



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
SCHOOL OF ENGINEERING
DEPARTAMENT OF COMPUTER SCIENCE

**EMPIRICAL BASIS FOR THE DEVELOPMENT OF
ADAPTIVE INTERFACES: BEHAVIORAL AND
NEUROPHYSIOLOGICAL EVIDENCES OF
DECISION-MAKING AND EXPERTISE
DEVELOPMENT IN A SEQUENTIAL CHOICE
SCENARIO**

CRISTÓBAL MATÍAS MOËNNE-LOCCOZ VARGAS

Thesis submitted to the Office of Research and Graduate Studies
in partial fulfillment of the requirements for the degree of
Doctor in Engineering Sciences

Advisor:

DOMINGO MERY

DIEGO COSMELLI

Santiago de Chile, October 2017

© MMXVII, CRISTÓBAL MATÍAS MOËNNE-LOCCOZ VARGAS



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
SCHOOL OF ENGINEERING
DEPARTAMENT OF COMPUTER SCIENCE

**EMPIRICAL BASIS FOR THE DEVELOPMENT OF
ADAPTIVE INTERFACES: BEHAVIORAL AND
NEUROPHYSIOLOGICAL EVIDENCES OF
DECISION-MAKING AND EXPERTISE
DEVELOPMENT IN A SEQUENTIAL CHOICE
SCENARIO**

CRISTÓBAL MATÍAS MOËNNE-LOCCOZ VARGAS

Members of the Committee:

DOMINGO MERY

DIEGO COSMELLI

VLADIMIR LÓPEZ

CRISTIAN TEJOS

MAX CHACÓN

JEAN-PHILIPPE LACHAUX

JORGE VÁSQUEZ

Thesis submitted to the Office of Research and Graduate Studies
in partial fulfillment of the requirements for the degree of
Doctor in Engineering Sciences

Santiago de Chile, October 2017

© MMXVII, CRISTÓBAL MATÍAS MOËNNE-LOCCOZ VARGAS

*Gratefully to my family
and friends*

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my advisors Professor Diego Cosmelli and Professor Domingo Mery for allowing me to make possible my Ph.D study and research: thanks for your thoughtful guidance. Also, I want to thank my Thesis Committee for all their helpful suggestions and time. I offer my sincere appreciation for the learning opportunities provided by my committee.

I am deeply grateful to Professor Vladimir López for helpful discussions and excellent comments, but also for listening to me on any occasion.

I would like to thank my fellow labmates: Rodrigo Vergara, Constanza Baquedano, Gonzalo Boncompte, and Catherine Andreu for being always around: thanks for all the chitchatting and procrastination time we spend together. Specially, I am grateful to Rodrigo Vergara for being a friend and partner throughout all this process.

Most importantly, none of this would have been possible without the love and patience of my family: my grandparents René Moenne and Melita Geister, my parents Juan Carlos Moenne and Silvia Vargas, my aunt Edith Moenne, and my sister Isabel Moenne and her husband Fabián Sepúlveda. Thank you for their faith in me and for supporting me spiritually throughout my studies.

Last but not least, to my caring, loving and supportive soul mate, Cristina Jara: my deepest gratitude, my heartfelt thanks.

This work was supported by the Chilean National Council of Scientific and Technological Research (CONICYT) National PhD grant number 21110823.

LIST OF PAPERS

Moënné–Loccoz, C., Vergara C. R., López, V., Mery, D., & Cosmelli, D. Modeling search behaviors during the acquisition of expertise in a sequential decision-making task. *Frontiers in Computational Neuroscience* (2017).

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iv
LIST OF PAPERS	v
LIST OF FIGURES	viii
LIST OF TABLES	x
LIST OF SYMBOLS	xii
ABSTRACT	xv
RESUMEN	xvii
1. Outline	1
1.1. Motivation	1
1.2. Approach	2
2. Introduction	4
2.1. Electrophysiology	9
3. Hypothesis	13
4. Objectives	14
5. Methods	15
5.1. Task	15
5.2. Participants	18
5.3. Modeling Framework	19
5.3.1. Low-level Actions vs Strategies	19
5.3.2. Behavioral Modeling	24
5.3.3. Scores	31
5.3.4. Clustering	32

5.4. Recordings	33
5.4.1. Mouse-Tracking Recordings	33
5.4.2. Electrophysiological Recordings	35
6. Results	39
6.1. Behavioral Results	39
6.2. Model Parameter Dependence	40
6.3. Individual Differences	42
6.4. Predicting Participants' Choices	46
6.5. Mouse-Tracking	49
6.6. Electrophysiological Results	53
6.6.1. Checkerboard Probes	53
6.6.2. Binary Choice ERP Components	56
6.6.3. Feedback ERP Components	65
7. Discussion	69
7.1. Modeling Framework	69
7.2. Task Structure	72
7.3. Temporal and Spatial Encoding Criteria	73
7.4. Electrophysiological Analysis	75
7.5. Individual Differences among Users	76
8. Conclusions and Future Outlook	78
References	79

LIST OF FIGURES

5.1	Schematic presentation of the BDT showing one possible instance of the task	16
5.2	Graphical representation of a Hidden Markov Model	25
5.3	Electrodes connectivity template	37
6.1	Reaction Times	39
6.2	Framework parameters modulation	41
6.3	Learners and non-learners average strategies across tasks	43
6.4	Study Cases	45
6.5	Participants performance, framework prediction, and 50/50 guesses averaged across tasks	47
6.6	Mouse-Tracking averaged data before and after clicking an icon separated by the most likely observation predicted by the model for learners and non-learners groups	49
6.7	Mouse-Tracking averaged data before and after clicking an icon separated by the action predicted by the model for learners and non-learners groups	50
6.8	Checkerboard probes ERP components at selected ROIs for learners group	54
6.9	Checkerboard probes ERP components at selected ROIs for non-learners group	55
6.10	Learners binary choice related ERPs separated by the most likely observation predicted by the model	59
6.11	Non-learners binary choice related ERPs separated by the most likely observation predicted by the model	60

6.12	Learners binary choice related ERPs separated by the most likely observation predicted by the model	63
6.13	Non-learners binary choice related ERPs separated by the most likely observation predicted by the model	64
6.14	Learners feedback related ERPs at selected ROIs and its corresponding scalp maps	67
6.15	Non-learners feedback related ERPs at selected ROIs and its corresponding scalp maps	68

LIST OF TABLES

5.1	Icon-concept mapping	17
5.2	Formal definitions of the strategies	22
6.1	Statistics for mouse movements clustered by the most likely observation predicted by the behavioral modeling	51
6.2	Statistics for mouse movements clustered by behavioral modeling prediction .	52
6.3	Summary of significant spatiotemporal statistics for binary choice EEG signals separated by observation prediction	58
6.4	Summary of significant spatiotemporal statistics for feedback EEG signals . .	66

LIST OF SYMBOLS

- t :
Time step
- s_i :
Strategy model
- α :
Learning rate. With subscript (α_i) represents the learning rate of strategy s_i
- v :
Specific observation
- $P(v|s_i)$:
Emission probability of observation v under strategy s_i
- $P(s_i \rightarrow s_j)$:
Probability of transiting between strategies s_i and s_j
- $w_i(t)$:
Weight of strategy s_i at time step t
- Q :
Sequential decision path of choices
- Q_t :
Selected path at time step t
- Q^* :
Target path of a task instance (leads to positive feedback)
- l_{ij}^t :
Transition links votes at time step t between strategies s_i and s_j

- $L_{ij}(t)$: Cumulative votes until time step t between strategies s_i and s_j
- τ : Temporal observation window of the process. With subscript (τ_v) represents the temporal relevance of observation v
- p : Place of the Binary Decision Tree (BDT)
- c : Concept of the Binary Decision Tree (BDT)
- a : Choice/action
- $h_i(t)$: Expert prediction of the strategy s_i at time step t . Predicted actions are defined as $a \in \{\text{left, right}\}$
- $|\cdot|$: Number of elements of a given set (operator)
- M : Set of strategy models. $M = \{\text{random, topological, generative, discriminative}\}$
- V : Set of observations. $V = \{\text{mistake, explore, hit}\}$
- $D_i(t)$: Domain of action of strategy s_i at time step t
- $R_i(t)$: Set of learned choices of strategy s_i at time step t
- $G_i(t)$: Set of all choices of strategy s_i at time step t

$H_i(t)$:

Set of all paths that do not violate the strategy's s_i rules at time step t , while simultaneously allowing the exploration of available choices (does not include target paths Q^*)

$H_i^*(t)$:

Set of paths that lead to positive feedback at time step t for strategy s_i (includes only target paths Q^*)

$B_v(t)$:

Set of best explanatory strategies of v (explained through emission probabilities) at time step t

$U(t)$:

Set of strategies that can leave the random strategy (those that do not see the current action as a mistake), thus $l_{random j}^t > 0$ at time step t , plus the random strategy

ABSTRACT

Our daily interaction with computer interfaces is plagued of situations in which we go from inexperienced to experienced users through self-motivated repetition of the same task. In many of these interactions, we must learn to find our way through a sequence of decisions and actions before obtaining the desired result. For instance, when drawing cash from an ATM machine, choices are presented in a set-by-step fashion and a specific sequence of actions must be performed in order to produce the expected outcome. But, as we become experts in the use of such interfaces, it is possible to identify specific search and learning strategies? And if so, can we use this information to predict future actions? In addition to better understanding the cognitive processes underlying sequential decision making, this could allow building adaptive interfaces that can facilitate interaction at different moments of the learning curve. Here we tackle the question of modeling sequential decision-making behavior in a simple human-computer interface that instantiates a 4-level binary decision tree (BDT) task. We record behavioral data from voluntary participants while they attempt to solve the task. Using a Hidden Markov Model-based approach that capitalizes on the hierarchical structure of behavior, we then model their performance during the interaction. Our results show that partitioning the problem space into a small set of hierarchically related stereotyped strategies can potentially capture a host of individual search behaviors. This allows us to follow how participants learn and develop expertise in the use of the interface. Moreover, using a Mixture of Experts based on these stereotyped strategies, the model is able to predict the behavior of participants that master the task. Furthermore, using behavioral indicators derived from our behavioral model, we are able to capture the rich structure of the learning process and expertise development in the participants' Electroencephalogram (EEG) recordings, revealing at brain level the different stages of the decision-making process through Event Related Potentials (ERP). Our long-term goal is to inform the construction of interfaces that can establish dynamic conversations with their users in order to facilitate ongoing interactions.

Keywords: Sequential Decision-Making, Hidden Markov Models, Expertise Acquisition, Behavioral Modeling, Search Strategies, Binary Decision Tree, Event Related Potentials.

