabilities are plotted as a contour plot using a kernel density method in Figures 4.5b and 4.5c.

The plots show that there is an advantage in selecting a design through DO methods. Indeed, the results suggest that very high reward probabilities are less efficient as experimental designs. This feature is more evident for the Beta/Gamma case. This implies that, in order to obtain good parameter estimations, we want to avoid trivial experiments where one of the arms or both arms are easily identified as the winning choice. Similarly, low probabilities should also be avoided since, in this case, arms can easily be identified as unprofitable. This is especially true when both

arms have very low probabilities since the subjects are faced with an unwinnable game with no learning opportunity. Intuitively, trivial designs require no learning over time and the MABT experiments are easily solved by the subjects. Subsequently, less information on the learning process — Q — is available to fit Model 1 compared to more challenging designs.

We repeated the analysis for Model 2, see Figure 4.9b. The model parameters are sampled similarly to the previous case from either Beta/Gamma distributions

$$\alpha_w \sim \text{Beta}(1.5, 5), \alpha_l \sim \text{Beta}(3, 7) \text{ and } \tau \sim \text{Gamma}(7.5, 0.5),$$

or Uniform distributions

$$\alpha_w \sim \text{Unif}(0, 1), \alpha_l \sim \text{Unif}(0, 1) \text{ and } \tau \sim \text{Unif}(0, 100).$$

The local utility is

$$u(\mathbf{d}, \mathbf{c}, \mathbf{r}, \boldsymbol{\theta}) = 1 - \frac{(\alpha_w - \hat{\alpha}_w)^2 + (\alpha_l - \hat{\alpha}_l)^2 + \left(\frac{1}{100}(\tau - \hat{\tau})\right)^2}{3}.$$

We obtain results — shown in Figure 4.6 — very similar to Model 1 thanks to the close similarity between Models 1 and 2. Very high and very low reward probabilities are less efficient. The difference between designs on the border and those closer to the centre are more marked when using Beta/Gamma distributions to sample the model parameters. Thus, the results suggest avoiding trivial experiments and submit challenging tasks to the subjects. Designs on the border of the design variable space D do not allow the subjects to learn over time due to the simplicity or impossibility of the task. The DO method suggests to select experimental designs which instead allow the subjects to learn over time according to the equations provided in both Models 1 and 2.

Finally, we show in Figure 4.7 the results relative to Model 3. The model parameter ϵ is sampled either from a Beta distribution — $\epsilon \sim \text{Beta}(2, 6)$ — or from

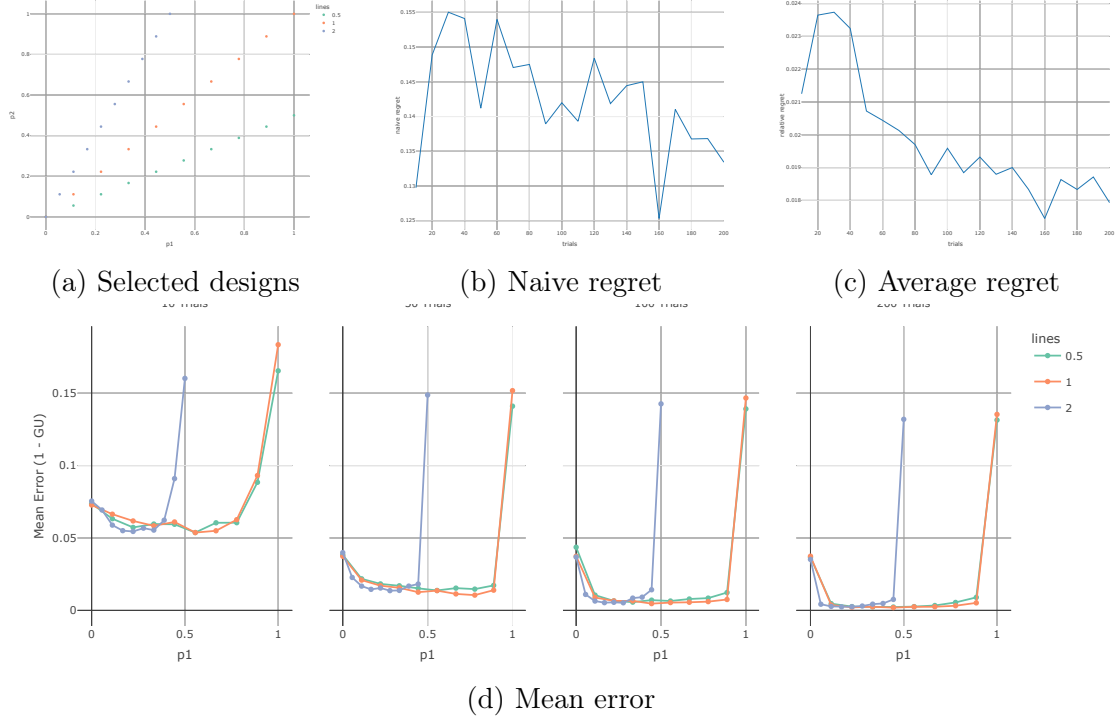


Figure 4.8: Regret analysis for Model 1 with Beta/Gamma sampling. Figure 4.8a shows the set of designs selected to perform additional regret analysis. We chose three lines passing through the origin and with positive coefficients. According to Figure 4.5a, along these directions we observe the most considerable changes in parameter estimation performance. We select 10 points across the three segments $D \cap \{y = \rho x\}$ with $\rho = 2, 1, 0.5$. Figure 4.8d shows the behaviour of the parameter estimation error for the designs in $\bar{D}$ for, in order, $N = 10, 50, 100, 200$ number of trials in the MABT experiment. For each design, we simulated 1000 experiments according to Equation (4.8) and learning Model 1. Hence, we compute the MLE estimator of the parameters, then calculate the error according to Equation (4.9). Figures 4.8b and 4.8c show the regrets defined, respectively, in Equations (4.10) and (4.11). The regrets are computed for different numbers of trials $\{10, 20, 30, \dots, 200\}$. Both regrets decrease, as expected, as the number of trials increases. The average regret decreases faster because the utility flattens as the number of trials increases.

a Uniform distribution — $\epsilon \sim \text{Unif}(0, 1)$. The local utility is

$$u(\mathbf{d}, \mathbf{c}, \mathbf{r}, \boldsymbol{\theta}) = 1 - (\epsilon - \hat{\epsilon})^2.$$

For this model, the results show a flat utility surface with insufficient curvature to inform the selection of an optimal design. This suggests that parameter recovery for Model 3 remains effective independently of the reward probability configuration chosen by the experimenter. Consequently, the MLE for this model is robust to

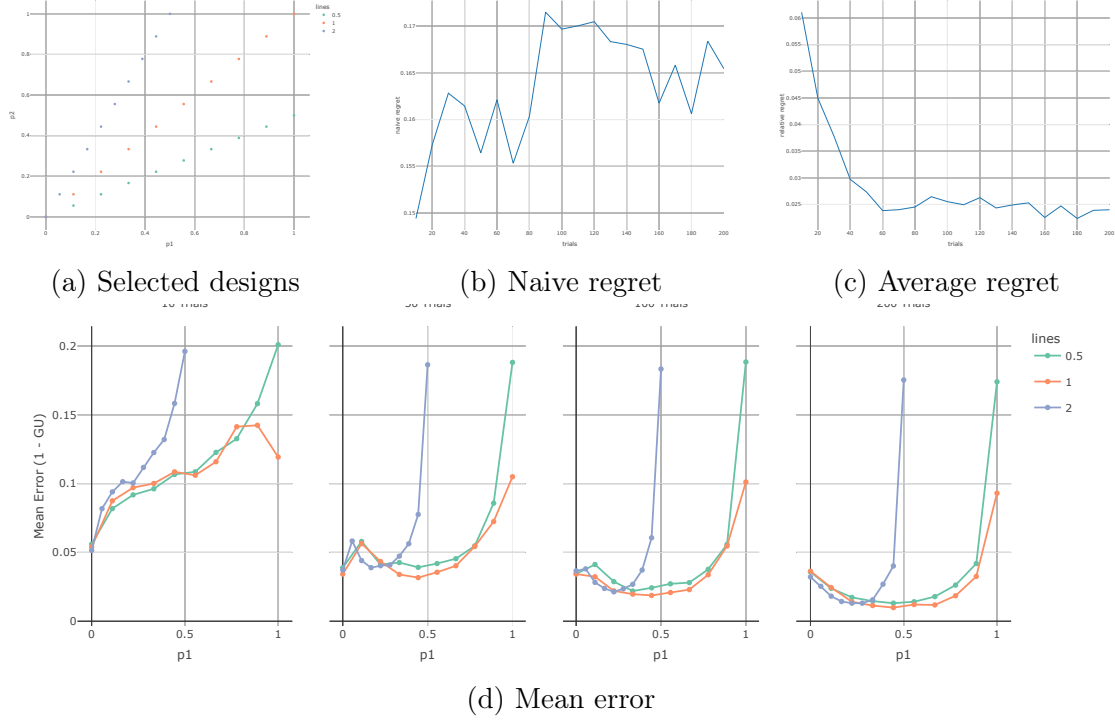


Figure 4.9: Regret analysis for Model 2 with Beta/Gamma sampling. The set of designs of interest $\bar{D}$ shown in Figure 4.9a is the same used for Model 1. Figure 4.9d shows the behaviour of the parameter estimation error for the designs in $\bar{D}$ for, in order, $N = 10, 50, 100, 200$ number of trials in the MABT experiment. For each design we simulated 1000 experiments. Figures 4.9b and 4.9c show the regrets defined, respectively, in Equations (4.10) and (4.11). The regrets are computed for different numbers of trials $\{10, 20, 30, \dots, 200\}$.

environmental shifts. This outcome is expected, as the strategy relies primarily on information from the previous trial. In contrast, Models 1 and 2 are likely more sensitive to reward probabilities shifts as they cumulatively integrate the entire history of outcomes in their decision strategy.

Regret The plots shown before give a qualitative perspective on the advantages of using DO methods for this problem. In order to provide a quantitative perspective, we evaluated the error for the parameter estimation in a finite subset of experimental designs $\bar{D} \subseteq D$. These were chosen based on the results of Algorithm 3 to highlight the behaviour of the augmented probability and, thus, the utility. Then, using these error evaluations, we define a measure of *regret* to better evaluate the performance of DO methods in the experiments. This measure quantifies the loss — in parameter estimation quality — of utilizing suboptimal designs rather than the optimal designs