RESUMEN

En el día a día nuestra interacción con interfaces computacionales está llena de situaciones en las cuales pasamos de ser usuarios inexpertos a expertos mediante la repetición de una misma tarea. En muchas de estas interacciones debemos aprender a encontrar una ruta, dentro de una secuencia de decisiones y acciones, la cual nos lleva al resultado buscado. Por ejemplo, cuando retiramos dinero de un cajero automático, las elecciones son presentadas paso a paso y una secuencia específica de acciones debe ser realizada en orden de obtener el resultado deseado. Entonces, a medida que nos hacemos expertos en el uso de estas interfaces, ¿es posible identificar estrategias específicas de búsqueda y aprendizaje? De ser así, ¿podemos usar esa información para predecir acciones futuras? Además de comprender mejor los procesos cognitivos que subyacen a la toma de decisiones secuencial, esto podría permitir construir interfaces adaptativas que puedan facilitar la interacción en diferentes momentos de la curva de aprendizaje. Aquí abordamos la pregunta de modelar el comportamiento de toma de decisiones secuencial usando una interfaz visual simple representada por un árbol de decisión binario (por sus siglas en inglés BDT) de cuatro niveles. Registramos datos conductuales de participantes voluntarios mientras tratan de resolver la tarea. Utilizando un enfoque basado en el modelo oculto de Markov, que se capitaliza la estructura jerárquica del comportamiento, luego modelamos el desempeño de los participantes durante la interacción. Nuestros resultados muestran que una partición del espacio del problema en un pequeño grupo de estrategias estereotipadas y relacionadas jerárquicamente pueden capturar potencialmente una serie de comportamientos de búsqueda. Esto nos permite seguir cómo los participantes aprenden y desarrollan habilidades en el uso de la interfaz. Más aun, usando una Mezcla de Expertos basadas en las estrategias, somos capaces de predecir el comportamiento de los participantes que aprenden la tarea. Además, usando indicadores conductuales derivados de nuestro modelamiento, somos capaces de capturar la compleja estructura de los procesos de aprendizaje y desarrollo de expertise presente en los registros de Electroencefalograma

(EEG) de los participantes, revelando a nivel cerebral las diferentes etapas del proceso de toma de decisión a través de Potenciales Relacionados a Eventos (por sus siglas en inglés ERP). Nuestra meta a largo plazo es informar acerca de la construcción de interfaces que puedan establecer una conversación dinámica con sus usuarios, en orden de facilitar la interacción con ellas.

Palabras Claves: Toma de Decisiones Secuencial, Modelo Oculto de Markov, Desarrollo de Expertise, Modelamiento Conductual, Estrategias de Búsqueda, Árbol de Decisión Binario, Potenciales Relacionados a Eventos.

1. OUTLINE

1.1. Motivation

Countless of our daily interactions with computer interfaces are based on repetitive sequences of actions. For instance, think about what you do to draw cash from an ATM machine. You will probably insert or swipe your card, input your pin code and then go through a series of button or screen presses in order to get the money. In most cases, different options offered to you by the machine will maintain a consistent spatial distribution for the same bank. Sometimes such distribution is even shared between banks. This obviously facilitates the task because it allows you to learn and exercise a sequence of actions for common requests. Indeed, when one is accustomed to using a given machine to withdraw a standard amount, it is not rare to quickly go through the sequence of button presses and successive screens, oblivious of the other available alternatives; one might even be able to reproduce such sequence from memory.

The scenario just described is a common situation where decision making and learning are intertwined. From the simple ATM case described above to looking up some specific information in a web site for the first time, we go from inexperienced to experienced users though self-motivated repetitions of the same task. However, as common as we may think these situations are, there are few studies of how behavior and decision-making processes operate in these kind of interactions, how we can quantify the level of expertise of a given user while she or he explores and learn the “ways” of the interface, or the way in which we modulate our attention as we learn the interface structure.

In this project, we want to address these questions using a binary decision tree task to represent a simplified version of the ATM situation. While identifying the decision-making process with a small set of stereotyped search and learning strategies, this simple interface will allow us to dynamically follow the behavior of participants as it unfolds in the context of repetitive sequential choices. It is worth noting that these types of tasks remain severely understudied in cognitive neuroscience, despite being representative of

most of human-machine interactions. The development of computational models that can deal with such scenarios can therefore have a significant impact in understanding how real-world decision making processes are made, and the cognitive mechanisms that underlie such decision making behavior.

Given the relevance and potential applications of adaptability in a simplified user-interface experience and the wide range of interactions mediated by sequential choice interfaces in modern society, we aim at yielding insight into how decisions are undertaken by a specific user, providing a knowledge-base to establish more ecological conversations between users and interfaces. We also expect to bring closer the study of decision-making to more ecological experimental formats, such as the sequential choice paradigm with limited feedback. Our long-term goal is contributing to a base of empirical knowledge to develop adaptive human-machine interfaces, which considers individual strategies as well as expertise acquisition over time.

1.2. Approach

To tackle a sequential decision-making scenario that represents a plausible model of standard human-computer interactions, participants are presented with a decision-making search task structured as a four-level Binary Decision Tree (BDT). In this task, participants use a computer mouse to explore and navigate through a series of binary choices. This task represents a sequential choice with limited feedback scenario, where participants are asked to discover an underlying concept-icon mapping, receiving feedback only at the end of the sequence of choices. This simple human-computer interaction makes learning the structure of the interface challenging enough to see a variety of individual behaviors as participants explore and learn the task.

To deal with the problem of modeling behavior during the process of learning of the BDT, we formalize a small set of stereotyped hierarchically related search strategies which

allows to solve the concept-icon mapping. Based on a Hidden Markov Model (HMM)-based, we build a framework that can track the online deployment of the search strategies as participants explore and learn the structure of the BDT. This allows us to follow how participants develop expertise in the use of the interface, obtaining context-driven learning curves of each strategy and the approximate learning curve of the learning process. In addition, using a Mixture of Experts based on the strategies, the model can predict the most likely next choice of the participant based on its previous behavior.

Taken together, these aspects support the idea that our framework could serve as a novel basis for the development of computational interfaces that dynamically react to the behavior of individual users. As such, our objective is to experimentally investigate the individual differences in the decision-making process, expertise acquisition, and the attentional dynamics of participants while exploring the interface. Using behavioral indicators and brain-related ones we want to characterize the different stages of learning and expertise development captured by the framework.

2. INTRODUCTION

Whether you are preparing breakfast or choosing a web link to click on, decision making processes in daily life usually involve sequences of actions that are highly dependent on prior experience. Consider what happens when you interact with an ATM machine: you have to go through a series of specific button presses (i.e. actions) that depend on whether you are interested in, for instance, drawing money or consulting your account balance (i.e. the outcome). Despite some commonalities, which sequence you use will depend on the specific ATM brand you are dealing with, while previous exposure will determine which behavioral strategy you deploy. Maybe you cautiously explore the available choices and hesitate before pressing each button; maybe this is the same machine you have used for the last year, so you deftly execute a well practiced sequence of actions to draw some cash.

Sequential choice situations such as the ATM example are pervasive in everyday behavior. Not surprisingly, its importance for the understanding of human decision making in real-world scenarios has been recognized for long time, as a wealth of studies attest (Rabinovich, Huerta, & Afraimovich, 2006; Gershman, Markman, & Otto, 2014; Cushman & Morris, 2015; Otto, Skatova, Madlon-Kay, & Daw, 2014); see (M. M. Walsh & Anderson, 2014) for review). Yet, despite its importance, the issue of how expertise is developed in the context of sequential choice situations remains still under-explored. While the study of optimal strategies or courses of action to solve sequential choice scenarios is a fundamental aim of such studies (Alagoz, Hsu, Schaefer, & Roberts, 2009; Friedel et al., 2015; Schulte, Tsiatis, Laber, & Davidian, 2014; Sepahvand, Stöttinger, Danckert, & Anderson, 2014), it is still necessary to better understand how agents learn and acquire such strategies as they interact with the world (W.-T. T. Fu & Anderson, 2006; Acuña & Schrater, 2010; Sims, Neth, Jacobs, & Gray, 2013).

Human-computer interfaces offer a privileged scenario to study the development of expertise in sequential decision making processes. As we learn to use an interface, isolated exploratory actions turn into expert goal-directed sequences of actions by repeatedly

testing and learning to adjust behavior according to the outcome (Solway & Botvinick, 2012). Furthermore, such scenarios are particularly well adapted to sampling behavioral data in varying degrees of ecological validity. They also allow for testing different ways to use such behavioral data to make the interface more responsive. Indeed, in addition to contributing to a better understanding of the psychological and neurobiological mechanisms underlying sequential decision making, taking into account how people learn to interact with novel systems could have relevant consequences for adaptive interface design. Here we tackle the question of expertise acquisition in a sequential decision making scenario. We aim to discover if individuals can be described in terms of specific behavioral strategies while they learn to solve the task and, if so, whether this can be used to predict future choices.

Several approaches have been developed to address the problem of modeling behavior in sequential decision making scenarios. The Reinforcement Learning (RL) paradigm has been successfully extended to model behavior during sequential choice tasks (M. M. Walsh & Anderson, 2014; Dezfouli & Balleine, 2013; Dayan & Niv, 2008; Acuña & Schrater, 2010; Daw, 2013). In general terms, RL techniques aim at finding a set of rules that represent an agent's policy of action given a current state and a future goal by maximizing cumulative reward. Because actions are chosen in order to maximize reward, it is necessary to assign value to the agent's actions. Reward schemes work well when gains or losses can be estimated (e.g. monetary reward). However, in many of our everyday interactions, reward in such an absolute sense is difficult to quantify. Accordingly, the accuracy of an arbitrary reward function could range from perfect guidance to totally misleading (Laud, 2004).

To overcome the difficulty of defining a reward function, the Inverse Reinforcement Learning based approaches try to recover a reward function from execution traces of an agent (Abbeel & Ng, 2004). Interestingly, this technique has been used to infer human goals (Baker, Saxe, & Tenenbaum, 2009), developing into methods based on Theory of Mind to infer peoples' behavior through Partially Observable Markov Decision Process

(POMDP) (Baker, Saxe, & Tenenbaum, 2011). Although these techniques are important steps in the area of plan recognition, they usually focus on which is the best action an agent can take given a current state, rather than determine high-level patterns of behavior. Also, they consider rational agents who think optimally. However, in learning scenarios, optimal thinking is achieved through trial and error, rather than being the de facto policy of action.

Alternative modeling approaches exist that do not require defining a reward function to determine behaviors. Specifically, Markov models have been adapted to analyze patterns of behavior in computer interfaces, such as in web page navigational processes (Singer, Helic, Taraghi, & Strohmaier, 2014; Ghezzi, Pezzè, Sama, & Tamburrelli, 2014), where no simple, unitary reward is identifiable. In general terms, Markov models are aimed at modeling the stochastic dynamics of a system which undergoes transitions from one state to another, assuming that the future state of the system depends only on the current state (Markov property). For example, a Markov chain of navigational process can be modeled by states representing content pages and state transitions representing the probability of going from one page to another (e.g. going from the login page to the mail contacts or to the inbox). Possible behaviors or cases of use can then be extrapolated from the structure of the model (see (Singer et al., 2014) for an example). In general, however, these behaviors are highly simplified descriptions of the decision-making process because they do not consider the rationale behind the user's actions, and only focus on whether the behavioral pattern is frequent or not. Accordingly, if the user scrolls down the page searching for a specific item in some order and then makes a decision, the psychological processes behind his or her actions are ignored.

An interesting extension of the simpler Markov models, which aims to capture the processes underlying decision making behavior, are the Hidden Markov Models (HMM) (Rabiner, 1989). In a HMM, the states are only partially observable and the nature of the underlying process is inferred only through its outcomes. This relationship between states and outcomes allows modeling a diversity of problems, including the characterization of

psychological and behavioral data (Visser, Raijmakers, & Molenaar, 2002; Duffin, Bland, Schaefer, & De Kamps, 2014). Of special interest is the use of HMMs to model the strategic use of a computer game interface (Mariano et al., 2015). Using sets of HMMs, Mariano and collaborators analyze software activity logs in order to extrapolate different heuristics used by subjects while they discover the game's rules. Such heuristics, which are extrapolated a posteriori, are represented by hidden states composed by the grouping of actions and the time taken to trigger such actions. These are then used to identify patterns of exploratory behavior and behaviors representative of the mastery of the game. Interestingly, such heuristics show a good adjustment with self-reported strategies used by the participants throughout the task (Mariano et al., 2015).

Generally speaking, however, Markov models use individual actions to represent hidden states (such as click this or that icon) and more complex high-level behavioral heuristics (such as search strategies) are only inferred after the experimental situation or setup (see (Mariano et al., 2015)). This represents a potential limitation if we are to build interfaces that are responsive to the learning process as it unfolds. Indeed, throughout the acquisition of different skills, there is abundant evidence that humans and other animals rely on grouping individual actions into more complex behavioral strategies such as action programs or modules to achieve a certain goal (Marken, 1986; Manoel, Basso, Correa, & Tani, 2002; Rosenbaum, Cohen, Jax, Weiss, & Van Der Wel, 2007; Matsuzaka, Picard, & Strick, 2007). This can happen while individual actions remain essentially unchanged and only the way they are organized changes. For instance, it has been shown that throughout the development of sensorimotor coordination, individual movements progressively become grouped into sequences of movements, as the child becomes adept at controlling goal-directed actions (Bruner, 1973; Fischer, 1980).

If expertise development and skill acquisition implies the hierarchical organization of individual actions into complex, high-level behavioral strategies, taking this into account could represent a relevant line of development for behavioral modeling. Recently, this

hierarchical structure of behavior has been considered in some approaches, such as Hierarchical Reinforcement Learning (HRL) (M. M. Botvinick, 2012; M. Botvinick & Weinstein, 2014). The idea behind HRLs is to expand the set of actions available to an agent to include a set of extended high-level subroutines which coordinate multiple low-level simple actions that otherwise the agent would have to execute individually. An interesting consequence of this is that it could potentially aid in reducing the dimensionality of the problem, a critical issue in behavioral modeling (Doya & Samejima, 2002; Shteingart & Loewenstein, 2014; M. M. Botvinick, 2012). Furthermore, if flexibility in membership between low-level actions and behaviors exist, modeling behaviors in a probabilistic way could help in better capturing the dynamical nature of expertise acquisition.

Here we present here a HMM-based approach that capitalizes on the hierarchical structure of behavior to model the performance of individuals as they develop expertise in a sequential decision making task that is structured as a 4-level Binary Decision Tree (BDT). We use the HMM structure to infer the distribution of probabilities of a modular set of pre-defined stereotyped high-level search strategies (represented by hidden states), while observing the outcome of the user's actions. As such, this approach is reminiscent of type-based methods where one focuses on a pre-specified group of behaviors among all possible behaviors (Albrecht, Crandall, & Ramamoorthy, 2016). Such "types" of behavior can be hypothesized based on previous knowledge of the interaction or on the structure of the problem. In our case, pre-defined strategies are selected in order to cover a series of increasingly efficient behaviors in the context of the BDT: from random unstructured exploration to goal-directed actions driven by feedback and knowledge of the task's structure. Additionally, this approach allows us to use the modular architecture as a Mixture of Experts (see (Doya & Samejima, 2002) for a similar idea). This enables us to model the evolution of the user's behavior while simultaneously asking the experts about the user's most probable next choice. As a consequence, it is possible to evaluate the model also in terms of its capacity to predict future actions, which would be desirable in the context of adaptive interface design.

We test this approach using a simple computational interface game where an underlying concept-icon mapping must be discovered in order to complete the task. As mentioned above, the game is structured as a four-level BDT with limited feedback (see Fig. 5.1; see also (W.-T. T. Fu & Anderson, 2006), p198 for a structurally similar design). Each node of the tree represents a decision, and the links between nodes represent the consequences of each decision. The depth of the tree represents the number of sequential choices that are needed to achieve a specific goal. This scenario captures sequential decision-making situations with limited feedback that are pervasive in real world human-computer interface interaction. Furthermore, it allows us to follow the development of expertise as the users discover the rules of the game and deploy different search strategies to solve the task.

2.1. Electrophysiology

One of the major substrates of the neurophysiological studies is the electrical activity of the brain. This activity can be sampled using an electrode array, which is placed non-invasively on the scalp. Noninvasive registration is known as Electroencephalogram (EEG), and mainly reflects the sum of postsynaptic potentials of cortical neurons. Stereotyped electrophysiological responses to a stimulus are called Event Related Potentials (ERPs) (S. J. Luck, 2014). ERPs are measured by averaging many trials of the EEG recording to retain the relevant brain's response to the stimulus while averaging out the random brain activity.

As the participant explores the interface, some ERPs are modulated by learning or attentional processes. In the BDT task, these modulations are mainly evoked at each sequential choice, and can be captured at the onset of stimuli such as responses (mouse clicks), feedback presentation, and attentional probes (checkerboard patterns). Checkerboard probes are external stimuli added to the task and are used to sample the early visuospatial attentional components as the P1 and N1 ERPs. This kind of attention is related to the orienting network, which is focused in the prioritization of sensory input by selecting a location (Petersen & Posner, 2012). P1 peak occurs around $\sim 70 - 90$ milliseconds

after the onset of the visual stimuli, and N1 peak occurs around $\sim 120 - 150$ milliseconds respectively. Both components are localized over the lateral occipital scalp. The presentation paradigm used in this work (see Methods section) is similar to the Posner paradigm (Posner, 1980; Posner, Snyder, & Davidson, 1980), where cues are presented lateralized to the left or right positions. Then, the comparison between these ERPs is carried out by comparing the same stimulus when attended or not. Without the behavioral modeling, checkerboard probes comes close to a classical alternative to sample the attentional state of the participants. However, the lack of clear attentional assignments, i.e. what and when an element is or is not attended, makes using this technique difficult in some cases.

Mouse responses, which also trigger the presentation of feedback in the last level of the BDT, are also related to visual stimuli. By changing the icons of the binary choice or by presenting feedback of the actions of the participant, a P2 appears. The P2 has a peak located around the centro-frontal area which occurs about 200 ms after the onset of the visual stimuli. Although the characterization of this component is unclear, in the context of visual search tasks its changes could be interpreted as facilitation of attentional deployment toward target locations (Anllo-Vento & Hillyard, 1996; Akyurek & Schubo, 2013); improved task performance (Ross & Tremblay, 2009); and outcome predictability (Polezzi, Lotto, Daum, Sartori, & Rumiati, 2008), among others. Besides the brain processes previously mentioned, this component has been also reported to be modified by feedback (Schuermann, Endrass, & Kathmann, 2012). Thus, it could behave different in the presence of explicit/implicit feedback.

Besides the visual ERP components, there are other ERPs related to the decision-making process that can be evoked by the mouse response. The Error-Related Negativity (ERN) is a component associated with error processing and occurs within 100 ms of an erroneous response (Gehring, Goss, Coles, Meyer, & Donchin, 1993; Larson, Clayson, & Clawson, 2014). This component is more prominent in frontal and central electrode sites. These errors may be conscious (aware errors) or unconscious (unaware errors), but even

in correct responses a small negativity could be detected (Shalgi & Deouell, 2012). In the context of BDT choices, the ERN can be understood as the expectations in the outcome of the current decision, which initiates behavioral adjustments in order to optimize the choice selection process (Maier, Yeung, & Steinhauser, 2011). In addition to the ERN and temporally later, other important components of the decision-making process are the P3 and the Late Positive Complex (LPC). P3 component is measured more strongly in the parietal scalp sites, is associated with the evaluation or categorization of a stimulus and has its peak around $\sim 300 - 500$ ms after stimuli onset. This component has also implications in cognitive workload (Donchin, 1981), and is modulated by the delivery of task relevant information (Polich, 2007). Perhaps the most relevant aspect of the P3 is its widespread use in BCI, mainly due to its robustness (Ahn, Lee, Choi, & Jun, 2014). In the presence of explicit feedback, a fronto-central P3 appears. The feedback-related P3 is suggested to be modulated by information which is motivationally significant or salient (Schuermann et al., 2012). Furthermore, it is also related to the feedback-guided learning by the context updating hypothesis (revisions of the mental model of the task) (San Martín, 2012; Donchin & Coles, 1988). Following the P3, the Late Positive Complex is a positive-going ERP component in the temporal window of $\sim 500 - 800$ ms after stimulus onset, and it is largest in parietal scalp sites. This component is associated with recognition memory (Friedman & Johnson, 2000; Wolk et al., 2006), which is usually divided into two processes: familiarity and recollection. Familiarity is an early process, about $\sim 300 - 500$ ms, thus it is modulated in the P3 effect. On the other hand, recollection modulates the LPC, such as in correct source or associative recognition (Curran, Schacter, Johnson, & Spinks, 2001), differences in the remember/known paradigms (Spencer, Abad, & Donchin, 2000), among others. Also, a more parietal LPC topology is associated to post-retrieval processing, as in the strategic processing related to the required task (Ranganath & Paller, 2000).

In addition to the previously mentioned ERPs, feedback stimuli are also characterized by the Feedback-Related Negativity (FRN). The FRN is a centro-frontal ERP component associated with the onset of a feedback signaling the outcome of a decision (Miltner,

Braun, & Coles, 1997; Sallet, Camille, & Procyk, 2013). FRN effects are modulated in decision-making tasks by risk (Schuermann et al., 2012), loss and gain (Liu, Nelson, Bernat, & Gehring, 2014), among other feedback-related properties. In the BDT task, this component is likely to be associated with feedback-guided learning (San Martín, 2012; M. M. Walsh & Anderson, 2012), which compares the difference between the obtained and expected value to trigger changes in behavior.

By capturing the evolution of the decision-making process over time, the above-mentioned ERPs could reveal how the development of expertise is related (or not) to the attentional resources used by the participant to learn the task. We expect to show these differences when comparing the different stages of the learning process, such as those produced by the formalization of stereotyped search strategies.

3. HYPOTHESIS

We hypothesize that the interaction between a user and an the interface can be described by stereotyped strategies, characterized by a set of predefined actions/decisions and the distinctive use of attentional resources. The level of expertise developed by the user, while exploring the interface, would correlate with changes in the use of such strategies, prompting different curves of performance.

Specifically, the dynamics of behavioral and electrophysiological indicators will correlate with distinct patterns over time:

- a) Behaviorally, as the user explores the interface and learns the underlying mapping between concepts and icons, his or her search process will be characterized by the use of different strategies. Moreover, the development of expertise in the use of the interface will correlate with a decrease in mouse movements variability, likewise, the reaction times and errors will be minimized throughout the course of the experiment.
- b) Neurophysiologically, the Event Related Potential (ERP) components' amplitude will be modulated in dependence with the attentional resources allocated by the user, correlating with the user's individual strategies and the development of expertise.

4. OBJECTIVES

The main objective of this proposal is to experimentally investigate the individual differences in the decision-making process, expertise acquisition, and the attentional dynamics of participants while exploring a novel interface. Using behavioral indicators (participant's choices, reaction times, mouse tracking, etc) and brain-related ones (event related potentials) we want to characterize the different stages of learning and expertise development. Specifically, we are interested in:

- a) Studying the deployment of putative search strategies and the development of expertise as a function of task acquaintance.
- b) Capturing the individual differences of participants in the use of the strategies throughout the task.
- c) Modeling the decision-making process of participants online (behavioral modeling).
- d) Predicting the participants' choices based on the online modeling.
- e) Studying the electrophysiological changes in relation to the behavioral modeling.

5. METHODS

5.1. Task

In order to represent a Human-Computer Interaction (HCI) scenario that requires learning, we implement a sequential decision making task that participants have to solve through active exploration. The task consists in presenting an instruction to participants and offering them a series of successive binary choices among abstract icons that, depending on the sequence of choices, can lead to either positive or negative feedback (see Figure 5.1 and below for details). In abstract terms, the task is a 4-level BDT where each node of the tree represents a binary decision and the links between nodes represent the consequences of each decision. The depth of the tree represents the number of available sequential choices. This scenario captures sequential decision-making situations with limited feedback that are pervasive in real world human-computer interface interaction (e.g. when drawing cash from an ATM machine).