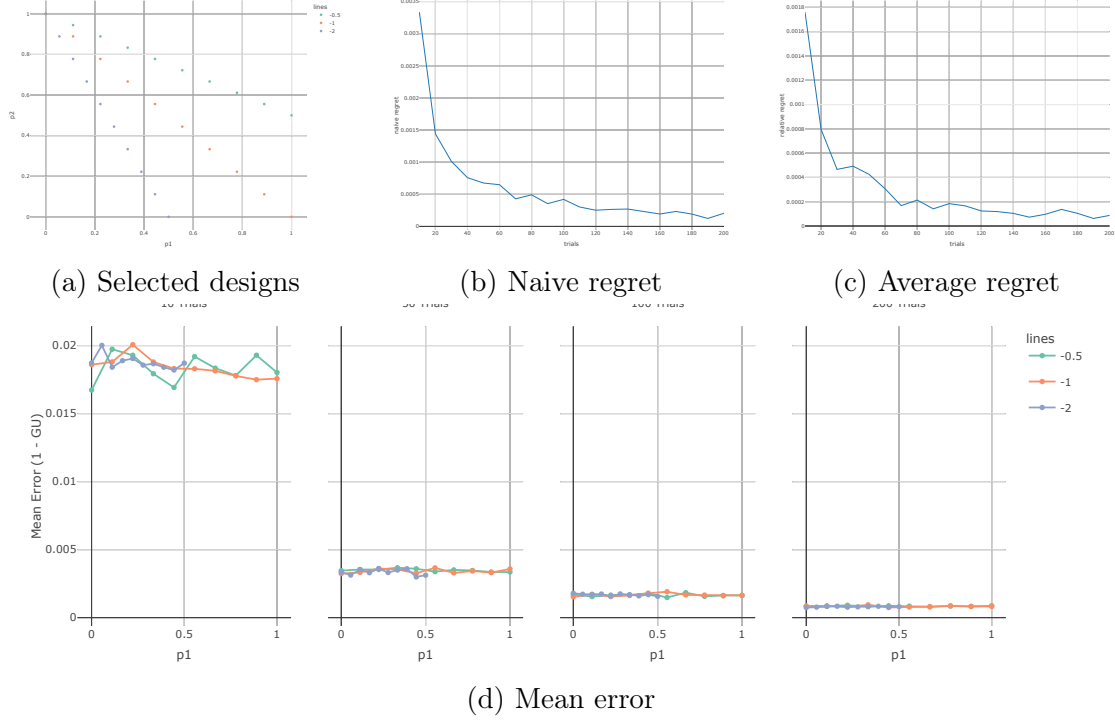


Figure 4.10: Regret analysis for Model 3 with Beta/Gamma sampling. The set of designs of interest $\bar{D}$ is shown in Figure 4.10a. The points are selected on the segments $D \cap \{y = 1 - \rho x\}$ with $\rho = -2, -1, -0.5$. Figure 4.10d shows the behaviour of the parameter estimation error for the designs in $\bar{D}$ for, in order, $N = 10, 50, 100, 200$ number of trials in the MABT experiment. For each design we simulated 1000 experiments. Figures 4.10b and 4.10c show the regrets defined, respectively, in Equations (4.10) and (4.11). The regrets are computed for different numbers of trials $\{10, 20, 30, \dots, 200\}$.

identified through the DO method. It allows both a comparison of good designs with close to optimal ones and very underperforming ones. This analysis allows us to better highlight the advantages brought by DO methods in the context of the MABT experiments considered here. Moreover, these results could help practitioners weigh the additional efforts and computation costs necessary to apply these methods against the resulting advantages. We first focus on Model 1, then repeat the analysis for the other two models.

Utilizing the same methods applied for step 2 in Algorithm 3, we simulate random experiments and compute the error

$$\text{Err}(\mathbf{d}) = \frac{\|\mathbf{w}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\|^2}{|\boldsymbol{\theta}|} = 1 - u(\mathbf{d}, \mathbf{c}, \mathbf{r}, \boldsymbol{\theta}). \quad (4.9)$$

For each design $\mathbf{d} \in \overline{D}$, we plot the mean of the errors $\text{Err}(\mathbf{d})$ for numbers of trials $N = 10, 50, 100, 200$, see Figure 4.8d. As expected, this plot confirms the behaviour seen in Figure 4.5a. The error increases as one or both the reward probabilities increase. High reward probabilities should be avoided as experimental designs for the problem considered. Small reward probabilities are also shown to be less effective design choices. Comparing the error across different numbers of trials, we notice that the error decreases as the number of trials increases. This implies that the average difference between any design and the best one decreases as the number of trials increases. Picking randomly, a high number of trials guarantees we are less likely to select a non-optimal design. Nevertheless, very high reward probabilities remain undesirable experimental designs even for a high number of trials. Thus, regardless of the number of trials, there are experimental designs which perform poorly and should be avoided.

This behaviour can be summarized by looking at a *naive regret* measure defined as

$$R_N = \max_{\mathbf{d} \in \overline{D}} \text{Err}(\mathbf{d}) - \min_{\mathbf{d} \in \overline{D}} \text{Err}(\mathbf{d}), \quad (4.10)$$

where N is the number of trials. High naive regret values imply that selecting the least optimal design can have a substantial negative impact on the parameter estimation analysis contrary to lower values which have less impact. Alternatively, it is possible to compute the *average regret* defined as

$$R'_N = \frac{1}{|\overline{D}|} \sum_{\mathbf{d} \in \overline{D}} \left(\text{Err}(\mathbf{d}) - \min_{\mathbf{d} \in \overline{D}} \text{Err}(\mathbf{d}) \right). \quad (4.11)$$

While the naive regret R_N approximates the difference between the worst and the best designs, the average regret R'_N approximates the average difference between the best design and any other option. The average regret better shows how the utility flattens as the number of trials increases. In particular, this definition highlights the regret of selecting an experimental design randomly compared to a better informed decision through the study of the design variable space D . Notice that the regrets

R_N and R'_N are defined and computed on the subset $\overline{D} \subseteq D$ for ease of computation. They are approximations of the real regrets of DO methods, which should be instead computed on the whole design variable space D .

The results concerning the naive and relative regret for Model 1 with Beta/Gamma sampling are shown in Figure 4.8. In this case, the naive regret R_N tends to increase for low trial counts and, then, it shows a decreasing trend. This implies that the difference between an optimal design and an underperforming one — selected within the set $\overline{D}$ — initially increases. Then, as the number of trials increases, this difference decreases. This suggests that a higher number of trials counterweights a poor design choice. This point is further confirmed by the average regret R'_N , which shows a consistent decreasing trend as the number of trials increases. Thus, a random selection of experimental designs becomes less likely to negatively impact the parameter estimation quality when the number of trials increases.

We repeated the same analysis for Model 2 with Beta/Gamma sampling; the results are shown in Figure 4.9. As for Model 1, the average regret of Model 2 shows a consistent decreasing trend as the number of trials increases, suggesting that, in general, the number of trials compensates for poor design choices. Moreover, this trend seems steeper compared to that observed in Model 1. Nevertheless, unlike the first model, in Model 2 the naive regret shows an increasing trend. Thus, in this case the selection of a bad design becomes more impactful as the number of trials increases.

Finally, an analysis of regret was performed for Model 3; the results are shown in Figure 4.10. In this case, we utilised a different set of designs $\overline{D}$. The naive regret R_N and the average regret R'_N behave very similarly and show a sharp decreasing trend as the number of trials increases.

In conclusion, the average regret R'_N shows a decreasing trend as the number of trials increases for all models considered in this work. This implies that, as expected, more data weakens the negative effects of suboptimal designs choices. Nevertheless, the possibility of selecting a very underperforming design remains an issue for Model

1 and 2. Indeed, the naive regret R_N either increases or decreases slowly in these cases. This implies that the difference between a poor design and the optimal one remains relevant for these models.

Another perspective on the results from the regret analysis can be seen in Figures 4.8d, 4.9d and 4.10d. In these plots, we show the error computed across a cohort of 1000 simulated subjects for different total numbers of trials considered (10, 50, 100, and 200). For Models 1 and 2, we notice that there are multiple near-optimal designs (designs which perform similarly to the optimal one). In particular, in these cases, the error plots show flat regions, especially for higher numbers of trials. This suggests that the optimal designs might not be unique for those cases. Nevertheless, the use of DO methods is still relevant for shorter experiments where suboptimal choices are not counterweighted by a higher amount of data. Moreover, DO methods are useful to avoid designs which perform poorly. Indeed, in both cases, the highest error — corresponding to the worst design — is double or more than the lowest error — corresponding to the optimal design. Finally, the results for Model 3 confirm the flat utility surface observed in the initial DO application: design choice appears to have negligible impact on parameter recovery across all trial counts.

Thanks to the regret analysis, information on a minimum number of trials for the data collection of each model is also available. Indeed, we notice that the error decreases sharply between 10 and 50 trials, then it stabilizes, see Figures 4.8d, 4.9d and 4.10d. This suggests that a higher number of trials results in decreasing benefits. Assuming a good design has been selected, a total number of trials between 50 and 100 is adequate — in simulated settings — for Model 1. For Model 2, a higher number of trials is needed as the error decreases more slowly. Model 3 reaches good parameter recovery already around 50 trials. A meticulous approach to this problem would involve simulating experiments corresponding to the selected optimal design and analysing how the parameter estimation error behaves as the number of trials increases. Notice that all these observations are based on simulated data. The number of trials adequate for an experiment is also impacted by the number and

type of models that a researcher includes in their study. In particular, if models have different optimal designs, a higher number of trials in a compromise design is a possible solution. On the other hand, if those models have similar optimal designs, a smaller number of trials might be sufficient. Nevertheless, DO methods that target model discrimination have been developed and might be a better fit for such problems.

Consequences for the subjects Experiments designed through DO methods can be perceived as frustrating by study participants. For example, in the context of this work, the results obtained for Model 1 and Model 2 suggest the design of non-trivial tasks for the subjects. Indeed, we find that for these two models neither of the arms should be easy to identify as profitable — high reward probability — or unprofitable — very low reward probability. Notice, however, that generating an excessively uncertain environment for the experiments can cause excessive frustration for study participants. So far, we have assumed perfect compliance from our simulated subjects, but in real applications participants may drop out because of this issue. This consequence has been recognized previously in the application of DO methods. DO methods are often found effective, but the possibility that their application may reduce adherence and increase dropout could potentially offset their benefits. Therefore, there is a trade-off between collecting optimal data and ensuring study participants are not overburdened by the experiment.

In order to consider this issue, the local utility can be redefined to include a burden measure for the experiment. A possible example is

$$u(\mathbf{d}, \mathbf{c}, \mathbf{r}, \boldsymbol{\theta}) = 1 - \frac{\|\mathbf{w}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\|^2}{|\boldsymbol{\theta}|} + L(\mathbf{d}), \quad (4.12)$$

where $L : D \rightarrow [0, 1]$ represents the burden placed on the subjects by the experiment. Then, small values of L correspond to stimulating experiments, while high values of L describe frustrating ones. For example, $L(\mathbf{d})$ could track the number of wins/losses and increase in value as the subject is facing repetitive losses. Alternatively, $L(\mathbf{d})$

Assumptions & practical deployment checklist

1. Assumptions

- **Model suitability:** the selected cognitive model must accurately reflect subject behaviour. The models analysed in this work assume that the parameters are fixed in time.

Diagnostic: *a posteriori* model fit on empirical data; comparison of the subject's estimated parameters between the first half versus second half of the experiment.

- **Utility alignment:** the utility function must map precisely to the experimental goal. This work only considers parameter recovery to measure the utility of a design.

Diagnostic: verify that optimizing the chosen utility directly translates to the desired real-world outcome (e.g. improved MLE parameter recovery).

- **Reward distribution matching:** simulated experimental conditions must match the empirical protocol. This work only considers bandit tasks with $K = 2$ arms and Bernoulli distributed rewards.

Diagnostic: ensure code parameters match the actual reward structures presented to subjects.

2. Practical deployment checklist

- **Known *a priori* (by practitioner):** the model, the trials' length T , the utility function $u(\mathbf{d})$, and the priors' parameters.
- **Prior selection:** informed by pilot data and/or uniform priors.
- **Real-world impact:** the framework acts as an "ideal scenario" validator.

could be inferred from real data based on subjects' adherence (Monacelli, Zhang, Schlee, Langguth, et al., 2023). While DO methods aim to improve data collection, ultimately the quality of those data depends on the adherence and focus of the subjects. Therefore, studying the impact these methods have on real data and including it in the DO method itself using tools such as Equation (4.12) is an interesting direction for future work.

Practical considerations As this chapter focuses purely on simulation work, its primary real-world utility serves as a pre-experimental validation tool. Essentially, it

allows researchers to simulate their planned experiment under ideal conditions to verify if their chosen design variables (e.g. reward probabilities) make sense before allocating resources to human trials. Prior to utilizing this method, the cognitive model, the experimental structure, and the utility function must be explicitly defined. Once these are set, the algorithm can simulate the experiment to suggest optimal design choices for the multi-armed bandit task. However, to assess the plausibility of these suggestions in real data, practitioners must perform several checks. First, *a posteriori* model fitting is required to ensure the chosen cognitive model adequately describes the actual subjects' behaviour. Second, the selected utility function must perfectly align with the specific research objective — which, in this chapter, is maximizing parameter recovery via Maximum Likelihood Estimation (MLE). Finally, researchers must carefully select appropriate parameter priors based on existing literature or pilot data. In the absence of such knowledge, uniform distributions may be employed, but comparing results under both prior specifications is highly recommended.