Specifically, participants were presented with a computer screen that had an instruction of the type “*verb_{L1} with a noun_{L2} an adjective_{L3} noun_{L4}*” (i.e. “*Feed a cat with a small fish*”, see inset in Fig. 5.1 and Table 5.1 for more details). Note that because the task was performed by native Spanish speakers, the the final adjective-noun pair is inverted regarding the previous instruction example and would read “*Alimenta un gato con un pez pequeño*”. After clicking anywhere on this first screen, participants were confronted with the first binary choice (*level 1* of the BDT) with two icons located 2.4° to each side of a central fixation spot. Each of the variable words in the instruction had a fixed mapping to an abstract icon and level of the BDT (as indexed by subscripts in the instruction example above), but participants were not informed of this fact. They were instructed to click using the computer mouse on one of the two icons to proceed to the next screen where the next set of two icons was presented (*level 2* of the BDT). The overall arrangement of the icons remained constant and only their identity changed according to the specific task mapping. This branching structure was repeated until level 4 was reached. The last level was always

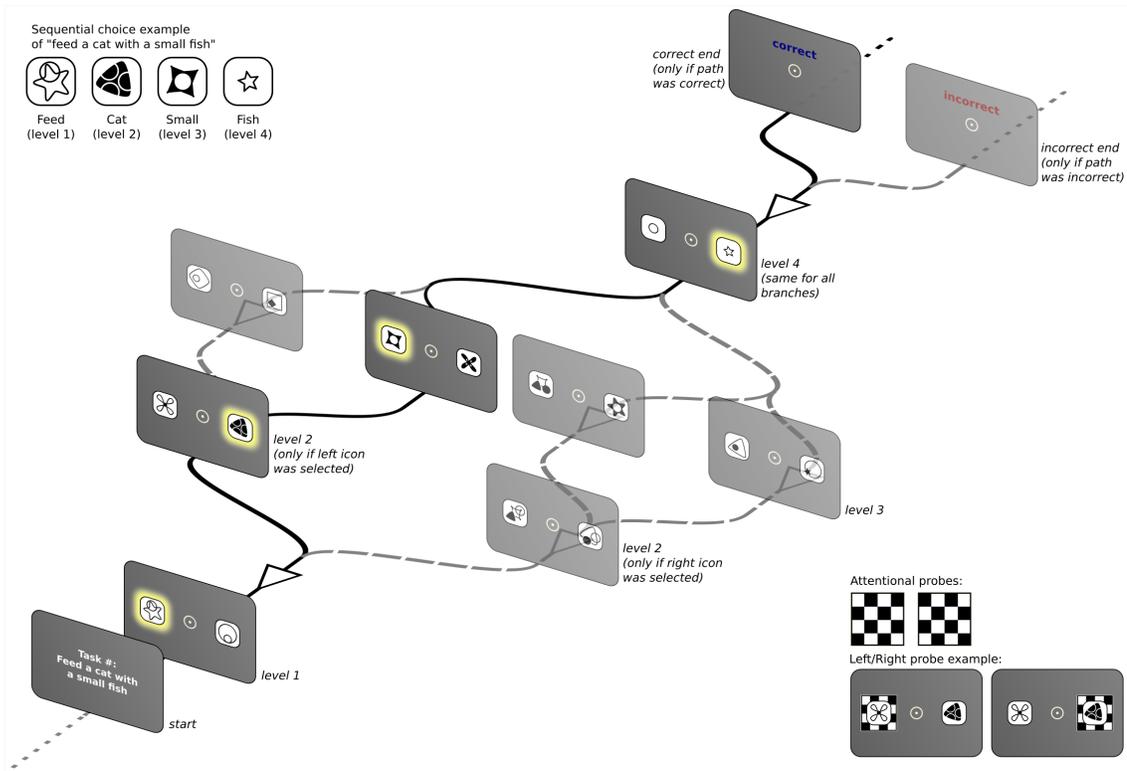


Figure 5.1. **Schematic presentation of the BDT showing one possible instance of the task.** Only one screen per level is presented, depending on the icon clicked previously. Highlighted icons along the black continuous lines represent the correct icon-to-concept mapping (see left inset) that, when clicked in the correct sequence, produces a positive feedback. Right inset presents the attentional probes used to sample the attentional state of the participants.

the same regardless of the branch because only two possible adjectives were used: “small (pequeño)” or “large (grande)”. Once subjects clicked an icon in level 4 they received feedback informing them whether the path they had chosen was correct. In the case of a negative feedback, subjects had no way of knowing, a priori, at which level they had made the wrong choice. Therefore, they had to discover the mapping based exclusively on their exposure to successive iterations of the task and the feedback received at the end of each chosen path. Importantly, as we chose a set of instructions in which the last binary choice (L4) was always the same, there are a total of 16 different instructions in the BDT. Each instruction (corresponding to a single task instance) was presented repeatedly until

level	icon	word	icon	word
level 1 (verb)		Capturar (con una)		Liberar (en)
level 2 (noun)		Red		Trampa
		Domo		Campo
level 3 (noun)		Libélula		Mariposa
		Escarabajo		Hormiga
		Alacrán		Cucaracha
		Luciérnaga		Abeja
level 4 (adjective)		Grande		Chico

Table 5.1. **Icon-concept relationships used in the task.** Each row represents a binary choice screen and its related icon-concept mapping. Note that not all binary choices are reachable from one particular icon (except for the fourth level), thus the first icon can only reach the first row (binary choice) of the second level, and the first and second rows of the third level, and so forth. There are 16 possible instances of the structure “(verb_{L1}) with a (noun_{L2}) an (adjective_{L3}) (noun_{L4})”. Examples of sentences are: “Capturar con una red una libélula grande”, “Liberar en el domo un alacrán chico”, “Capturar con una trampa una mariposa chica”, etc.

the participant was able to choose the correct path 5 times in a row before moving onto the next possible instruction. After 15 successive wrong answers, participants were asked whether they wanted to move onto the next task. If they refused, after 10 further wrong answers they were asked again. If they decided to persist in the same task, they only had 5 additional chances to find the correct path else they were forced to move onto the next one. The maximum time allowed to complete the entire task was set to 40 minutes.

In addition, to be able to sample the attentional state of participants in the electrophysiological record, checkerboard patterns (see Fig. 5.1 right inset) were added as attentional probes across trials. These probes were presented non-simultaneously and centered to the

left/right icons. Each probe has a duration of 75 milliseconds, and the temporal separation between them was randomized in the 400–600 milliseconds interval.

5.2. Participants

Twenty-two participants were recruited (12 females) of ages ranging from 20 to 32 years old, with a mean age of 26 ± 3 years (mean \pm SD). All participants reported normal or corrected-to-normal vision and no background of neurologic or psychiatric conditions. Participants performed on average 162 ± 51 trials (mean \pm SD) during the course of the experiment (range 89-325). Trials are defined as repetitions of an instance of the task, from the instruction to the feedback after the sequence of choices. Each task instance contains on average 10.79 ± 1.47 trials (mean \pm SD). In addition, two groups of participants were defined according to whether they managed to solve the task (learners, $N=14$) or not (non-learners, $N=8$) in the allowed time window (40 min). We consider learners all participants who managed to complete the 16 tasks instances, reaching full knowledge of the interface as evidenced by consistent positive feedback. Conversely, the non-learners group is composed by participants who did not complete the task within the time limit, or failed to reach full knowledge of the interface.

The study was approved by the Ethics Committee of the School of Psychology of the Faculty of Social Sciences, Pontificia Universidad Católica de Chile. All participants gave written informed consent. The nature of the task was explained to participants upon arrival to the Laboratory. All experiments were performed in the Psychophysiology Lab of the School of Psychology of the same University. Participants sat in a dimly illuminated room, 60 cm away from a 19-inch computer screen with a standard computer mouse in their right hand. All participants were right handed. Prior to starting the task, subjects were fitted with a 32-electrode Biosemi ActiveTwo © digital electroencephalographic (EEG) system, including 4 electrooculographic (EOG) electrodes, two of them placed in the outer canthi of each eye and two above and below the right eye. Continuous EEG was acquired at 2048 Hz and saved for posterior analysis. Throughout the task, participants were instructed to

maintain fixation on a central spot in order to avoid eye-movement related artifacts in the EEG data. All stimuli were presented around fixation or 2.4° to each side of the central fixation spot. This ensured that, despite the instruction to maintain fixation, all participants could easily see all stimuli and perform the task without difficulty in perceptual terms.

5.3. Modeling Framework

Subjects (i.e. users) that undertake a decision-making process act based on information that is, in great part, unavailable to an observer (i.e the interface). While perception of the task environment can be available to both, the subject's prior knowledge, beliefs, preferences and other psychological processes remain hidden from the observer. In other words, the nature of the decision-making processes (i.e. its psychological underpinnings) can only be inferred based on overt actions and, eventually, patterns of overt actions. Importantly, because such hidden information is what makes the difference between different subjects, it represents a natural target for any interface trying to adapt to its current user.

To identify these patterns, we develop a Hidden Markov Model (HMM) based approach that capitalizes on the hierarchical structure of behavior. Specifically, (a) inputs captured by the interface such as mouse clicks, correspond to low-level actions; (b) systematic combinations of low-level actions into high-level subroutines correspond to strategies and (c) the expectation of the use of specific strategies or a combination of them corresponds to complex behavior, which we refer as behavioral modeling in this work. In the following sections we will present the modeling framework in detail taking into account each of these concepts.

5.3.1. Low-level Actions vs Strategies

The simplest action that can be taken on the BDT is to click one of the two possible icons of the binary choice. We will therefore consider these two as the only low level actions for the model. What such low level actions mean or represent, in terms of learning

of the BDT, depends on whether the participant is using them in some systematic way to obtain positive feedback (i.e. a strategy). We consider such systematic combination of low-level actions the high-level strategies of the model.

Different strategies can be used to search a BDT of the characteristics used here. In the following we model four high-level strategies in the form of well defined search strategies, that account for increasingly sophisticated ways to solve the task:

- (i) **Random Search Strategy:** If the participant displays no systematic use of low level actions, we label this as random behavior. In other words, overt actions seem to be unrelated to the task's demands so that we can only assume ignorance regarding the underlying decision making strategy.
- (ii) **Topological Search Strategy:** If the participant shows evidence of searching based exclusively on information regarding the spatial arrangement of the BDT, regardless of the identity of the presented icons (for instance, by choosing to explore from the leftmost to the rightmost branch), we label this as a topological behavior. When clicking based on spatial features, the participant iteratively discards paths of the BDT so that complete knowledge can be obtained only when the 16 paths of the BDT are correctly recognized. Accordingly, as paths share common information, the learning curve of a user invested exclusively in this strategy will grow exponentially as the search space becomes smaller.
- (iii) **Generative Search Strategy:** If the participant shows evidence of considering the identity of individual icons to guide her choice of actions to reach positive feedback, we label this as a generative behavior. In other words, it implies a first level of successful mapping between current task instruction and the specific BDT instance that is being explored (i.e. when the participant learns that a given icon means a given concept). When clicking based on generative relationships, the participant discards subtrees of the BDT where it is not possible to reach positive feedback. Complete knowledge of the BDT can be obtained when the 16 icons are correctly mapped. All the generative relationships can be

learned by being exposed to positive feedback in the 8 paths that contains all of them. Therefore, the learning curve of this strategy is represented by a sigmoid function.

- (iv) **Discriminative Search Strategy:** Here the participant uses a generative model, but adds the ability to learn and relate the negative form of a concept-icon relationship (i.e. learn A by a generative association and then label the neighbor as not-A). In other words, the discriminative search strategy is one that predicts concepts that have not yet been seen in the scope of positive feedback. Concepts are deduced from the context and the understanding of the rules of how the interface works. When clicking based on discriminative relationships, the participant can prune the BDT subtrees more aggressively to obtain positive feedback. As in the generative case, complete knowledge of the BDT can be obtained when the 16 icons are correctly mapped. However, all discriminative relationships can now be learned by being exposed to positive feedback in the 4 paths that contains all of them. Accordingly, the learning curve of this strategy is represented by a sigmoid function that is steeper than in the generative case.

Each of the above models has an initial domain of action that corresponds to the set of actions that can be performed on the BDT according to the strategy's rules. Once exploration of the interface is underway, the initial domain of action of each strategy will necessarily change. This can happen because the user learns something about the specific task instance he is currently solving, or because he learns something about the overall structure of the interface. It is therefore necessary to define criteria that, according to each strategy's rules, allow one to update their domain of action depending on local (task-instance) and global (task-structure) knowledge. Local updates criteria will coincide with the rules of the topological strategy for all systematic strategies, because according to our hierarchical definition, the simplest way to discard places of the BDT systematically is using topological information. Conversely, for global updates –and for the sake of simplicity– we will define knowledge in terms of optimal behavior, (i.e. learning places or concepts of the task instances by repeating the correct path 5 times in a row). Tab. 5.2

presents the formal definition of the domain of action D_i and both updating schemes for each strategy s_i .

To track the BDT knowledge of the specific strategy s_i , we define α_i as a measure of what still needs to be mapped (place or concept) of the BDT at a given time t :

$$\alpha_i(t) = 1 - \frac{|R_i(t)|}{|G_i(t)|} \quad (5.1)$$

Table 5.2. Formal definitions of the strategies are presented according to their initial domain, domain of action updates (task instance learning), and knowledge updates (task structure learning).

Initialization of the Domain of Action		
<i>(defines D_i for a given task instance starting at time t' and target path Q^*)</i>		
<i>strategy</i> <i>(i index)</i>	<i>rule/definition</i>	<i>description</i>
random	$D_i(t) = \{p p \in G_i(t)\}$	The random strategy does not consider learning, thus its domain of action is open to all BDT locations p at any time t .
topological generative discriminative	$D_i(t = t') = \{p p \notin R_i(t') \vee p \in Q^*\}$	At the beginning of a given task, valid actions can be taken only in unexplored locations p or those belonging to the target path Q^* .
generative discriminative	$\forall c c \in R_i(t') \wedge c \in Q^*$ $\rightarrow \text{subtree}(\text{neighbor}(c)) \notin D_i(t')$	Additionally, discard tree branches by using previous known c concept-icon relationships. $\text{subtree}(c)$ is a function that yields c and the set of all locations below c , and $\text{neighbor}(p)$ yields the neighbor of a given location/icon p .
Domain of Action Updates		
<i>(updates D_i while searching for correct feedback in path Q_t at time t)</i>		
topological generative discriminative	$\forall p \text{leaf}(p) \wedge p \in Q_t \rightarrow p \notin D_i(t+1)$ $\forall p p \in Q_t \wedge \text{neighbor}(p) \notin D_i(t)$ $\rightarrow p, \text{parent}(p) \notin D_i(t+1)$	The final selection of a path Q_t is always out of domain in the next iteration. $\text{leaf}(p)$ is a function that is true for each BDT leaf (any 4^{th} level location/icon). Bottom-up rule to discard parent nodes of Q_t when they have already being explore. $\text{parent}(p)$ is a function that yields the location/icon in the level immediately above which leads to p .
Knowledge Updates		
<i>(updates R_i when learning the task instance of target path Q^* at time t)</i>		
topological	$\forall p \text{leaf}(p) \wedge p \in Q^* \rightarrow p \in R_i(t+1)$ $\forall p p \in Q^* \wedge \text{neighbor}(p) \in R_i(t)$ $\rightarrow p, \text{parent}(p) \in R_i(t+1)$	The final selection of a learned path Q^* is always in the learned set of upcoming tasks. Bottom-up rule to set parent nodes of Q^* as learned.
generative discriminative	$\forall c c \in Q_a^* \cap Q_b^* \wedge c \notin \{Q_j t_a^* < j < t_b^*\}$ $\wedge c \in \{Q_j t_b^* \leq j \leq t\} \rightarrow c \in R_i(t+1)$	Concept c is considered learned if there are no selection mistakes between two tasks a and b which intersect in that concept. Task a is learned at time t_a^* , and task b starts at time t_b^* .

where $R_i(t)$ is the set of learned choices, $G_i(t)$ is the set of all distinct choices in the strategy's domain, and $|\cdot|$ is the number of elements of a given set. This parameter is the complement of the specific learning curve of each strategy. Specifically, $\alpha_i = 1$ indicates complete lack of knowledge about the interface, and $\alpha_i = 0$ indicates full knowledge.

Note that high-level strategies evolve according to the user's iterative interactions with the task. For instance, if the participant does not show evidence of learning any path, the learning curve of each strategy is a straight constant line at $\alpha_i = 1$. When feedback becomes available (i.e. when the participant reaches the end of a path producing either a correct or incorrect answer), we ask each model how such observation changes or violates its expected probabilities regarding the nature of future feedback. As long as no learning is involved, all active models will answer equally to this query. However, as evidence of learning becomes available, each model will restrict the domain of possible future actions that are consistent with what the model predicts the participant's knowledge should be. A topological model will label as a mistake any repetition of a path that previously gave positive feedback in the context of a different instruction. A generative model will label as mistakes actions that are inconsistent with a successful icon-concept mapping for which there is prior evidence. The discriminative model inherits the restrictions imposed by the generative model, but will also consider mistakes as those actions that do not take into account not-A type knowledge that the participant should have, given the history of feedback. An important consequence of the above is that strategies can yield the probability of clicking a given icon of the BDT without further training or modeling at any moment throughout the task.

It is worth noting that defining all possible strategies to solve the BDT is not necessary. To delimit the knowledge level of the participant, only the lower and higher bounds of the problem must be defined and more strategies in-between will only increase the framework's resolution. In the BDT case used here, the lower bound is necessarily the random strategy. The upper bound is set by the discriminative strategy because it is the best possible strategy to solve the BDT task (i.e. it requires the least exposure to positive feedback).

Accordingly, we make the assumption that at any given moment, the user has all strategies at his disposal but that overt behavior is best captured by a weighted mix of them. We call this level the behavioral model of the framework.

5.3.2. Behavioral Modeling

Once the high-level strategies are defined, we turn to modeling the expectation about the use of a specific strategy or a combination of them by the participant. Such behavioral model is composed by a modular architecture of the four possible strategies, which interact between them in a HMM-like structure. This modular architecture has the advantage of modeling complex strategies as if they were a single abstract state in the behavioral model.

Formally, a HMM is defined by a finite set of hidden states, s_1, s_2, \dots, s_n , and each time a relevant information arises (e.g. feedback) the system moves from one state $s(t) = s_i$ to another $s(t + 1) = s_j$ (possibly the same). The transition probabilities, $P(s_i \rightarrow s_j)$, determine the probability of transiting between states: $P(s_i \rightarrow s_j) = P(s(t + 1) = s_j | s(t) = s_i)$. The observable information of the process is a finite set of distinct observations, v_1, v_2, \dots, v_m , and a probabilistic function of the states. Accordingly, each observation has an emission probability $P(v_k | s_i)$, $k \in \{1, \dots, m\}$, of being seen under a state s_i . As the system must start somewhere, it is necessary to define an initial probability distribution of the states: $w_i = P(s(t = 0) = s_i)$, $i \in \{1, \dots, n\}$. Given that the sets of hidden states and observations are defined a priori, the only values to estimate are the initial distribution and the transition and emission probabilities. Each strategy therefore takes the role of a hidden state at the behavioral level, which then yields the probability distribution of each observation for each trial. See Fig. 5.2 for a graphical example of a HMM.

In our framework, the states s_i will be represented by a set of strategies that contains search behaviors (strategy models) to solve the task. The transition probabilities will determine the probability that a participant changes her or his search behavior depending on

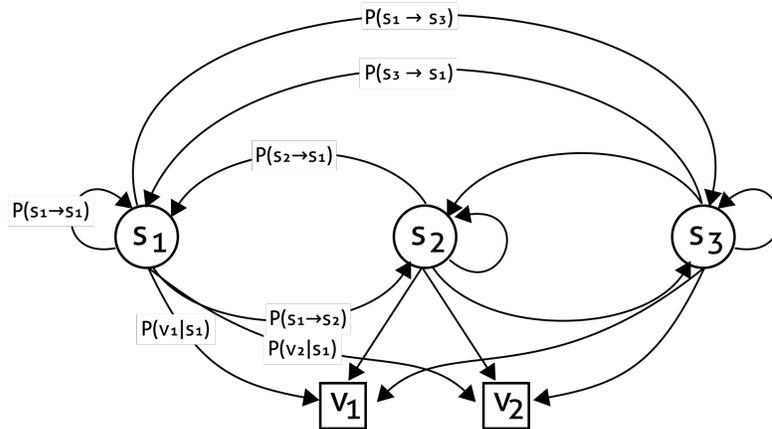


Figure 5.2. **Graphical representation of a Hidden Markov Model.** A HMM with 3-states ($\{s_1, s_2, s_3\}$) and 2 observations ($\{v_1, v_2\}$). Labels for the transition and emission probabilities of the state s_1 are included in the diagram.

her/his level of expertise, and the observable information of the process will be determined by the sequential choice outcome.

5.3.2.1. Emission Probabilities

Although the task can yield positive and negative feedback, an observer can interpret these observations in different ways depending on the situation. While positive feedback is unambiguous (a hit observation), negative feedback can have two different connotations: it is a mistake if, given previous actions, the observer is warranted to assume that the participant should have had the knowledge to avoid performing the action that produced such outcome. Observing such feedback will therefore mean evidence in favor of random behavior. Else, negative feedback is consistent with exploratory search behavior prior to the first positive feedback and, accordingly, not considered a mistake. The set of observations V is therefore defined as: $V = \{\text{mistake, explore, hit}\}$.

Since strategies are sensitive to the context, their emission probabilities change as the participant makes choices. At each sequence step, we calculate the emission probabilities

within the subtree of possible future choices. Considering the last choice of the participant as the root of the subtree, we enumerate all possible future paths of actions Q , defining the following sets at step t :

$$H_i(t) = \{Q | (\forall a \in Q)[a \in D_i(t)] \wedge (\exists a \in Q)[a \notin Q^*]\} \quad (5.2)$$

$$H_i^*(t) = \{Q | (\forall a \in Q)[a \in Q^*]\} \quad (5.3)$$

where a represents a specific action needed to generate path Q , and Q^* the target path. $H_i(t)$ is the set of all paths that do not violate the strategy's rules, while simultaneously allowing the exploration of available choices. $H_i^*(t)$ is the set of paths that lead to positive feedback. Thus, emission probabilities for strategy s_i are defined as follows:

$$P(V|s(t) = s_i) = \begin{cases} \left\{ 0, \frac{|H_i(t)|}{|H_i(t)|+|H_i^*(t)|}, \frac{|H_i^*(t)|}{|H_i(t)|+|H_i^*(t)|} \right\}, & \text{if } |H_i(t)| > 0 \vee |H_i^*(t)| > 0 \\ \{1, 0, 0\}, & \text{otherwise} \end{cases} \quad (5.4)$$

The first case represents emission probabilities for those strategies that, given their rules, allow for future exploration or exploitation. In the second case, when the rules of the strategy cannot explain the current actions, the emission probabilities are fixed to explain mistakes. This is also the case for the random strategy, which is assumed when the observer has no knowledge about the participant's strategy.

5.3.2.2. Transition Probabilities

It is possible, but not necessary, that a participant moves progressively through each of the increasingly complex strategies as he learns the structure of the task. Although such progression may seem as discrete steps (i.e. first using a topological strategy and then abandoning it altogether when conceptual knowledge becomes available), it is most likely that at any given moment of the task, the participant's strategy will fall somewhere in between, being better represented by a mix of strategy models. This is precisely the

distribution captured by the behavioral model. To estimate it, it is necessary to model the interactions between strategies s_i , in terms of transition probabilities and relative weights.