While the environment is highly flexible, we consider only a limited number of cases. This introduces some additional limitations. First, we only consider bandit tasks with $K = 2$ arms and Bernoulli distributed rewards. Some empirical applications might utilise continuous rewards (e.g. Beta-distributed) or more than two arms. These cases would necessitate appropriate modifications to the underlying code. Second, for all the models considered, the parameters are assumed fixed in time. This might be incompatible with real settings where subject performance might drop as the experiment goes on due to decreasing engagement. For this purpose, ensuring the model accurately describes the data we plan to collect is extremely important.

4.5 Discussion and outlook

We utilised the algorithm developed in Müller et al. (2004) to identify optimal regions of experimental designs in order to improve the point estimation of model parameters. We considered a static two-armed bandit task and three different learning models:

a TD model similar to the Rescorla-Wagner model (Model 1), a TD model which accounts for wins and losses separately (Model 2) and the stochastic win-stay lose-shift model (Model 3).

The results suggest that challenging tasks are preferable for TD learning models as they allow the subjects to learn over time. The learning process is, indeed, the main focus of these models thanks to the iterative estimation of the reward probabilities they rely on. In particular, high or low reward probabilities should be avoided as, in these cases, the arms are easily identified as profitable or not, impeding the learning experience. On the other hand, our results suggest that any reward probabilities configuration is good for the stochastic win-stay lose-shift model. Indeed, in this case, the model parameter can be recovered effectively independently of the experimental design thanks to a flat utility surface.

To further quantify the possible impact of DO methods in this context, we performed a regret analysis on a subset of experimental designs selected based on the results previously obtained. We find that, for Models 1 and 2, the average regret decreases as the number of trials increases. Hence, the selection of a suboptimal design is less impactful if more data is collected. Nevertheless, the naive regret decreases more slowly or increases, suggesting that DO methods remain relevant regardless of the number of trials to identify and avoid underperforming designs.

Our results are for the very specific study case of data obtained through a static two-armed bandit task and described by three specific learning models. The field of computational psychology is currently developing so that a more general framework would greatly benefit the applicability of these techniques. Based on the specific research question of the researcher, many different versions of the multi-armed bandit tasks are being used and new statistical models specifically designed to describe data collected through these experiments are being studied. This implies that both the static two-armed bandit task data and the models considered here could not be the intended tools of the practitioners who are likely to use these results.

Thanks to the versatility of the design optimization framework, this work can

be adapted to different variations of MABT experiments and models based on the needs of the researcher. The extension of the tools developed here is a possible area of future work. The code developed in this work can be extended to include different models and/or more design variables — such as the number of trials or an additional reward probability relative to a third arm. In particular, regarding the models, streamlining the implementation could make these tools more accessible to the interested practitioners. For example, it is possible to utilise widely used tools in the computational psychology field such as the HBayes package for the definition and evaluation of these models (Ahn et al., 2017). A second option is to utilise state-space models to describe any model needed by the researcher through a well-established and thoroughly studied mathematical framework (Yousefi et al., 2019). Nevertheless, to the best of our knowledge, the translation of models currently used in the computational field into state-space models is not always straightforward and more research in this area could be needed.

Furthermore, as already proposed by Melinscak and Bach (2020), it is of interest to investigate an adaptive generalization of the algorithm. Bayesian methods are a possible solution. Sensitivity to priors is a disadvantage in a traditional design optimization setting, but it can be an advantage in an adaptive design optimization setting. Exploiting Bayesian priors allows us to extend Algorithm 3 to its adaptive version (Müller et al., 2004). Another approach was recently suggested by Blau et al. (2022). The authors focus on searching for a policy to explore the design variable space rather than an optimal design, obtaining a non-myopic algorithm. In particular, they re-frame the DO problem as a Markov Decision Process and, within their framework, they avoid the computation of the posterior distribution. As highly computationally expensive posteriors are common in computational psychology due to the complexity of MABT experiments and their models, this new approach might solve a common challenge of DO applications in this area.

The results of this work are theoretical: we focus on the simulation and analysis of possible experimental design questions for the interested practitioners in the context

of MABT experiments. A relevant area of future research is the application of these methods in real-world experiments. While outside the scope of this work, real-world studies can provide valuable information about the effectiveness of the approaches simulated here in real settings and evaluate the potential frustration caused by the sometimes challenging designs selected through the DO method.

Chapter 5

Adaptive data collection through an inhomogeneous MCMC algorithm

In Chapter 4, we investigated the effects of DO methods through the application of inhomogeneous Markov chains to decision-making experiments. In this chapter, we utilise the tools developed in Chapter 4 to obtain an adaptive extension. This allows us to tailor the data collection process to the individual subject, as we iteratively update the experimental design using a design optimization framework and the information accumulated about each subject. To achieve this goal, we divide a two-armed bandit task into batches. After each batch, the experimental design — the reward probabilities p_1 and p_2 — is updated based on the knowledge accumulated about the subject’s behaviour up to that point using Algorithm 3. This results in a non-static version of the two-armed bandit task.

A similar approach has been used by Melinscak and Bach (2020). However, while we use the algorithm developed by Müller et al. (2004), they use a Gaussian-based optimization method to select the next optimal design. Although we expect the latter option to have a computational advantage, our objective was to use the information stored in the MCMC simulations to further analyse the impact of (adaptive) DO methods for MABT experiments. In particular, we use the results provided by the MCMC simulations to evaluate the regret in our experimental choices.

The code and the simulated data used in this chapter are available in the GitHub repository (<https://github.com/gretamonacelli/docmab/releases/tag/v1.0.0>).

5.1 The experiment

As in Chapter 4, we consider a two-armed bandit task, but focus only on Model 1 described in Section 4.2. During the task, each subject is asked to choose between two arms for a fixed number of trials $T \in \mathbb{N}^+$. Each arm gives a reward according to a probability distribution. In particular, the two arms are modelled as two Bernoulli random variables $S_1 \sim \text{Ber}(p_1)$ and $S_2 \sim \text{Ber}(p_2)$ where $p_1, p_2 \in [0, 1]$. However, unlike the previous chapter, the reward probabilities are no longer fixed. Instead, they can be changed by the researcher at fixed intervals. Therefore, in this chapter we are no longer using static MABT as in Chapter 4, but rather a non-static, piecewise MABT as in Melinscak and Bach (2020). In this case, we have a sequence of Bernoulli random variables $S_{1,i} \sim \text{Ber}(p_{1,i})$ and $S_{2,i} \sim \text{Ber}(p_{2,i})$ where $p_{1,i}, p_{2,i} \in [0, 1]$ and $i = 1, \dots, k$.

The division of the two-armed bandit task into intervals is formally done as follows. Let $k, b \in \mathbb{N}^+$ such that the total number of trials is equal to $T = kb$. Then b represents the *batch* or interval of the experiment. The first change of the reward probabilities through the application of Algorithm 3 is performed after data from the first b trials has been collected. Then, another $k - 1$ batches of length b follow, each followed by an experimental design update. In total, there are k batches of length b . It is possible to consider a more general division of the experiment, but that is outside the scope of this work. In any case, it is important to ensure that at least the first batch has sufficient length to allow a smooth estimation of the parameters α and τ . A graphical representation of this setup is given in Figure 5.1. Examples of data simulated in this experimental setting are given in Figures 5.2 and 5.3.

After each batch, we utilise Algorithm 3 to select a new optimal design. The update is then based on the results of a design optimization method applied on the data available up to that point in the experiment. The new optimal design is selected

by solving

$$\hat{\mathbf{d}}_j \in \arg \max_{\mathbf{d} \in D} U(\mathbf{d}), \quad (5.1)$$

where

$$U(\mathbf{d}) = \int u(\mathbf{d}, \mathbf{c}, \mathbf{r}, \alpha, \tau) p(\mathbf{c}, \mathbf{r}, \alpha, \tau) d\mathbf{c} d\mathbf{r} d\alpha d\tau. \quad (5.2)$$

The local utility is defined to improve the MLE estimates of the learning rate α and the inverse temperature parameter τ of the Rescorla-Wagner model:

$$u(\mathbf{d}, \alpha, \tau) = 1 - \frac{(\alpha - \hat{\alpha})^2 + \left(\frac{1}{100}(\tau - \hat{\tau})\right)^2}{2}. \quad (5.3)$$

Thanks to this, the reward probabilities gradually adapt to the behaviour of the subjects using their prior history in order to improve parameter estimation.

In conclusion, for each subject, we obtain and store the following data:

- $\alpha_{\text{true}} \in [0, 1]$, $\tau_{\text{true}} \in [0, 100]$ the ground truth of the Rescorla-Wagner parameters, that is, the learning rate and the inverse temperature parameters;
- $\hat{\alpha}_a \in [0, 1]^k$, $\hat{\tau}_a \in [0, 100]^k$ the sequence of the MLE estimates of the Rescorla-Wagner parameters during the application of the adaptive DO algorithm;
- $(\mathbf{d}_1, \dots, \mathbf{d}_k) \in [0, 1]^{2 \times k}$ the sequence of the experimental designs updated throughout the adaptive DO case.

To evaluate these methods, we simulated a cohort of subjects participating in this experiment. We utilise the Rescorla-Wagner model to generate data. The values of the learning rate and the inverse temperature parameters are assumed fixed. They are sampled, respectively, from a $\text{Beta}(a, b)$ and a $\text{Gamma}(v, z)$. They represent the ground truth of the experiment and are assumed — sometimes naively — known throughout this chapter.

To evaluate the performance of our adaptive design optimization method (as described above), in which reward probabilities are updated after each batch using inhomogeneous MCMC informed by participant behaviour, we simulated the same

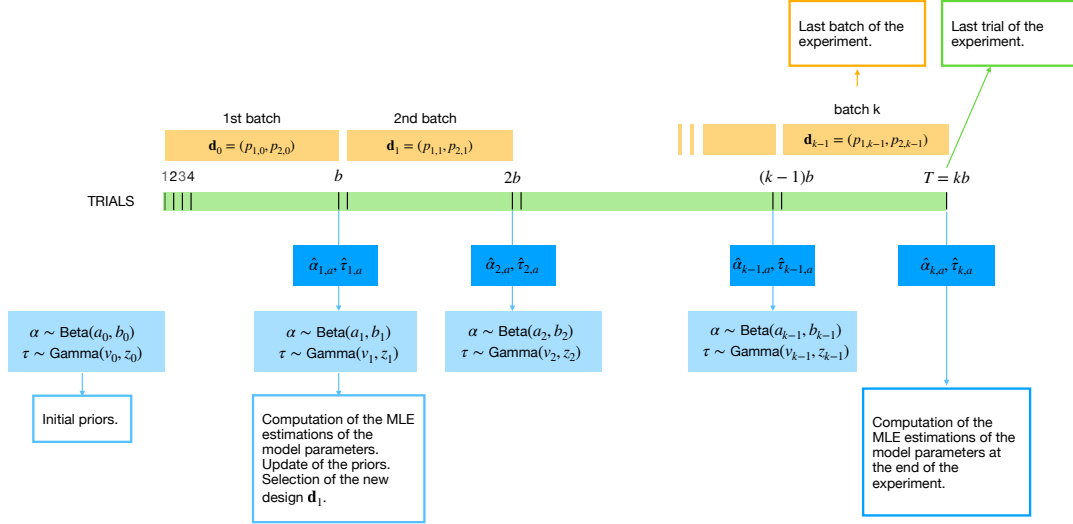


Figure 5.1: The figure shows a graphical representation of the experiment presented to a subject. The timeline of the experiment is represented in green and consists of $T = kb$ trials. These are divided into k batches for data collection, each of length b . The goal of the experiment is to use — after each batch — the algorithm developed by Müller et al. (2004) (see also Algorithm 3 in Chapter 4) in order to inform the data collection for the following batch. The first experimental design $\mathbf{d}_0$ is chosen thanks to the results in Chapter 4. Subsequently, after each batch, the experimental designs $\mathbf{d}_i$, for $i = 2, \dots, k - 1$ — that is, the reward probabilities of the two-armed bandit task — are updated using the data collected up to that point. In order to save computational power, the update rule and Algorithm 3 are not used on the last batch of data, for which we only compute the estimates $\hat{\alpha}_k, \hat{\tau}_k$ to evaluate the performance of the algorithm. This process is summarised in Algorithm 4. As the local utility is defined to improve the MLE estimation of the learning rate and the inverse temperature parameters, the new experimental designs are selected towards that goal. Algorithm 3 is informed of the prior history of the subjects in two ways. First, the choices and rewards experienced by the subject up to the current batch are passed to the algorithm. Second, the priors of Algorithm 3 are updated through the online estimations of the parameters $\hat{\alpha}_i, \hat{\tau}_i$ with $i = 1, \dots, k - 1$. This allows us to tailor the experimental designs to the individual subject's behaviour.