To obtain the transition probabilities we use a voting scheme based on emission probabilities, where each strategy distributes its own $P(V|s_i)$ depending on the ability of the strategy to explain a specific type of observation. Thus, for each observation $v \in V$ we define the set of best explanatory strategies of v as follows:

$$B_v(t) = \{s_i | s_i \in \arg \max_{s_i} P(v|s_i)\} \quad (5.5)$$

Then, every step t in which the participant performs an action, transition links l_{ij}^t between strategies s_i and s_j gain votes according to the following rules:

$$l_{ij}^t = \sum_{v \in V} P(v|s_i) \begin{cases} 1, & \text{if } s_i \in B_v(t) \wedge i = j \\ \frac{1}{|B_v(t)|}, & \text{if } s_j \in B_v(t) \wedge s_i \notin B_v(t) \\ 0, & \text{otherwise} \end{cases} \quad (5.6)$$

These rules represent three cases: (a) Any strategy that belongs to $B_v(t)$ strengthens its self-link, not sharing its $P(v|s_i)$ with other strategies. (b) Strategies that do not belong to $B_v(t)$ generate links to those that best explain the observation v , losing their $P(v|s_i)$ in equal parts to those that best explain v . (c) Strategies that do not explain v or do not belong to $B_v(t)$ do not receive votes for observing v .

In the case of the random strategy, its emission probabilities are fixed to $\{1, 0, 0\}$, therefore the above rules do not generate links with other strategies in the case of exploring or exploiting the BDT knowledge. In order to overcome this limitation, we define the set of strategies that can leave the random strategy as those that do not see the current action as a mistake, plus the random strategy itself:

$$U(t) = \{s_i | P(\text{mistake}|s_i) = 0\} \cup \{s_{\text{random}}\} \quad (5.7)$$

Then, the links from random strategy are voted as:

$$l_{random,j}^t = \frac{1}{|U(t)|} \begin{cases} 1, & \text{if } s_j \in U(t) \\ 0, & \text{otherwise} \end{cases} \quad (5.8)$$

Note that $\sum_v P(v|s_i) = 1$ for each strategy, thus the vote sharing scheme is always normalized.

The value of these links represent only what happens at the current time. The participant's actions, however, can be tracked historically by defining an observation window of the process τ , which modulates the weight of the votes over time. Therefore, at time t , the amount of cumulative votes L between strategies i and j is defined by:

$$L_{ij}(t) = \frac{\sum_{n=0}^t l_{ij}^n \tau(n)}{\sum_{n=0}^t \tau(n)} = P(s_i \rightarrow s_j) \quad (5.9)$$

Here we define τ as a Gaussian function with a standard deviation of σ steps, normalized to a maximum of 1 at the current time t :

$$\tau(n) = e^{-\frac{(n-t)^2}{2\sigma^2}} \quad (5.10)$$

The result is a set of directed interactions (i.e. transition probabilities) among different strategies, storing historical information of the participants' behavior.

5.3.2.3. Weights Optimization

Weights represent the probability distribution across strategy models at each iteration. Once emission and transition probabilities are known, weights are updated by comparing the cost, in terms of probabilities, to start in some strategy and end in the best transition link that explains an observation type. This allows us to compare the transition that best represents the current state of the behavioral model (to where the system is moving) for a specific observation, and how much it costs for each model to reach that point.

The cost, in terms of probability of moving from the state s_k to s_j (which could be the same), given an observation $v \in V$, is represented by $P(s_k \rightarrow s_j)P(v|s_j)$. Note that there may be more than one best transition for the system. Accordingly, we define set of best transition links as:

$$\{(s_k, s_j) | \arg \max_{s_k, s_j \in M} P(s_k \rightarrow s_j)P(v|s_j)\} \quad (5.11)$$

where M represents the set of all strategy models. Eq. 5.11 implies that the behavioral model identifies the transition from s_k to s_j as one of the most representative in case of observing v at time t . In case more than one best transition exists, the most convenient path is optimized. Then, for an initial state s_i , the cost of reaching the best transition link is defined as the path that maximizes the following probability:

$$L_{ii}(t)P(v|s_i)P(s_a \rightarrow s_b)P(v|s_b) \cdots P(s_k \rightarrow s_j)P(v|s_j) \quad (5.12)$$

We use L_{ii} as the prior for the initial probability $w_i(t)$, so that models with $w_i(t-1) = 0$ can be incorporated in the optimization at time t . Note that Eq. 5.12 is equivalent to the Viterbi path constrained to a fixed observation v .

Then, the weight of strategy s_i at time $t+1$ is represented by the total cost for the set of observations V :

$$w_i(t+1) = \sum_{v \in V} w_i(t)P(v|s_i) \cdots P(s_k \rightarrow s_j)P(v|s_j)\tau_v(t) \quad (5.13)$$

where $\tau_v(t)$ yields the relevance of observation v in the total cost function, given an observation window τ (such as defined in Eq. 5.10):

$$\tau_v(t) = \sum_{n=0}^t \tau(n) \begin{cases} 1, & \text{if } v(t=n) = v \\ 0, & \text{otherwise} \end{cases} \quad (5.14)$$

The most likely observation at time step $v(t)$ is defined by consensus regarding the observation with the best average emission probabilities:

$$v(t) = \arg \max_{v \in V} \sum_{s_i \in M} P(v|s_i) \quad (5.15)$$

Finally, in order to obtain a probability distribution over the search strategies, each weight is normalized by the coefficient W defined as:

$$W = \sum_{i \in M} w_i(t + 1) \quad (5.16)$$

Note that if two or more models have the same domain of action (e.g. both concept-based strategies have the same domain of action for target path Q^*), we consider only the one with greatest knowledge. Otherwise the normalization is unfair to the other models.

5.3.2.4. Learning Curve

Each strategy's individual knowledge of the BDT, α_i , can be combined with its respective weight w_i to produce a mixture of basal strategies α^* . This is accomplished by a weighted sum:

$$\alpha^*(t) = \sum_{i \in M} \alpha_i(t) w_i(t) \quad (5.17)$$

Recall that α_i is the complement of the specific learning curve of each strategy. Therefore, the approximate learning curve of a given participant can be obtained as: $1 - \alpha^*$.

5.3.2.5. Predicting Participants' Choices

As the models' weights are known previous to icon selection, we can build a Mixture of Experts (Jacobs, Jordan, Nowlan, & Hinton, 1991) where search strategy models become the experts that must answer the question "Which icon is the participant most likely to click in the next step?". Because each model can produce any of two possible actions, i.e clicking on the left or the right icon, the most probable next choice at step t will be the one that has the largest support as expressed by the weighted sum rule of the ensemble:

$$\arg \max_{a \in \{\text{left}, \text{right}\}} \mu_a(t) = \sum_{i \in M} w_i(t) \begin{cases} 1, & \text{if } a \in h_i(t) \\ 0, & \text{otherwise} \end{cases} \quad (5.18)$$

where $h_i(t)$ represents the expert-prediction of the strategy i . Prediction can be based on dichotomous decisions or random selection. Dichotomous decisions occur when there is only one possible action in the strategy's domain of action so that it will be selected with probability equal to 1. Alternatively, when both possible actions belongs to the strategy's domain (i.e. have equal probability of being selected) or when neither of the actions belongs to the domain (i.e. the strategy is in the inactive set), a coin is tossed to choose at random (50/50 guess).

It is worth noting that the prediction capability of each strategy depends on the size of its domain of action. Strategies with wider exploratory behaviors often have larger domains, and consequently, less predictive power due to the number of paths that can be selected. This aspect is captured by the number of 50/50 guesses, because, as mentioned above, whenever the strategy has more than one equally likely possibility of action, it must choose at random. This does not mean that the strategy itself is failing to capture the participants' choice of action, but that the conditions are ambiguous enough to keep looking for the correct path.

5.3.3. Scores

To better visualize the search process of participants (and groups of participants), we introduce three scores. As any sequence of choices can be the consequence of different degrees of expertise (from fully random behavior to goal-directed exploitation), we define a scale where we assign points depending on the most likely observation $v(t)$ that yields the behavioral modeling (Eq. 5.15) at each step of a sequence of choices Q .

To measure the degree of expertise for path Q , realized between time steps t_a and t_b , we define the expertise score as:

$$\text{expertise} = \frac{1}{|Q|} \sum_{t_a \leq t \leq t_b} \begin{cases} 1, & \text{if } v(t) = \text{hit} \\ 0.5 & \text{if } v(t) = \text{explore} \\ 0, & \text{if } v(t) = \text{mistake} \end{cases} \quad (5.19)$$

where expertise equal to 1 means exploitation behavior and expertise equal to 0 means completely random behavior. Values in-between represent various degrees of exploration.

If only quantifying exploration and exploitation rates, we assign points equal to 1 only if $v(t)$ match the explore/hit observation respectively:

$$\text{exploration} = \frac{1}{|Q|} \sum_{t_a <= t <= t_b} \begin{cases} 1, & \text{if } v(t) = \text{explore} \\ 0, & \text{otherwise} \end{cases} \quad (5.20)$$

Likewise, exploitation score is defined as:

$$\text{exploitation} = \frac{1}{|Q|} \sum_{t_a <= t <= t_b} \begin{cases} 1, & \text{if } v(t) = \text{hit} \\ 0, & \text{otherwise} \end{cases} \quad (5.21)$$

5.3.4. Clustering

The behavioral modeling provides interesting indicators to grasp the complexity of the decision-making processes behind the participant's actions. Here we describe the prediction of choices and the learning curve as potential indicators to cluster task events or EEG epochs.

5.3.4.1. Observation Clustering

The most likely observation predicted by the behavioral model (Eq. 5.15) allows to separate each icon selection into three categories: random, exploration, and exploitation. The random condition considers cases when participants make mistakes during the search of the correct path ($v(t) = \text{mistake}$). The exploration condition applies to systematic exploration scenarios ($v(t) = \text{explore}$), and the exploitation condition applies to goal-directed exploitation of the BDT knowledge ($v(t) = \text{hit}$).

5.3.4.2. Prediction Clustering

The prediction of actions allows to separate each icon selection into three categories: likely choices, unlikely choices, and 50/50 guesses. Using the same formulation of the Mixture of Experts (Eq. 5.18), we define likely/unlikely choice as the case when the participant chooses the same/opposite action predicted by the behavioral model, and the model certainty about its own prediction is greater than 50%. In the 50/50 guess case, independent of the action predicted by the model, the uncertainty of the prediction is greater than 50%, so we label the action decided by the model as random.

5.4. Recordings

5.4.1. Mouse-Tracking Recordings

Screen x and y pixel coordinates of the mouse pointer were acquired at 60 Hz (the same refresh rate of the screen). The center of the screen, of 1280x1024 pixels, was used as the origin of the coordinate system. The distance in pixels from the center of the screen to the center of each icon was 120 pixels, and the width and height of each icon was 64 pixels.

5.4.1.1. Mouse-Tracking Pretreatment

Pixel coordinates were converted to absolute values in order to make symmetrical the movements towards the left/right icon positions. Distance covered between pixel samples were calculated using the Euclidean metric. Continuous mouse-tracking data was then epoched starting 800 ms prior to response (click) onset and ending 800 ms after.

5.4.1.2. Mouse-Tracking Analysis

Two main mouse analyses were undertaken based on averaging the previously mentioned mouse-tracking epochs: response clustered by the prediction capabilities of the

modeling framework (Eq. 5.18), and by response clustered by the most likely observation (Eq. 5.15). All conditions with less than 20 trials were discarded to obtain averages and from the statistical analysis.

5.4.1.3. Statistical Analysis

Our main goal is to establish if different behavioral modeling measurements are related to different mouse movement behaviors. Thus, we use the distance covered before and after clicking an icon as the analyzed variables. We therefore have a three-way ANOVA mixed design with two level by each factor, with two repeated measures factors (cluster condition and before/after mouse response) and one inter-subject factor (participant's group). In order to analyze the possible interaction among these three factors we used an ANOVA instead of non-parametric versions (Kruskal-wallis and Friedman's test). To assure that ANOVA would work well under low sample size and eventually unbalanced by condition (depending on cluster analysis result), we evaluated all assumptions. When sphericity assumption was not met, we corrected the p-value based in Greenhouse-Geisser method, reporting the Greenhouse-Geisser epsilon (GGe), followed by the corrected p-value (p[GG]). For all ANOVA models, effect size is reported using Generalized Eta-Squared denoted as η_G^2 (Bakeman, 2005). Also, normality of dependent variable is reported using Shapiro-Wilk Test, included at the end of each ANOVA report. We maintained the ANOVA approach when normality assumption was not met, as ANOVA is robust under such scenario (Schmider, Ziegler, Danay, Beyer, & Bühner, 2010). Nonetheless, we still reported if normality of dependent variable was violated by using Shapiro-Wilks test. Finally, we rerun all the analysis using a Permutation ANOVA method. This method is a good call to confirm results specially under low sample sizes and unbalanced conditions. All post hoc analyses were run using a Holm-Sidak correction considering the repeated measures.

5.4.2. Electrophysiological Recordings

Prior to starting the task, participants were fitted with a 32-electrode Biosemi ActiveTwo© digital electroencephalographic (EEG) system in an extended 10-20 configuration (Jasper, 1958), including 4 electrooculographic (EOG) electrodes, two of them placed in the outer canthi of each eye and two above and below the right eye. Continuous EEG was acquired at 2048 Hz and saved for posterior analysis.

5.4.2.1. EEG Pretreatment

Specific data processing were made to use the Event Related Potential (ERP) technique (Coles & Rugg, 1996; S. Luck, 2005), using the MNE-python software (Gramfort et al., 2013). Data was re-referenced to mastoids and bandpass filtered between 0.5 and 40 Hz. Then, continuous data were submitted to independent component analysis (ICA) to remove vertical and horizontal EOG components of the signal.

For the analysis of ERPs associated with the attentional checkerboard probes, continuous EEG was epoched starting 100 ms prior to probe onset and ending 400 ms after, to avoid contamination by successive probes. In the case of ERPs obtained to responses (mouse click) and feedback presentation (L4 of the BDT) continuous EEG was epoched starting 200 ms prior to response/feedback onset and ending 800 ms after. Artifact detection was performed on segmented data using a peak-to-peak threshold algorithm, with a voltage threshold of 150 μV . All epochs containing detected artifacts were rejected.

5.4.2.2. ERP Analysis

Five main ERP analyses were undertaken based on averaging the previously mentioned EEG epochs: ERP waveforms evoked by the attentional probe stimulus on left/right sides of the clicked icon, by the feedback obtained in the sequence of choices, by response clustered by the prediction capabilities of the modeling framework (Eq. 5.18), and by response clustered by the most likely observation (Eq. 5.15). All ERPs were baseline

corrected using the time window before the onset of the respective stimulus (100 ms in the case of checkerboard probes and 200 ms in the rest of the cases).

In the ERPs clustered by mouse responses, only responses from levels 1 to 3 of the BDT were considered. This is because the last response (last level of the BDT) shows explicit feedback of the sequence of choices, whereas the previous levels show only implicit feedback. All conditions with less than 20 trials were discarded to obtain evoked ERPs and from the statistical analysis.

5.4.2.3. Statistical Analysis

For statistical analysis, we use a non-parametric permutation clustering F-test (Maris & Oostenveld, 2007) to determine the significant spatiotemporal differences between conditions. We used this approach to handle the Multiple Comparison Problem (MCP) of the EEG data in order to find significant groups of electrodes, and also the temporary window where they occur.

In this method, all significances are family-wise error rate (FWER) corrected, and as being non-parametric, assumptions of Gaussian data distributions do not need to be satisfied. Specifically, clusters of significant areas were binded together by pooling neighboring sites that showed the same effect. The cluster-level statistic was computed as the sum of the F-values of all sites within the corresponding cluster, ending up with one statistic for each cluster. With a p-value of 0.05, we use an F distribution to compute the threshold to determine the significance of each cluster. Specifically, we used the Percent-Point function (PPF) of the F distribution with the degrees of freedom of the comparison. The permutation distribution was then obtained by randomly permuting conditions between samples and recomputing the clusters and the test statistics. After 2000 permutations, the cluster-level significance for each observed cluster was estimated as the proportion of elements of the permutation distribution greater than the cluster-level statistics of the corresponding cluster.

We also use a template of connectivity as a spatial prior to the clustering. The template was based on automatic symmetric triangulation of the 2D layout template (see Fig. 5.3 for a graphical representation of the connectivity template.).

Finally, it is worth noting that the statistical analyzes presented with the ERP results are exploratory, as the sample sizes of learners ($N = 14$) and non-learners ($N = 8$) are small and asymmetrical. This is because a priori we can not tell in which group a participant will be. Furthermore, the previous separation was found a posteriori, therefore a larger sample was not collected to address this problem. From the statistical point of view, this translates into incomplete distributions, so that the N is insufficient to properly determine

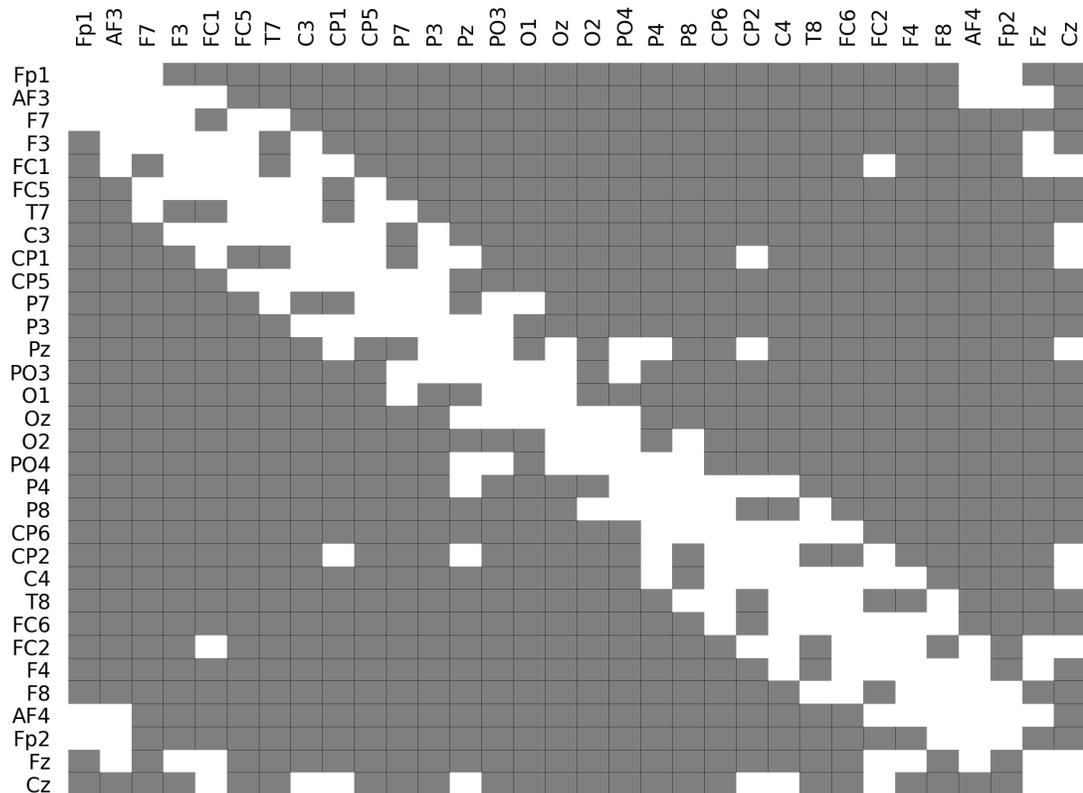


Figure 5.3. **Electrodes connectivity template.** White squares of the matrix represent electrodes that are connected. Gray squares represent electrodes that are not connected to each other. The template was built on an automatic symmetric triangulation of the 2D layout template.

whether a participant is an outlier or not (since the standard deviation is not representative of a normal distribution). In addition, there is no suitable normality test which operates in this range of N. However, we expect that an increase in the sample size could improve the results.

6. RESULTS

In this study we recruited twenty-two participants which were confronted with a decision-making search task structured as a four-level Binary Decision Tree (BDT). Participants were asked to discover an underlying concept-icon mapping, receiving feedback only at the end of the sequence of choices. In the following sections we will examine the behavioral and neurophysiological changes observed throughout the experiment.

6.1. Behavioral Results

Reaction times (RT) averaged over all participants for individual repetitions of each task instance and for the overall experiment are presented in Fig. 6.1. As different participants have different number of trials per task instance, we calculated a representative

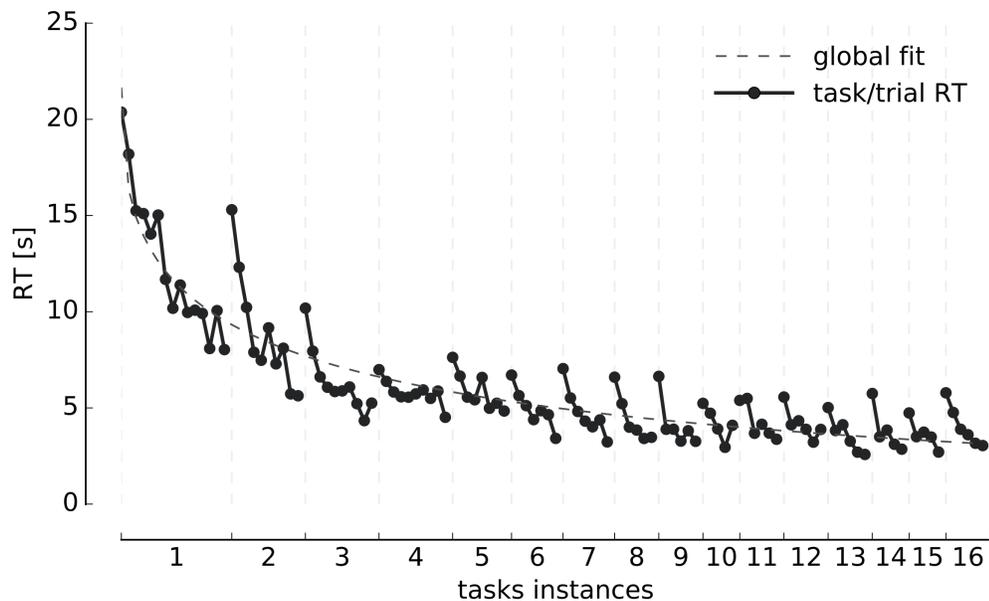


Figure 6.1. **Reaction Times.** Evolution of trial reaction times (y-axis) grouped by tasks instances (x-axis) and averaged over all participants. The global RT curve fit, corresponds to an exponential decay function of the type: $\lambda_1 \exp(-\lambda_2 x^{\lambda_3})$, where $\lambda_1 = 21.6$ is the starting average RT, $\lambda_2 = 0.28$, and $\lambda_3 = 0.4$. The coefficient λ_3 is necessary since the drop in time is not as steep as when $\lambda_3 = 1$ (the usual exponential decay constant).

average of each instance through the following interpolation scheme: we estimated a bin size for each task by calculating the average number of trials per task over the group of participants. Each individual’s task-related vector was then linearly interpolated in that common space. This preserves the relative time that it takes –on average– for the participants to complete each task instance, while it allows to visualize the average of the reaction time and learning curves (Fig. 6.3) more naturally.