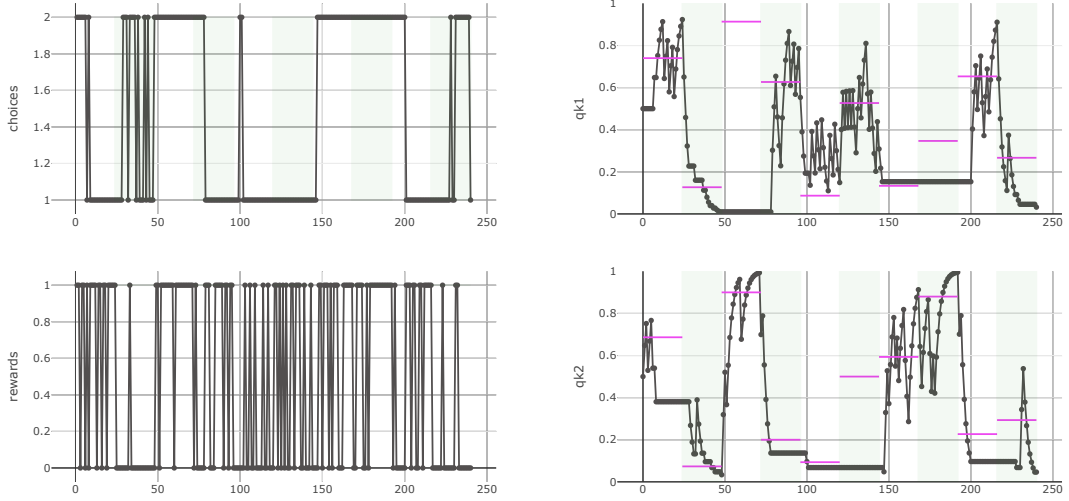


Figure 5.2: Example of data collected through the adaptive settings described in Figure 5.1 with $b = 24$.

cohort of subjects under two comparison settings. First, a static two-armed bandit task, where the reward probabilities remain fixed throughout the experiment at the same initial values used in the adaptive setting. Second, a random setting, where reward probabilities are randomly resampled from a uniform distribution $\text{Unif}(0, 1)$ after each batch.

The experiment framework provided above can be adapted to the different needs of researchers by changing its design features such as the total number of trials or the priors for the ground truth parameters. For the purposes of this work, we restricted our attention to some specific choices. The reward probabilities for the initial data collection have been chosen as the optimal results obtained in Chapter 4. In particular, $(p_1, p_2) = (0.74, 0.69)$. The total number of trials is set to $T = 240$; this is because 240 has a high number of exact divisors and allows for easier division into batches of our experiment. We consider $b = 24$ as a short batch length and $b = 80$ as a long batch length. The real values of the learning rates α and the inverse temperature parameters τ are sampled, respectively, from $\text{Beta}(2, 6)$ and $\text{Gamma}(7.5, 0.5)$ as in Chapter 4.

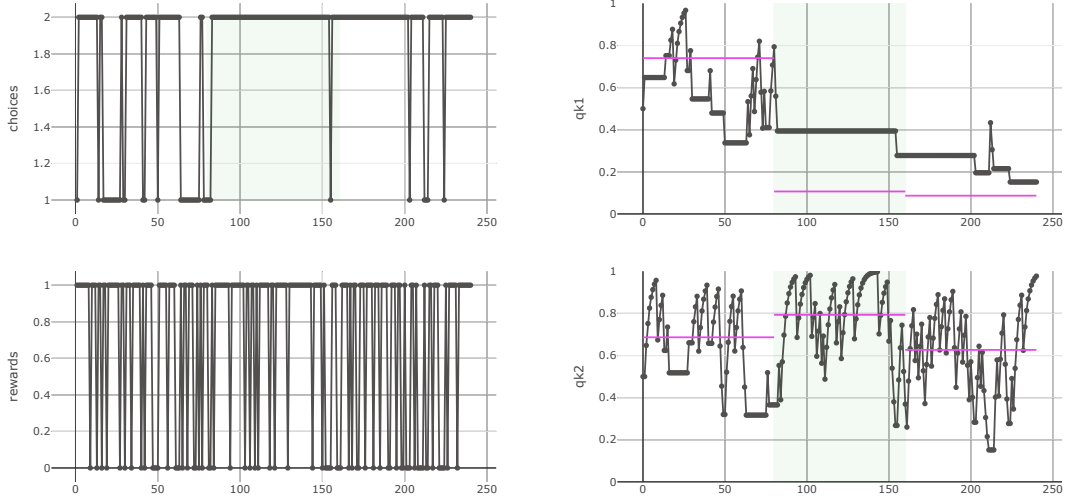


Figure 5.3: Example of data collected through the adaptive settings described in Figure 5.1 with $b = 80$.

Algorithm 4

- 1: Collect b trials.
 - 2: Compute the MLE estimates and update the priors.
 - 3: Apply Algorithm 3.
 - 4: Select the new optimal design using kernel density estimation.
 - 5: Collect b trials with the updated optimal design.
 - 6: Repeat steps 2-5 until T trials have been collected.
-

5.2 Selection of the new optimal design

After each batch, we adaptively change the experimental designs — the reward probabilities of the two-armed bandit task — in order to tailor the experiment to the individual subjects. The update consists of three main steps: the update of the priors, the use of Algorithm 3 and the selection of the next designs based on a kernel density estimation method. The new designs are then used to collect a new batch of data. We repeat these steps until the number of trials set by the researcher has been collected, see Algorithm 4.

Prior update After each batch, new information about the behaviour of a subject is available thanks to the new choices and rewards collected. Our aim is to include this new information into the selection process of the next experimental design so that it is better tailored to the individual subject. The selection is based on Algorithm 3,

see Section 4.3. The prior history of the subject is passed to this algorithm in two different ways.

First, the past history of choices and rewards from all previous batches is assumed known within Algorithm 3. Thus, this information is included in the MLE estimations of the Rescorla-Wagner parameters during the MCMC simulations. As the utility has been defined using these estimations, it follows that the selection of the new experimental design considers the prior history of choices and rewards performed by specific subjects.

Second, after each batch, we update the priors used in Algorithm 3. These priors are used to simulate data in the MCMC chain. The priors are updated iteratively to describe the subject. This ensures that we generate data closer to the expected behaviour of this specific subject. Subsequently, the algorithm returns a new optimal design specifically tailored to the individual.

We explain below how the update of the priors is performed. In order to develop code that is easy to interpret and implement, we relied on simple updating formulas based on the sample mean, the unbiased variance and the method of moments. In particular, after each batch we compute the sample mean and unbiased variance estimates for the Rescorla-Wagner parameters. Then we recover the parameters for the Beta and Gamma priors using the method of moments. The initial priors $\text{Beta}(a_0, b_0)$ and $\text{Gamma}(v_0, z_0)$ are considered as initial data points for the updates.

As in Chapter 4, we choose as prior for the learning rate a Beta distribution, i.e. $\alpha \sim \text{Beta}(a_0, b_0)$ and

$$p(\alpha) = \frac{1}{B(a_0, b_0)} \alpha^{a_0-1} (1 - \alpha)^{b_0-1},$$

where $a_0, b_0 > 0$, $B(a_0, b_0) = \frac{\Gamma(a_0)\Gamma(b_0)}{\Gamma(a_0+b_0)}$ and Γ is the Gamma function. For the inverse temperature parameter we choose a Gamma distribution, i.e. $\tau \sim \text{Gamma}(v_0, z_0)$ and

$$p(\tau) = \frac{z_0^{v_0}}{\Gamma(v_0)} \tau^{v_0-1} e^{-z_0\tau},$$

where v_0 is the shape and z_0 is the rate.

Let $\hat{\alpha}_i, \hat{\tau}_i$ with $i = 1, \dots, k$ be the sequence of estimated learning rates and inverse temperature parameters. First, we utilise the iterative formulas for the sample mean and the unbiased variance

$$\begin{aligned}\hat{\mu}_n &= \frac{1}{n} \sum_{i=1}^n x_i = \frac{(n-1)\hat{\mu}_{n-1} + x_n}{n} & n > 1 \\ \hat{\sigma}_n^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{\mu}_n)^2 = \frac{n-2}{n-1} \hat{\sigma}_{n-1}^2 + \frac{(x_n - \hat{\mu}_{n-1})^2}{n} & n > 1,\end{aligned}\tag{5.4}$$

where $\hat{\mu}_1 = x_1$ and $\hat{\sigma}_1^2 = 0$.

For the learning rate α , Equations (5.4) can be rewritten as

$$\begin{aligned}\hat{\mu}_{0,\alpha} &= \frac{a_0}{a_0 + b_0} \\ \hat{\sigma}_{0,\alpha}^2 &= \frac{a_0 b_0}{(a_0 + b_0)^2 (a_0 + b_0 + 1)} \\ \hat{\mu}_{i,\alpha} &= \frac{1}{(i+1)} \hat{\alpha}_i + \frac{i}{(i+1)} \hat{\mu}_{i-1,\alpha} & i = 1, \dots, k-1 \\ \hat{\sigma}_{i,\alpha}^2 &= \frac{(i-1)}{i} \hat{\sigma}_{i-1,\alpha}^2 + \frac{(\hat{\alpha}_i - \hat{\mu}_{i-1,\alpha})^2}{(i+1)} & i = 1, \dots, k-1.\end{aligned}$$

The iterative formulas in Equations (5.4) are usually used from the first data point onward. In this work, we replace the first data point with chosen priors, so that the first estimates of the model parameters are considered as the second data point in the iterative formulas update. This explains the shift in the index from n to $i+1$. Note that the variance from the initial priors is not considered in the update rule, and the application of methods which do not encounter this limitation might be an interesting area of future research.

Assume that for all $i = 1, \dots, k$ it holds that

$$\hat{\mu}_{i,\alpha}(1 - \hat{\mu}_{i,\alpha}) > \hat{\sigma}_{i,\alpha}^2.\tag{5.5}$$

Then, using the method of moments, it is possible to recover the parameters to

update the priors $\text{Beta}(a_i, b_i)$ as follows

$$\begin{aligned} a_i &= \hat{\mu}_{i,\alpha} \left(\frac{\hat{\mu}_{i,\alpha}(1 - \hat{\mu}_{i,\alpha})}{\hat{\sigma}_{i,\alpha}^2} - 1 \right) & i = 1, \dots, k-1 \\ b_i &= (1 - \hat{\mu}_{i,\alpha}) \left(\frac{\hat{\mu}_{i,\alpha}(1 - \hat{\mu}_{i,\alpha})}{\hat{\sigma}_{i,\alpha}^2} - 1 \right) & i = 1, \dots, k-1. \end{aligned}$$

If Condition (5.5) is not satisfied, our implementation naively resets to the priors used in the previous batch.

A similar approach is used to update the priors $\text{Gamma}(v_i, z_i)$ for the inverse temperature parameter τ . First, we iteratively estimate the sample mean and the unbiased variance

$$\begin{aligned} \hat{\mu}_{0,\tau} &= \frac{v_0}{z_0} \\ \hat{\sigma}_{0,\tau}^2 &= \frac{v_0}{z_0^2} \\ \hat{\mu}_{i,\tau} &= \frac{1}{(i+1)} \hat{\tau}_i + \frac{i}{(i+1)} \hat{\mu}_{i-1,\tau} & i = 1, \dots, k-1 \\ \hat{\sigma}_{i,\tau}^2 &= \frac{(i-1)}{i} \hat{\sigma}_{i-1,\tau}^2 + \frac{(\hat{\tau}_i - \hat{\mu}_{i-1,\tau})^2}{(i+1)} & i = 1, \dots, k-1. \end{aligned}$$

Then, assuming $\hat{\sigma}_{i-1,\tau}^2, \hat{\mu}_{i,\tau} > 0$ for all $i = 1, \dots, k$, the parameters v_i, z_i for $i = 1, \dots, k$ are updated using the method of moments as follows

$$\begin{aligned} v_i &= \frac{\hat{\mu}_{i,\tau}^2}{\hat{\sigma}_{i,\tau}^2} & i = 1, \dots, k-1 \\ z_i &= \frac{\hat{\mu}_{i,\tau}}{\hat{\sigma}_{i,\tau}^2} & i = 1, \dots, k-1. \end{aligned}$$

In this work, the initial priors are defined as $\alpha \sim \text{Beta}(a_0, b_0)$ and $\tau \sim \text{Gamma}(v_0, z_0)$, where

$$a_0 = 2, \quad b_0 = 6, \quad v_0 = 7.5, \quad z_0 = 0.5.$$

Then, the mean and variance for the Beta distribution $\text{Beta}(a_0, b_0)$ are

$$\mu_{0,\alpha} = 0.25, \quad \sigma_{0,\alpha}^2 = 0.02,$$

while for the Gamma distribution $\text{Gamma}(v_0, z_0)$ they are

$$\mu_{0,\tau} = 17, \quad \sigma_{0,\tau}^2 = 30.$$

Evaluation of the design variable space As a design optimization method, we utilise the MCMC algorithm developed in Müller et al. (2004), see Algorithm 3. We previously utilised the algorithm in Chapter 4 to identify the optimal design for a static two-armed bandit task. In such a context, the algorithm is applied before the experiment starts. Then, the experimental designs remain fixed for the duration of the experiment once they are selected. In this chapter, however, the design optimization analysis is performed multiple times throughout the experiment to continuously update the experimental designs. The code utilised in Chapter 4 was slightly modified to achieve this goal by including data collected in previous batches in the analysis.

As in Chapter 4, the utility function can be selected to fit the purposes of the researcher. To analyse the performance of this method in its adaptive form, we chose the same utility used in Chapter 4 for the static counterpart, see Equation (4.6). The utility focuses on improving the MLE estimation of the Rescorla-Wagner parameters. Thus, the sequence of designs selected through Algorithm 3 is tailored to the individual subjects in order to improve the estimation of these parameters through a better-informed data collection.

Kernel density estimation Algorithm 3 provides simulations of the reward probabilities. These are also an approximation of the density for the global utility of the DO problem due to the definition of the augmented probability in the algorithm. Therefore, it is necessary to define a decision method to select a single new optimal design based on the results provided by this density approximation. For this purpose, we utilised a black-box kernel density estimation method in R using the `ks` package (Duong, 2007).

5.3 Results

In order to analyse the performance of Algorithm 4, we create a dataset of simulated experiments as described in Section 5.1. Let C be the set of all the simulated subjects. Denote the ground truth — the learning rate and the inverse temperature parameter — as, respectively, $\alpha_{\text{true}}^j \in [0, 1]$ and $\tau_{\text{true}}^j \in [0, 100]$ for any subject $j \in C$. Then, using the Rescorla-Wagner model, we simulate the response of these subjects to a two-armed bandit task. We use a static optimal design selected through the DO method in Chapter 4, a sequence of randomly sampled designs and an adaptive sequence of optimal designs selected using Algorithm 4.

The MCMC simulations required by Algorithm 3 represent the majority of the computational costs of the adaptive method described in Algorithm 4. Furthermore, Algorithm 3 is applied multiple times throughout the experiment — $(k-1) \cdot |C|$ times. Thus, the computational costs of the adaptive approach increase as the number of subjects increases and/or the length of the batches b decreases. At the same time, a higher number of simulations increases the total computational cost, which can be summarized as $(k-1) \cdot |C| \cdot N$, where N is the number of simulations. We utilised $N = 10\,000$ simulations (with a burn-in of 1 000) in an attempt to balance the computational costs of the experiments and the necessity to run the MCMC chains long enough to obtain reliable results with the DO method.

In all cases, we record the sequence of estimated parameters denoted as

$$\begin{aligned}\hat{\boldsymbol{\alpha}}_m^j &= (\hat{\alpha}_{m,1}^j, \dots, \hat{\alpha}_{m,k}^j) = (\hat{\alpha}_{m,i}^j)_{i=1}^k \in [0, 1]^k \\ \hat{\boldsymbol{\tau}}_m^j &= (\hat{\tau}_{m,1}^j, \dots, \hat{\tau}_{m,k}^j) = (\hat{\tau}_{m,i}^j)_{i=1}^k \in [0, 1]^k,\end{aligned}$$

where $j \in C$ denotes the subject and $m \in \{a, r, s\}$ indicates the method applied to select the experimental designs — respectively, adaptive, random or static. For the adaptive and random cases, we also save the trajectory of the experimental designs

$$D_m^j = (\mathbf{d}_{m,1}^j, \dots, \mathbf{d}_{m,k}^j) = (\mathbf{d}_{m,i}^j)_{i=1}^k \in [0, 1]^{2 \times k},$$

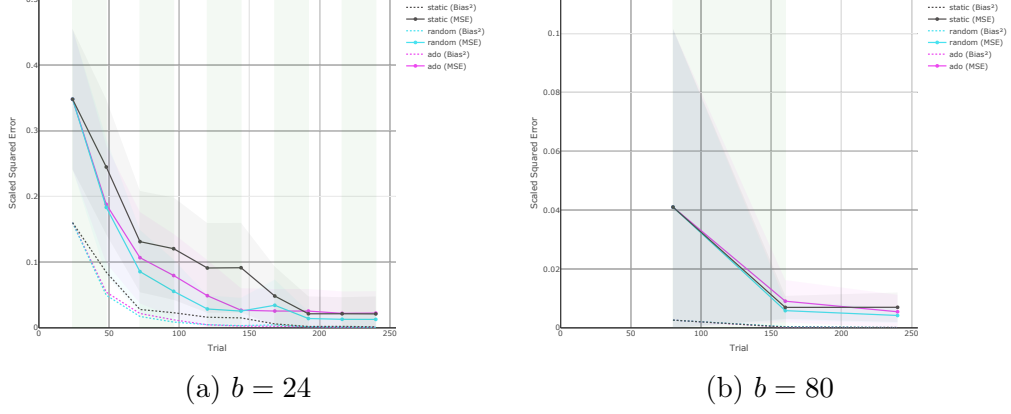


Figure 5.4: Comparison of the parameter estimation error between the adaptive, the random and the static approaches for batch lengths of 24 and 80 for a total of 240 trials. The plot on the left shows the average errors for batches of length 24, while the plot on the right concerns batches of length 80. The x-axis represents the trials of the simulated two-armed bandit experiment. Errors were computed after each batch. The adaptive results are shown in pink, the random in light blue and the static ones are shown in black. In both cases, the cohort is composed of 40 subjects. The total error is decomposed into squared bias and variance components. The dotted lines represent the squared bias, while the variance is represented by the difference between the mean squared error (solid line) and the bias line. For longer batch sizes, the bias remains consistently low (and overlaps across different algorithms), meaning that most of the overall error consists of variance. For the shorter batch length case, the bias is substantially higher, especially during the earlier trials.

where $i \in \{1, \dots, k\}$ identifies the batch and $m \in \{a, r\}$ denotes the method.

For the analysis of the results we focus on three different aspects of the Algorithm 4 performance: the impact on the parameter estimation compared to a static and random version, the pattern in the experimental design updates and the regret.

Estimation error First, we evaluate the impact on the estimation of the learning rate and the inverse temperature parameters for the Rescorla-Wagner model and compare it against a static and random approaches. In particular, for $i \in \{1, \dots, k\}$ and $m \in \{a, r, s\}$, we consider the average squared errors

$$\text{Err}_{m,i} = \frac{1}{|C|} \sum_{j \in C} \left((\hat{\alpha}_{m,i}^j - \alpha_{\text{true}}^j)^2 + \left(\frac{1}{100} (\hat{\tau}_{m,i}^j - \tau_{\text{true}}^j) \right)^2 \right).$$

Notice that the inverse temperature parameter error is scaled back between 0 and 1 using a factor $\frac{1}{100}$; the ground truth α_{true}^j and τ_{true}^j are assumed known for all $j \in C$.

We plot the errors in Figure 5.4 for a shorter batch length of 24 and a longer one of 80. For long batch lengths, our results suggest that all algorithms achieve similar performance. Conversely, for short batch lengths, the adaptive algorithm shows promising results, indicating that dynamically adjusting designs can indeed improve parameter recovery. Nevertheless, these gains appear moderate, particularly when compared to a random approach. This behaviour likely stems from the presence of many near-optimal experimental designs and a scarcity of severely underperforming ones (see Figure 4.5 in Chapter 4). Consequently, the random algorithm likely performs comparably to the adaptive algorithm in most scenarios. While the adaptive approach may offer greater robustness against underperforming design choices, dedicated analysis would be required to fully verify this impact.

Remark It is possible to segment the analysis to include only $\alpha \in [x_1, x_2] \subsetneq [0, 1]$ or $\tau \in [x_1, x_2] \subsetneq (0, 100]$. This could be useful in highlighting the differences between the two approaches for specific cases of α and τ such as "unusual" true values of these parameters according to our initial priors (values far away from $\alpha = 0.25$ and $\tau = 15$). Furthermore, our analysis only considers an optimal design for the static algorithm. This implies prior knowledge of the optimal conditions for the experiment at least in static settings. While DO methods are effective tools to inform researchers on this topic, the inherent variability of real-world scenarios may be better addressed through adaptive data collection strategies, which avoid relying solely on a single experimental design.

Design trajectories Second, we focus on the experimental designs D^j , with $j \in C$. For each subject, the sequence of experimental designs updated through the adaptive approach can be seen as a trajectory in the design variable space. Investigating these trajectories allows us to highlight patterns in the updates performed by the algorithm and which designs become more desirable as the experiment progresses over time. In order to achieve this goal, we defined a set of measures which summarize in a scalar value various aspects of the trajectories. Then, we analyse their distribution across

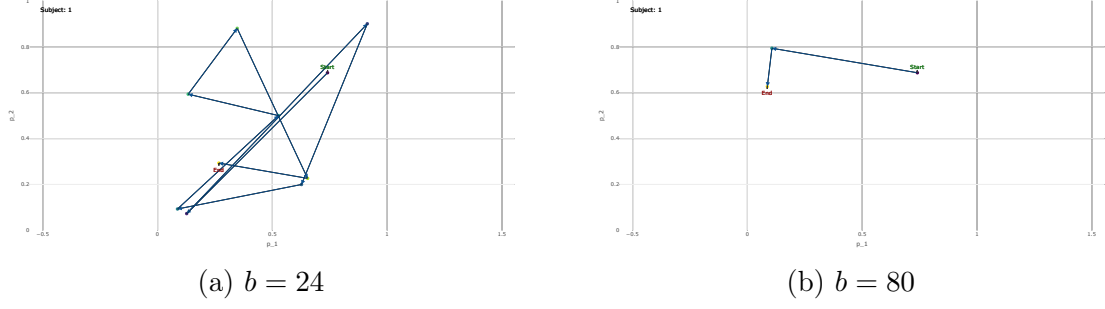


Figure 5.5: Example of an experimental design trajectory D^j . In Figures 5.5a and 5.5b, we plot an example of the design trajectory for, respectively, batches of length 24 and 80. The first experimental design $\mathbf{d}_1^j$ is fixed for all $j \in C$. The sequence of designs selected by the algorithm for this simulated subject is plotted in the design variable space D . Two consecutive points are connected by arrows, which indicate the direction of the trajectory.

the cohort of subjects.

As mentioned before, any sequence of designs D^j can be seen as a discrete trajectory in the design variable space $D = [0, 1]^2$. Indeed,

$$D^j : \{1, \dots, k\} \longrightarrow [0, 1]^2$$

$$i \longrightarrow \mathbf{d}_i^j = (p_{1,i}^j, p_{2,i}^j).$$

In Figure 5.5 we plot an example of experimental design trajectory D^j for batch lengths equal to 24 and 80.

Define $L_0 : [0, 1]^{2 \times k} \longrightarrow \mathbb{R}$ as

$$L_0(D^j) = \frac{1}{k-1} \sum_{i=1}^{k-1} \left((p_{1,i}^j - p_{1,i+1}^j)^2 + (p_{2,i}^j - p_{2,i+1}^j)^2 \right)^{\frac{1}{2}}.$$

The measure L_0 provides information about the length of each segment of the trajectory D^j . In particular, it quantifies the variability in the trajectory and measures how much it moves around the design variable space.

It is possible to modify the definition of L_0 to weight differently the segments of the trajectory. Consider $L_1 : [0, 1]^{2 \times k} \longrightarrow \mathbb{R}$ defined as

$$L_1(D^j) = \frac{2}{(k-1)k} \sum_{i=1}^{k-1} i \left((p_{1,i}^j - p_{1,i+1}^j)^2 + (p_{2,i}^j - p_{2,i+1}^j)^2 \right)^{\frac{1}{2}}.$$

The terms i for $i = 1, \dots, k-1$ assign more weight to distances between experimental designs encountered in later stages of the experiment, as i increases. Therefore, the measure $L_1(D^j)$ represents how much the trajectory moves in the design variable space, with particular focus on the end of the experiment. A comparison between $L_0(\cdot)$ and $L_1(\cdot)$ shows if the adaptive approach suggests very different designs from the previous ones consistently throughout the experiment — $L_0(\cdot) \sim L_1(\cdot)$ — or if it tends to take longer steps at the end of the experiment — $L_0(\cdot) \leq L_1(\cdot)$.

We include in our analysis two additional measures. We define $L_2 : [0, 1]^{2 \times k} \rightarrow \mathbb{R}$ as

$$L_2(D^j) = \frac{1}{k} \sum_{i=1}^k \left(|p_{1,i}^j| + |p_{2,i}^j| \right).$$

This measure quantifies how high the values of p_1 and p_2 tend to be. On the other hand, $L_3 : [0, 1]^{2k} \rightarrow \mathbb{R}$ defined as

$$L_3(D^j) = \frac{1}{k} \sum_{i=1}^k \left(|p_{1,i}^j - p_{2,i}^j| \right),$$

measures the average difference between the two arms' reward probabilities. As before, it is possible to add weights based on the time/index to highlight the behaviour of the trajectories over time.

In Figures 5.6a and 5.6b, we summarize the measures' results: we plot boxplots of the measures' distributions across the cohort C . The average distance between two consecutive experimental designs $\mathbf{d}_i^j$ tends to increase as the experiment progresses, as the boxplot relative to L_1 is higher than the boxplot relative to L_0 for both the batch lengths considered. This can be explained by the decreasing pattern of the regret observed in Chapter 4. Indeed, as the utility flattens as the number of trials increases, the algorithm can select experimental designs in a larger area. The sum of the two reward probabilities is, for both scenarios, close to 1, while the difference of the reward probabilities in each experimental design is, on average, relatively low. This suggests that, as shown in the experimental trajectory examples, the next experimental designs are chosen in the central area of the design variable space D so

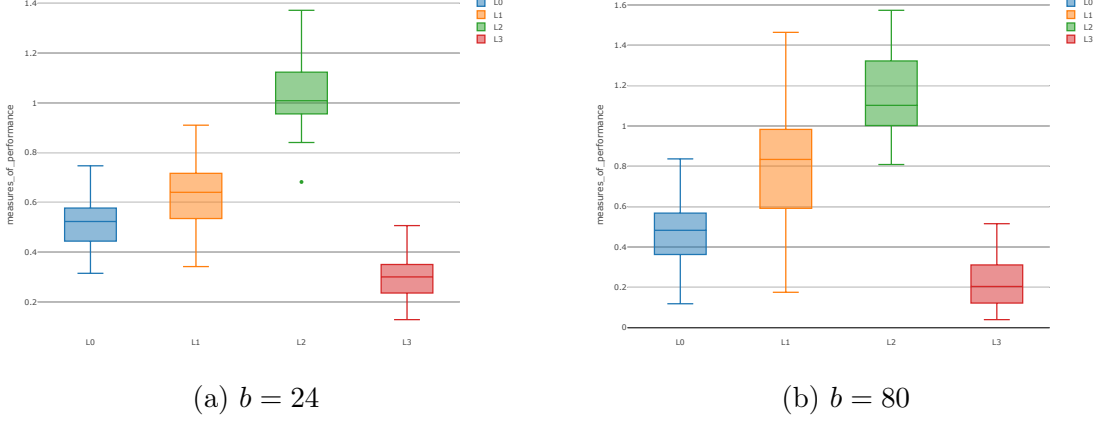


Figure 5.6: In Figures 5.6a and 5.6b we plot the boxplot of the performance measures L_i , $i = 0, \dots, 3$ across our cohort of subjects.

that one arm is just slightly better than the other.

Regret To further analyse the performance of Algorithm 4, we investigate the regret of continuing the experiment without updating the experimental designs to the optimal choice according to the DO results. Thanks to the use of MCMC chains in Algorithm 3, we can obtain the kernel density estimation of the simulated reward probabilities density. As this density is proportional to the utility of the design optimization problem, the kernel density estimation also provides an indication of the regret about selecting a design over another.

Therefore, for each subject $j \in C$ and each batch update $i = 1, \dots, k - 1$, we save three quantities: the maximum value of the kernel density $\delta_{\max}^{i,j}$, the minimum value $\delta_{\min}^{i,j}$ and the kernel density value at the previous design $\delta_{\text{old}}^{i,j}$. For each batch, we define the *relative regret* R_i as

$$R_i = \frac{1}{|C|} \sum_{j \in C} \frac{\delta_{\max}^{i,j} - \delta_{\text{old}}^{i,j}}{\delta_{\max}^{i,j} - \delta_{\min}^{i,j}}.$$

This function measures the average regret of not updating the optimal design after the DO application at batch i relative to the range of the reward probabilities density. Note that the regret in this chapter is computed differently from Chapter 4. This definition of regret is computationally cheaper as the adaptive case is computationally more intensive than the static case.

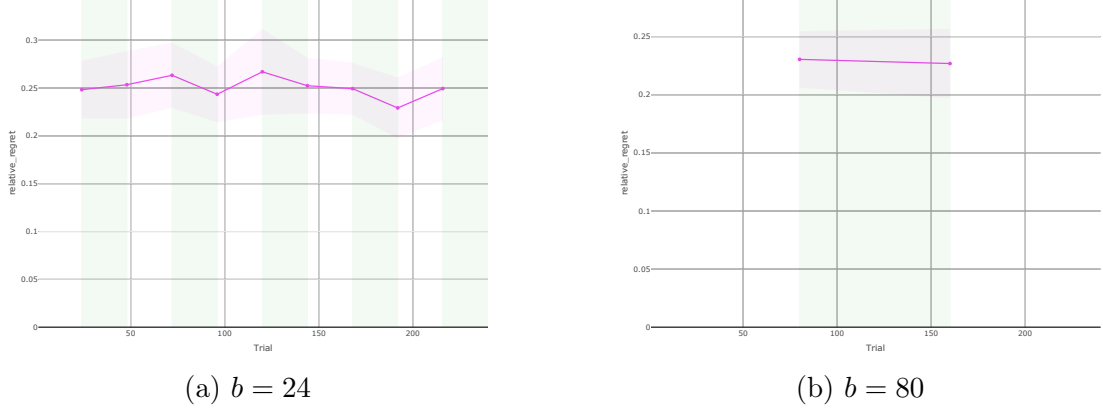


Figure 5.7: Regret for two cohorts of 40 subjects with respectively batch lengths equal to 24 or 80.

The relative regret R_i for $i = 1, \dots, k - 1$ is plotted in Figure 5.7 for batch lengths of 24 and 80. For both batch lengths, the regret remains almost constant and positive during the experiment. Thus, the design optimization method always finds a better design, for a similar gain for all batch lengths.

Practical Considerations Because the adaptive framework in this chapter builds directly upon the learning models introduced in Chapter 4, it shares the same core assumptions and limitations. We suggest carefully reviewing them as well. Moreover, the adaptive nature of this approach increases the importance of one of these assumptions: the stationarity of the subject’s underlying cognitive model and parameters. It is assumed that the subject’s behaviour is described by a fixed model and that the true parameters (α and τ) remain constant over time. This is a strong assumption in a static environment, but even more so in a volatile one. As the reward probabilities change adaptively from batch to batch, we would expect the subject’s behaviour to shift as well. This could translate into non-static parameters or even different models for different batches. We do not address this issue in this work, but refer to it as an interesting future direction. *A posteriori* checks on real data, such as fitting the model independently on early versus late batches, can be used to verify that the subject’s fundamental learning strategy has not shifted over time.

The translation of this adaptive approach into real-world applications requires the

Diagnostics & practical deployment checklist

1. Assumptions

- **Assumptions from the static case:** model suitability, utility alignment and reward distribution matching.
Diagnostics: see Chapter 4.
- **Model and parameter stationarity:** subjects do not change their fundamental learning strategy when the task design is adaptively updated.
Diagnostic: fit the model independently on different batches to verify that the parameter estimates remain stable over the course of the experiment.
- **Density estimation stability:** the kernel density estimation accurately identifies the peak utility for the new design.
Diagnostic: visually inspect the density maps from the MCMC output to ensure a well-defined global optimum exists before updating the design.

2. Practical deployment checklist

- **Known *a priori* (by practitioner):** the batch length b , the total trials T , and the parameters of the initial priors.
- **Computational cost:** the algorithm requires running computationally expensive MCMC simulations between batches. In a live empirical setting, this latency must be managed carefully to avoid disrupting the subject's focus.
- **Batch size trade-off:** adaptive settings are likely more valuable for shorter batch sizes, but longer batch sizes provide more data and are easier to deliver.

experimenter to set the batch length b and the total number of trials T . Furthermore, the same experimental decisions as in Chapter 4 must be made: the model, the utility and the priors. In a real-world application of this approach, the practitioner would face two major challenges: the computational cost and the choice of a batch length b . First, the MCMC simulations must be run between batches to update the design, which introduces significant computational latency during a live experiment. For this method to work, multiple hours, if not days, are needed between batches, and such delays might shift the subject's behaviour, breaking the stationarity assumption. Second, there is a critical trade-off in the selection of the batch length b . On the one hand, longer batch sizes provide better estimations, condense experimental data

collection and, likely, simplify the deployment of the experiment. On the other hand, shorter batches likely provide the best framework to exploit adaptive settings as the smaller number of trials makes the careful selection of the next experimental design more relevant (see Chapter 4 and Figure 5.4).

5.4 Conclusions and future directions

In conclusion, we analysed how a design optimization approach based on Müller et al. (2004) performs in a non-static two-armed bandit task for the Rescorla-Wagner model. In particular, we considered an adaptive approach that iteratively updates the reward probabilities based on the prior history of the subject. Our results suggest that the adaptive approach performs slightly better than a static approach for shorter batch sizes. The effects of this approach are not noticeable for longer batch lengths. The adaptive approach tends to select reward probabilities in the centre of the design variable space, avoiding very high or very low reward probabilities. Moreover, the algorithm tends to select the reward probability for the first arm close to that of the second arm, likely to make the two-armed bandit task non-trivial. Using the information collected through the MCMC simulations, we also considered the regret of not changing the reward probabilities according to the design optimization results in the adaptive approach. Our results show that changing the design is always valuable, and the regret of not doing so remains almost constant regardless of the batch lengths. Nevertheless, the major gains from employing this method seem to be in the shorter batch length case.

We highlight here some limitations and future directions of this work. As with the static approach described in Chapter 4, the results shown here are limited to one specific model — the Rescorla-Wagner model. While the method can be adapted to other models, these changes were not implemented in our code or considered in this work. The extension of this work to other models, such as the ones already considered in Chapter 4, is a possible future direction of investigation. Furthermore, as in Chapter 4, we did not use the Bayesian approach to estimate the parameters,

but rather worked around the natural updates of the Bayesian priors using method-of-moments updates. This allowed for an easier implementation of our code, but an extension to the Bayesian case using tools such as the `hBayesDM` package could be an interesting area of research (Ahn et al., 2017). Another limitation of this approach is the computational cost of running the MCMC chain. The MCMC approach allows for a greater understanding of the value of an adaptive approach compared to a static one, but it could be quite expensive compared to cheaper optimization algorithms. Thus, a comparison in terms of both performance and computational costs is a possible direction of future work. The code implemented to obtain the results in this chapter can be easily adapted to an R package, which would increase the usability of the tools utilised here for interested practitioners.

Finally, the work presented in this chapter is strictly theoretical and simulative. Nevertheless, experiments with human subjects might provide useful insights into the effectiveness of the methods analysed here in applicative settings. In particular, observing the possible reactions of subjects to the adaptive experimental design settings is an interesting area for future work.

Chapter 6

Conclusion

This work focuses on applying design optimization methods to experiments based on multi-armed bandit tasks (MABTs) within computational psychology. These experiments often require a high degree of preparation and lengthy data collection to provide reliable estimations of the parameters. Thus, ensuring data quality is particularly important when subjects may struggle to remain engaged in the study due to the duration of the experiments. We aimed to evaluate the impact of design optimization methods on parameter recovery, both before the experiment starts and during it. Indeed, mobile applications provide new opportunities for outside-the-lab data collection and, in particular, enable the tailoring of experimental designs to individual subjects during online experiments. We address this issue from three different perspectives.

First, we considered when an MABT experiment should be administered to subjects, assuming that this data has to be collected for high or low values of another control variable. We developed an algorithm that tracks adherence and the control variable history to select optimal times for the administration of a burdensome task, such as an MABT experiment. We applied this algorithm to both simulated settings and a real dataset of tinnitus-affected patients. This approach is particularly effective when adherence impacts data collection.

Second, we investigated how the selection of the reward probabilities through a design optimization method affects parameter recovery in MABTs. We considered

three different models: the Rescorla-Wagner model, an extension of the Rescorla-Wagner model and a Win-Stay-Lose-Shift model. We applied a DO method based on Müller et al. (2004) and investigated the regret of selecting non-optimal designs. We find that, for the Rescorla-Wagner model and its extension, reward probabilities close to each other, but not too high or low, are preferred. On the other hand, for the Win-Stay-Lose-Shift model, parameter recovery does not seem affected by the reward probabilities chosen as experimental designs.

Finally, we extended our analysis to a non-static setting and investigated how frequent changes in the reward probabilities of a two-armed bandit task impact the parameter estimation. In this case, we only considered the Rescorla-Wagner model. We divided the experiments into equal-length batches and extended the methods applied for the static settings to this online case. We identified an optimal experimental design after each batch and provided an analysis of the regret. We found that adaptive methods slightly improve the parameter estimation for shorter batch lengths.

According to our results, adaptive design optimization methods improve data collection; however, we only considered a small set of cases. More work in this area is needed to provide practitioners with efficient and accessible tools to inform their experimental design choices for MABT experiments in the psychology field. There are two main future lines of research. First, the results of this work can be improved from a technical viewpoint, focusing on the computational/statistical tools and simulative results. For example, future work might integrate Bayesian methods, evaluate the impact of different choices for the priors, use different statistical models, or consider different optimization methods. Second, more work can concentrate on the applicative settings, focusing on the collection and analysis of real-world data, as well as the delivery of accessible packages. Progress in this area might have a major impact on the appeal of these methods for practitioners.

One of the possible areas of research is the extension of the results of this work to different statistical models. Design optimization methods rely on predetermined

statistical models to select optimal designs. In Chapter 3, we utilised a Bernoulli random variable to model the adherence of the subjects. However, time series might substantially improve the algorithm. Similarly, we considered a restricted number of statistical models for the simulation and evaluation of MABT data compared to the options available in the literature. The extension of the tools developed in this work to these models could provide additional insights for interested practitioners.

Furthermore, it is worth evaluating other optimization methods in Chapters 4 and 5. The methods used in this work, based on the algorithm developed by Müller et al. (2004), are computationally expensive. They have the benefit of providing information on the regret, but they might be too demanding in real-world — especially adaptive — experiments. In particular, a novel approach proposed by Kim et al. (2017), based on reinforcement learning models could have a significant impact on design optimization in this context. While the algorithm is black-box like many AI tools, it avoids a computationally expensive step of the design optimization framework. Nevertheless, we believe the methods used here will remain a valuable source of information for practitioners as they allow researchers to quantify the effective gains of DO methods.

While this thesis mainly focuses on the simulation aspect, a clear area of future development in the applicative area is the collection of real-world data. In Chapter 3, we included a real subject study thanks to the collaboration with the University of Regensburg. However, no real data have been collected for Chapters 4 and 5. While the simulations performed in this work support the application of these methods, the study of their effectiveness in real settings can further highlight the limitations and benefits of DO methods in the area of MABT experiments. For example, we briefly mentioned the subjects' frustration sometimes caused by DO methods. However, the magnitude of these effects in the specific area of MABT experiments remains to be tested in real scenarios. Moreover, DO methods might introduce bias into the data, and studies in this area could reassure interested practitioners. Of particular interest would be an experiment that combines all the results of this work and enables the analysis of intraindividual differences in decision-making using batch-by-batch

MABT experiments.

In this work, we contributed to the application and evaluation of design optimization methods in MABT experiments. The goal was to define new methods and improve data collection by tailoring the experimental design to individual subjects. These tools are especially relevant when data is difficult or expensive to obtain. In these cases, an optimized data collection process can potentially reduce experimental costs or decrease the burden on the subjects whenever the process requires considerable engagement. We hope the tools developed and evaluated here will be useful to interested practitioners and will help design experiments that are more affordable for researchers and less burdensome for subjects.

Bibliography

- Ahn, W.-Y., Dai, J., Vassileva, J., Busemeyer, J. R., & Stout, J. C. (2016). Computational modeling for addiction medicine. In *Progress in Brain Research* (pp. 53–65, Vol. 224). Elsevier. <https://doi.org/10.1016/bs.pbr.2015.07.032>
- Ahn, W.-Y., Haines, N., & Zhang, L. (2017). Revealing neurocomputational mechanisms of reinforcement learning and decision-making with the hBayesDM package. *Computational Psychiatry*, 1, 24–57. https://doi.org/10.1162/CPSY_a_00002
- Aylward, J., Valton, V., Ahn, W.-Y., Bond, R. L., Dayan, P., Roiser, J. P., & Robinson, O. J. (2019). Altered learning under uncertainty in unmedicated mood and anxiety disorders. *Nature Human Behaviour*, 3(10), 1116–1123. <https://doi.org/10.1038/s41562-019-0628-0>
- Berry, D. A. (2006). Bayesian clinical trials. *Nature Reviews Drug Discovery*, 5(1), 27–36. <https://doi.org/10.1038/nrd1927>
- Blanchard, T. C., & Gershman, S. J. (2018). Pure correlates of exploration and exploitation in the human brain. *Cognitive, Affective, & Behavioral Neuroscience*, 18(1), 117–126. <https://doi.org/10.3758/s13415-017-0556-2>
- Blau, T., Bonilla, E. V., Chades, I., & Dezfouli, A. (2022). Optimizing sequential experimental design with deep reinforcement learning. *Proceedings of the 39th International Conference on Machine Learning*, 162, 2107–2128. <https://proceedings.mlr.press/v162/blau22a.html>
- Browning, M., Behrens, T. E., Jocham, G., O'Reilly, J. X., & Bishop, S. J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature Neuroscience*, 18(4), 590–596. <https://doi.org/10.1038/nn.3961>
- Brunner, E., Bathke, A. C., & Konietzschke, F. (2018). *Rank and pseudo-rank procedures for independent observations in factorial designs: Using R and SAS*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-02914-2>
- Buelow, M. T., Okdie, B. M., & Kowalsky, J. M. (2024). Ecological validity of common behavioral decision making tasks: Evidence across two samples. *Journal of Clinical and Experimental Neuropsychology*, 46(3), 187–206. <https://doi.org/10.1080/13803395.2024.2337759>

- Cavagnaro, D. R., Myung, J. I., Pitt, M. A., & Kujala, J. V. (2010). Adaptive design optimization: A mutual information-based approach to model discrimination in cognitive science. *Neural Computation*, 22(4), 887–905. <https://doi.org/10.1162/neco.2009.02-09-959>
- Cavanagh, J. F., Frank, M. J., Klein, T. J., & Allen, J. J. (2010). Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. *NeuroImage*, 49(4), 3198–3209. <https://doi.org/10.1016/j.neuroimage.2009.11.080>
- Chaloner, K., & Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*, 10(3), 273–304. <http://www.jstor.org/stable/2246015>
- Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3), 1–58. <https://doi.org/10.1145/1541880.1541882>
- Cheng, X., Wang, L., Lv, Q., Wu, H., Huang, X., Yuan, J., Sun, X., Zhao, X., Yan, C., & Yi, Z. (2022). Reduced learning bias towards the reward context in medication-naïve first-episode schizophrenia patients. *BMC Psychiatry*, 22(1), 123. <https://doi.org/10.1186/s12888-021-03682-5>
- Dai, J., Kerestes, R., Upton, D. J., Busemeyer, J. R., & Stout, J. C. (2015). An improved cognitive model of the Iowa and Soochow Gambling Tasks with regard to model fitting performance and tests of parameter consistency. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00229>
- Dayan, P., & Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive, Affective, & Behavioral Neuroscience*, 8(4), 429–453. <https://doi.org/10.3758/CABN.8.4.429>
- De Vries, L. P., Baselmans, B. M. L., & Bartels, M. (2021). Smartphone-based ecological momentary assessment of well-being: A systematic review and recommendations for future studies. *Journal of Happiness Studies*, 22(5), 2361–2408. <https://doi.org/10.1007/s10902-020-00324-7>
- Dezfouli, A., Griffiths, K., Ramos, F., Dayan, P., & Balleine, B. W. (2019). Models that learn how humans learn: The case of decision-making and its disorders. *PLOS Computational Biology*, 15(6), e1006903. <https://doi.org/10.1371/journal.pcbi.1006903>
- Duong, T. (2007). **ks** : Kernel density estimation and kernel discriminant analysis for multivariate data in *R*. *Journal of Statistical Software*, 21(7). <https://doi.org/10.18637/jss.v021.i07>
- Ekin, T., Naveiro, R., Ríos Insua, D., & Torres-Barrán, A. (2023). Augmented probability simulation methods for sequential games. *European Journal of Operational Research*, 306(1), 418–430. <https://doi.org/10.1016/j.ejor.2022.06.042>

- Fricker, R. D., Hegler, B. L., & Dunfee, D. A. (2008). Comparing syndromic surveillance detection methods: EARS' *versus* a CUSUM-based methodology. *Statistics in Medicine*, 27(17), 3407–3429. <https://doi.org/10.1002/sim.3197>
- Gershman, S. J. (2015). A unifying probabilistic view of associative learning. *PLOS Computational Biology*, 11(11), e1004567. <https://doi.org/10.1371/journal.pcbi.1004567>
- Gigli, M., & Stella, F. (2025). Multi-armed bandits for performance marketing. *International Journal of Data Science and Analytics*, 20(1), 151–165. <https://doi.org/10.1007/s41060-023-00493-7>
- Haines, N., Vassileva, J., & Ahn, W.-Y. (2018). The outcome-representation learning model: A novel reinforcement learning model of the Iowa gambling task. *Cognitive Science*, 42(8), 2534–2561. <https://doi.org/10.1111/cogs.12688>
- Hekler, E. B., Rivera, D. E., Martin, C. A., Phatak, S. S., Freigoun, M. T., Korinek, E., Klasnja, P., Adams, M. A., & Buman, M. P. (2018). Tutorial for using control systems engineering to optimize adaptive mobile health interventions. *Journal of Medical Internet Research*, 20(6), e214. <https://doi.org/10.2196/jmir.8622>
- Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, 15(8), 534–539. <https://doi.org/10.1111/j.0956-7976.2004.00715.x>
- Hulme, W. J., Martin, G. P., Sperrin, M., Casson, A. J., Bucci, S., Lewis, S., & Peek, N. (2021). Adaptive symptom monitoring using hidden Markov models – an application in ecological momentary assessment. *IEEE Journal of Biomedical and Health Informatics*, 25(5), 1770–1780. <https://doi.org/10.1109/JBHI.2020.3031263>
- Jami, A., Abbaszade, S., & Vahabie, A.-H. (2025). A review on exploration–exploitation trade-off in psychiatric disorders. *BMC Psychiatry*, 25(1), 420. <https://doi.org/10.1186/s12888-025-06837-w>
- Ji, L., Chen, M., Oravecz, Z., Cummings, E. M., Lu, Z.-H., & Chow, S.-M. (2020). A Bayesian vector autoregressive model with nonignorable missingness in dependent variables and covariates: Development, evaluation, and application to family processes. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(3), 442–467. <https://doi.org/10.1080/10705511.2019.1623681>
- Karvelis, P., Paulus, M. P., & Diaconescu, A. O. (2023). Individual differences in computational psychiatry: A review of current challenges. *Neuroscience & Biobehavioral Reviews*, 148, 105137. <https://doi.org/10.1016/j.neubiorev.2023.105137>
- Kenning, C., Bower, P., Small, N., Ali, S. M., Brown, B., Dempsey, K., Mackey, E., McMillan, B., Sanders, C., Serafimova, I., Van Der Veer, S. N., Dixon, W. G., & McBeth, J. (2024). Users' views on the use of a smartwatch app to collect daily symptom data in individuals

- with multiple long-term conditions (multimorbidity): A qualitative study. *Journal of Multimorbidity and Comorbidity*, 14, 26335565231220202. <https://doi.org/10.1177/26335565231220202>
- Kim, W., Pitt, M. A., Lu, Z.-L., & Myung, J. I. (2017). Planning beyond the next trial in adaptive experiments: A dynamic programming approach. *Cognitive Science*, 41(8), 2234–2252. <https://doi.org/10.1111/cogs.12467>
- Koizumi, D., Takeshi, M., & Michio, S. (2012). On the automatic detection algorithm of Cross Site Scripting (XSS) with the non-stationary Bernoulli distribution. *The 5th International Conference on Communications, Computers and Applications (MIC-CCA2012)*, 131–135.
- Kruschke, J. K. (2008). Bayesian approaches to associative learning: From passive to active learning. *Learning & Behavior*, 36(3), 210–226. <https://doi.org/10.3758/LB.36.3.210>
- Kwon, M., Lee, S. H., & Ahn, W.-Y. (2023). Adaptive design optimization as a promising tool for reliable and efficient computational fingerprinting. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 8(8), 798–804. <https://doi.org/10.1016/j.bpsc.2022.12.003>
- Leqi, L., Zhou, G., Kilinc-Karzan, F., Lipton, Z., & Montgomery, A. (2023). A field test of bandit algorithms for recommendations: Understanding the validity of assumptions on human preferences in multi-armed bandits. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–16. <https://doi.org/10.1145/3544548.3580670>
- Lin, B., Cecchi, G., Bouneffouf, D., Reinen, J., & Rish, I. (2020). A story of two streams: Reinforcement learning models from human behavior and neuropsychiatry. *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent System*, 744–752. <https://dl.acm.org/doi/10.5555/3398761.3398850>
- Lookman, T., Balachandran, P. V., Xue, D., & Yuan, R. (2019). Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. *npj Computational Materials*, 5(1), 21. <https://doi.org/10.1038/s41524-019-0153-8>
- Loosen, A. M., Kato, A., & Gu, X. (2025). Revisiting the role of computational neuroimaging in the era of integrative neuroscience. *Neuropsychopharmacology*, 50(1), 103–113. <https://doi.org/10.1038/s41386-024-01946-8>
- MacDonald, S. W., Nyberg, L., & Bäckman, L. (2006). Intra-individual variability in behavior: Links to brain structure, neurotransmission and neuronal activity. *Trends in Neurosciences*, 29(8), 474–480. <https://doi.org/10.1016/j.tins.2006.06.011>
- McClaine, S., Fedor, J., Bartel, C., Chen, L., Durica, K. C., & Low, C. A. (2024). Engagement with daily symptom reporting, passive smartphone sensing, and wearable device data collection during chemotherapy: Longitudinal observational study. *JMIR Cancer*, 10, e57347. <https://doi.org/10.2196/57347>

- Melinscak, F., & Bach, D. R. (2020). Computational optimization of associative learning experiments (A. Soltani, Ed.). *PLOS Computational Biology*, 16(1), e1007593. <https://doi.org/10.1371/journal.pcbi.1007593>
- Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin*, 117(3), 363–386. <https://doi.org/10.1037/0033-2909.117.3.363>
- Mohan, S. (2021). Exploring the role of common model of cognition in designing adaptive coaching interactions for health behavior change. *ACM Transactions on Interactive Intelligent Systems*, 11(1), 1–30. <https://doi.org/10.1145/3375790>
- Müller, P., Sansó, B., & De Iorio, M. (2004). Optimal Bayesian design by inhomogeneous Markov chain simulation. *Journal of the American Statistical Association*, 99(467), 788–798. <https://doi.org/10.1198/016214504000001123>
- Nogas, J., Li, T., Yanez, F. J., Modiri, A., Deliu, N., Prystawski, B., Villar, S. S., Rafferty, A., & Williams, J. J. (2021). Algorithms for adaptive experiments that trade-off statistical analysis with reward: Combining uniform random assignment and reward maximization. <https://neurips.cc/virtual/2021/33886>
- Padoa-Schioppa, C. (2011). Neurobiology of economic choice: A good-based model. *Annual Review of Neuroscience*, 34(1), 333–359. <https://doi.org/10.1146/annurev-neuro-061010-113648>
- Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, 442(7106), 1042–1045. <https://doi.org/10.1038/nature05051>
- Pillow, J., & Park, M. (2016). Adaptive Bayesian methods for closed-loop neurophysiology. In *Closed Loop Neuroscience* (pp. 3–18). Elsevier. <https://doi.org/10.1016/B978-0-12-802452-2.00001-9>
- Pitt, M. A., & Myung, J. I. (2019). Robust modeling through design optimization. *Computational Brain & Behavior*, 2(3-4), 200–201. <https://doi.org/10.1007/s42113-019-00050-1>
- Probst, T., Pryss, R., Langguth, B., & Schlee, W. (2016). Emotional states as mediators between tinnitus loudness and tinnitus distress in daily life: Results from the “TrackYourTinnitus” application. *Scientific Reports*, 6(1), 20382. <https://doi.org/10.1038/srep20382>
- Ryan, E. G., Drovandi, C. C., McGree, J. M., & Pettitt, A. N. (2016). A review of modern computational algorithms for Bayesian optimal design: A review of modern algorithms for Bayesian design. *International Statistical Review*, 84(1), 128–154. <https://doi.org/10.1111/insr.12107>
- Sazhin, D., Dachs, A., & Smith, D. V. (2025). Meta-analysis reveals that explore–exploit decisions are dissociable by activation in the dorsal lateral prefrontal cortex, anterior insula, and

- dorsal anterior cingulate cortex. *European Journal of Neuroscience*, 61(6), e70081. <https://doi.org/10.1111/ejn.70081>
- Shadlen, M. N., & Kiani, R. (2013). Decision making as a window on cognition. *Neuron*, 80(3), 791–806. <https://doi.org/10.1016/j.neuron.2013.10.047>
- Sievert, C. (2020). *Interactive web-based data visualization with R, plotly, and shiny*. CRC Press, Taylor; Francis Group. <https://doi.org/https://doi.org/10.1201/9780429447273>
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>
- Sottas, P.-E., Robinson, N., Rabin, O., & Saugy, M. (2011). The athlete biological passport. *Clinical Chemistry*, 57(7), 969–976. <https://doi.org/10.1373/clinchem.2011.162271>
- Steingroever, H., Fridberg, D. J., Horstmann, A., Kjome, K. L., Kumari, V., Lane, S. D., Maia, T. V., McClelland, J. L., Pachur, T., Premkumar, P., Stout, J. C., Wetzels, R., Wood, S., Worthy, D. A., & Wagenmakers, E.-J. (2015). Data from 617 healthy participants performing the Iowa gambling task: A “Many Labs” collaboration. *Journal of Open Psychology Data*, 3. <https://doi.org/10.5334/jopd.ak>
- Steingroever, H., Pachur, T., Šmíra, M., & Lee, M. D. (2018). Bayesian techniques for analyzing group differences in the Iowa gambling task: A case study of intuitive and deliberate decision-makers. *Psychonomic Bulletin & Review*, 25(3), 951–970. <https://doi.org/10.3758/s13423-017-1331-7>
- Steingroever, H., Wetzels, R., & Wagenmakers, E.-J. (2014). Absolute performance of reinforcement-learning models for the Iowa gambling task. *Decision*, 1(3), 161–183. <https://doi.org/10.1037/dec0000005>
- Stroud, J. R., Müller, P., & Rosner, G. L. (2001). Optimal sampling times in population pharmacokinetic studies. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(3), 345–359. <https://doi.org/10.1111/1467-9876.00239>
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (Second edition). The MIT Press.
- Sylvester, R. J. (1988). A Bayesian approach to the design of phase II clinical trials. *Biometrics*, 44(3), 823–836. <https://doi.org/10.2307/2531594>
- Team, R. C. (2023). R: A Language and Environment for Statistical Computing. <https://www.R-project.org/>
- Thomas, J. G., & Bond, D. S. (2015). Behavioral response to a just-in-time adaptive intervention (JITAI) to reduce sedentary behavior in obese adults: Implications for JITAI optimization. *Health Psychology*, 34(Suppl), 1261–1267. <https://doi.org/10.1037/hea0000304>

- Valentin, S., Kleinegesse, S., Bramley, N. R., Seriès, P., Gutmann, M. U., & Lucas, C. G. (2024). Designing optimal behavioral experiments using machine learning. *eLife*, *13*, e86224. <https://doi.org/10.7554/eLife.86224>
- Yang, J., Pitt, M. A., Ahn, W.-Y., & Myung, J. I. (2021). ADOPy: A Python package for adaptive design optimization. *Behavior Research Methods*, *53*(2), 874–897. <https://doi.org/10.3758/s13428-020-01386-4>
- Yousefi, A., Paulk, A. C., Basu, I., Mirsky, J. L., Dougherty, D. D., Eskandar, E. N., Eden, U. T., & Widge, A. S. (2019). COMPASS: An open-source, general-purpose software toolkit for computational psychiatry. *Frontiers in Neuroscience*, *12*, 957. <https://doi.org/10.3389/fnins.2018.00957>
- Zhang, S., & Lee, M. D. (2010). Optimal experimental design for a class of bandit problems. *Journal of Mathematical Psychology*, *54*(6), 499–508. <https://doi.org/10.1016/j.jmp.2010.08.002>
- Zimek, A., Schubert, E., & Kriegel, H.-P. (2012). A survey on unsupervised outlier detection in high-dimensional numerical data. *Statistical Analysis and Data Mining*, *5*(5), 363–387. <https://doi.org/10.1002/sam.11161>