A consistent two-fold exponential structure is visible revealing the development of expertise. Each individual task instance (same instructions) takes progressively less time to solve as participants repeat it. Likewise, successive instances of the task (different instructions) take progressively less time to solve as the session unfolds. Participants take an average of 29 ± 9 minutes (mean \pm s.d.) to complete the entire experimental session, with a minimum of 13 minutes and a maximum of 40 minutes, which was the maximum time allowed to solve the task.

6.2. Model Parameter Dependence

To illustrate how the modeling results are influenced by the observation window of the process τ (Eq. 5.10), we analyze the dynamics of τ while tracking the weight of a participant’s discriminative strategy (Fig. 6.2).

The observation window τ affects how observations and the voting scheme impact the point of view of the observer when determining the use of a given strategy by the user. This parameter directly affects the transition probabilities of the behavioral modeling, and the weights through the temporal relevance of the observations. To test its sensitivity, we built Gaussian kernels of σ equal to 1, 12 and 20 steps, which represent 1, 3, and 5 trials/sequences of choices (see inset in Fig. 6.2 for a graphical representation of the kernels). These kernels cover cases between the maximum temporal weight in the current choice ($\sigma = 1$) to 50% of the temporal weight at approximately 5 trials of distance ($\sigma = 20$). The highest σ value was chosen such that it includes the temporary extension of the

average number of trials to find positive feedback (8 paths), when the domain is the entire BDT.

As seen in Fig. 6.2, these cases capture the general dynamics of the learning process. For example, task instances 1 and 2 are statements which belong to the same branch of the BDT. The participant uses the information learned in instance 1 to answer instance 2, what is reflected positively in the weight of concept-based strategies such as the discriminative strategy. Conversely, task instance 3 does not belong to the same branch of instances 1 and 2. This time the participant is faced with a more explorative situation, which is reflected in a decrease in the weight of the discriminative strategy. Finally, around task instances 4 and 5 the weight of the strategy is consolidated around the maximum weight, which is expected in a fully discriminative behavior, where 4 paths of positive feedback can describe the full knowledge of the BDT.

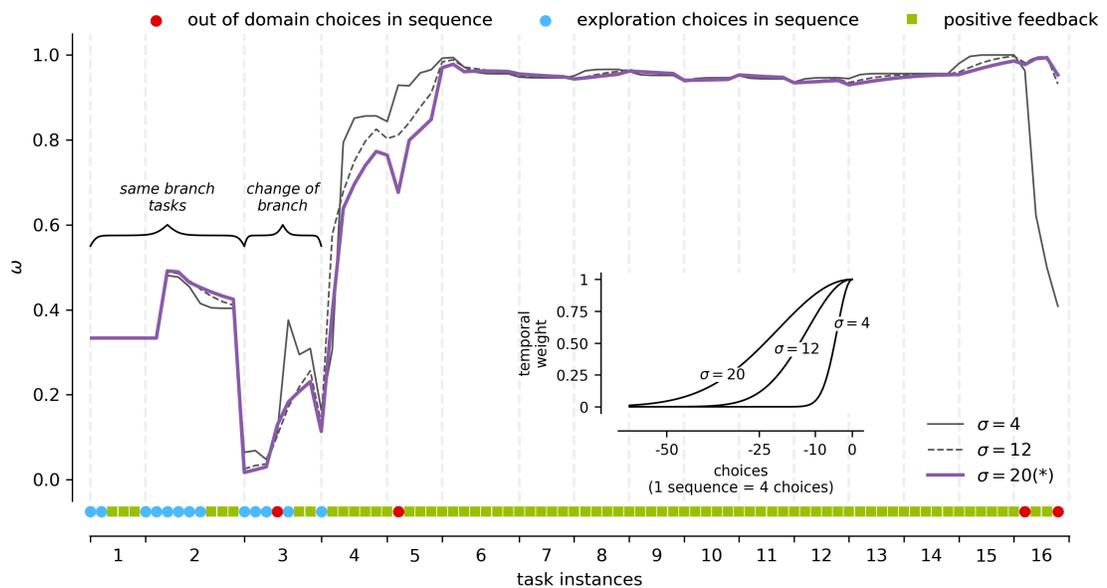


Figure 6.2. **Framework parameters modulation.** An example for a participant’s discriminative strategy weight (y-axis) for each task instance (x-axis) at different τ values (see inset). The inset x-axis presents the distance in choices from the current time and the y-axis the respective kernel weight. The curve marked by (*) represents the σ used for modeling the rest of results presented in the section.

Regarding the specific sensitivity of τ across kernels, small kernels tend to react faster to changes in the style of the participant, punishing or rewarding the weight of a strategy very quickly. For example, a $\sigma = 1$ kernel produces “spikes” in Fig. 6.2 in tasks instance 3 when the participant goes from mistake to exploration (rewarding case). Likewise, in tasks instance 16, when participant performs incorrect actions in a context of full knowledge, the weight of the strategy is punished abruptly to match other strategies that are compatible with such behavior. On the contrary, large kernels tend to produce smoother transitions, as a more extensive history is taken into account when calculating the relevance of the current observation. For example, task instance 4 in Fig. 6.2 represents a shift in the strategies’ weights towards discriminative behavior. As expected, the $\sigma = 1$ kernel has a steeper curve than the $\sigma = 20$ kernel around shift time. Finally, when a strategy is irrelevant in the current weights distribution, as seen from task instance 6 to 15 of Fig. 6.2, distinct kernels make no difference in the strategy’s final weight curve.

Although τ can be set for different usage scenarios and types of tasks, we use $\tau(\sigma) = 20$ for modeling our data hereafter. This choice produces smoother changes in the α curves, clearly showing the process of expertise acquisition.

6.3. Individual Differences

It was possible to distinguish two groups of participants according to whether they managed to solve the task (learners, N=14) or not (non-learners, N=8) in the allowed time window (40 min). We consider learners, Fig. 6.3A, all participants who managed to complete the 16 tasks instances, reaching full knowledge of the interface as evidenced by consistent positive feedback. Conversely, the non-learners group, Fig. 6.3B, is composed by participants who did not complete the task within the time limit, or failed to reach full knowledge of the interface.

For the learners group, the left panel in Fig. 6.3A shows the distinctive shape of each strategy as they reach full knowledge. The right panel in Fig. 6.3A shows the average

weight of each strategy model across the sixteen different task instances. Recall that the absence of learning under the random strategy is represented as a straight, constant line at

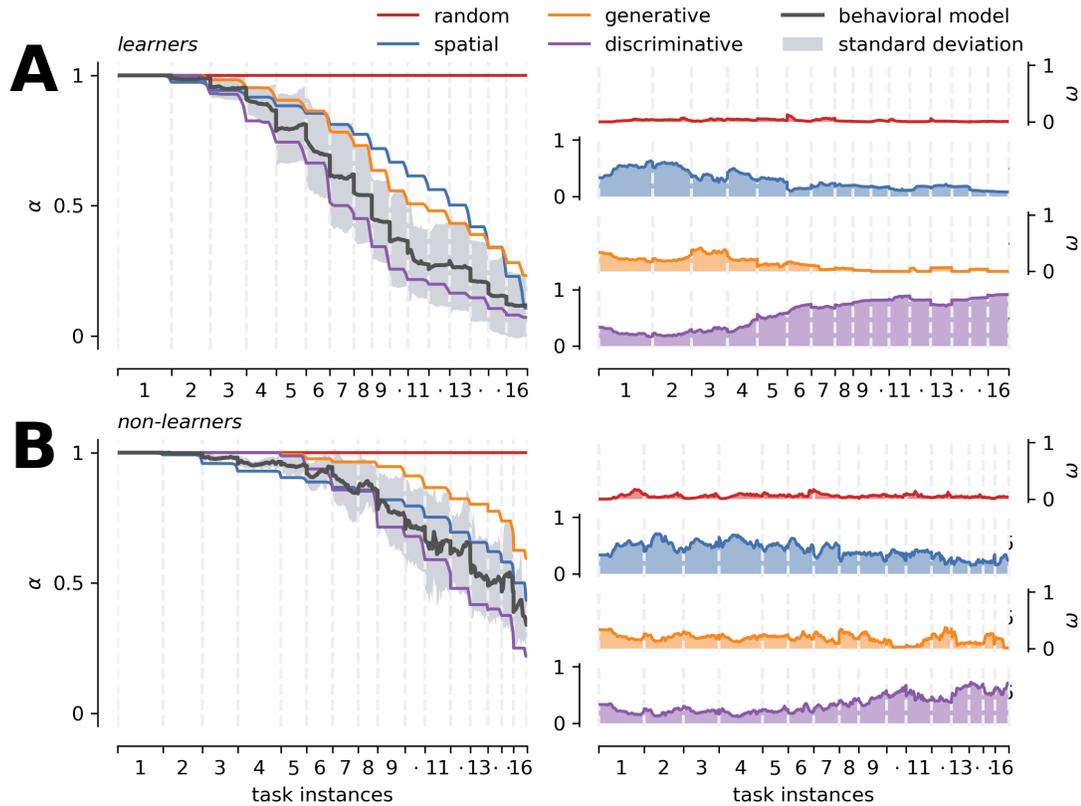


Figure 6.3. **Learners and non-learners average strategies across tasks.** (A) Right panel: characterization of the strategy models averaged over all participants (66.7%) that managed to solve the sixteen different instances of the task (x-axis). Left panel: average weights of each strategy models over the same group of participants. (B) Right panel: characterization of the strategy models averaged over all participants (33.3%) that didn't fully learn the task. Left panel: average weights of each strategy models over the same group of participants. The behavioral modeling represents the averaged individual approximation (basal strategies weighting) and is shown with its standard error across tasks. Dispersion of α curves for basal strategies is not presented to facilitate visualization. The latter can be seen in greater detail in Fig. 6.4 below. The same interpolation scheme as in Fig. 6.1 is used.

$\alpha = 1$. In the topological model, as learning progresses, the number of attempts necessary to encounter positive feedback decreases exponentially. Finally, generative and discriminative models are represented by increasingly steep sigmoid functions, as complete knowledge of the BDT can be obtained with less positive feedback exposure. It is worth noting that the definition of learning for concept-based strategies (no selection mistakes between two consecutive paths with the same target concept), causes that not all learning curves reach zero for all participants. This is mainly because not all concepts can be checked for the learning condition in one run of the task. The average group-level behavioral modeling remains between the space delimited by the generative and discriminative strategies, following more closely the discriminative strategy. The right panel in Fig. 6.3A shows that the weight of the discriminative strategy starts outperforming the rest of the strategies around task instance 4. Before that, the topological strategy is used to locate positive feedback. Generative strategy has a short transition period between topological and discriminative strategies around task 3, losing prominence quickly after that.

In contrast, non-learners tend to go through successive task instances without obtaining positive feedback (i.e. they are prompted to continue after persistent mistakes, see Methods section). Accordingly, in this group all strategies suffer substantially regarding knowledge completion, never reaching full knowledge (Fig. 6.3B left panel). In general, non-learners do not generate as much explicit knowledge as learners do, focusing on topological exploration over conceptual mapping in order to find positive feedback. This is reflected in the fact that the topological α curve surpasses the generative α curve. Importantly, however, some implicit knowledge seems to be generated based on the tasks that they do manage to complete. This would explain why the discriminative strategy gains weight towards the end of the task (Fig. 6.3B right panel task instance 9). In this group, the average behavioral modeling moves between topological and discriminative strategies.

Fig. 6.4 presents the deployment of strategies and their corresponding weights for representative cases of the two groups. In these cases, each strategy reaches different levels of knowledge depending on the participants' context. A highly proficient learner, Fig. 6.4A,

usually goes from random to discriminative strategies progressively, as seen in the weights

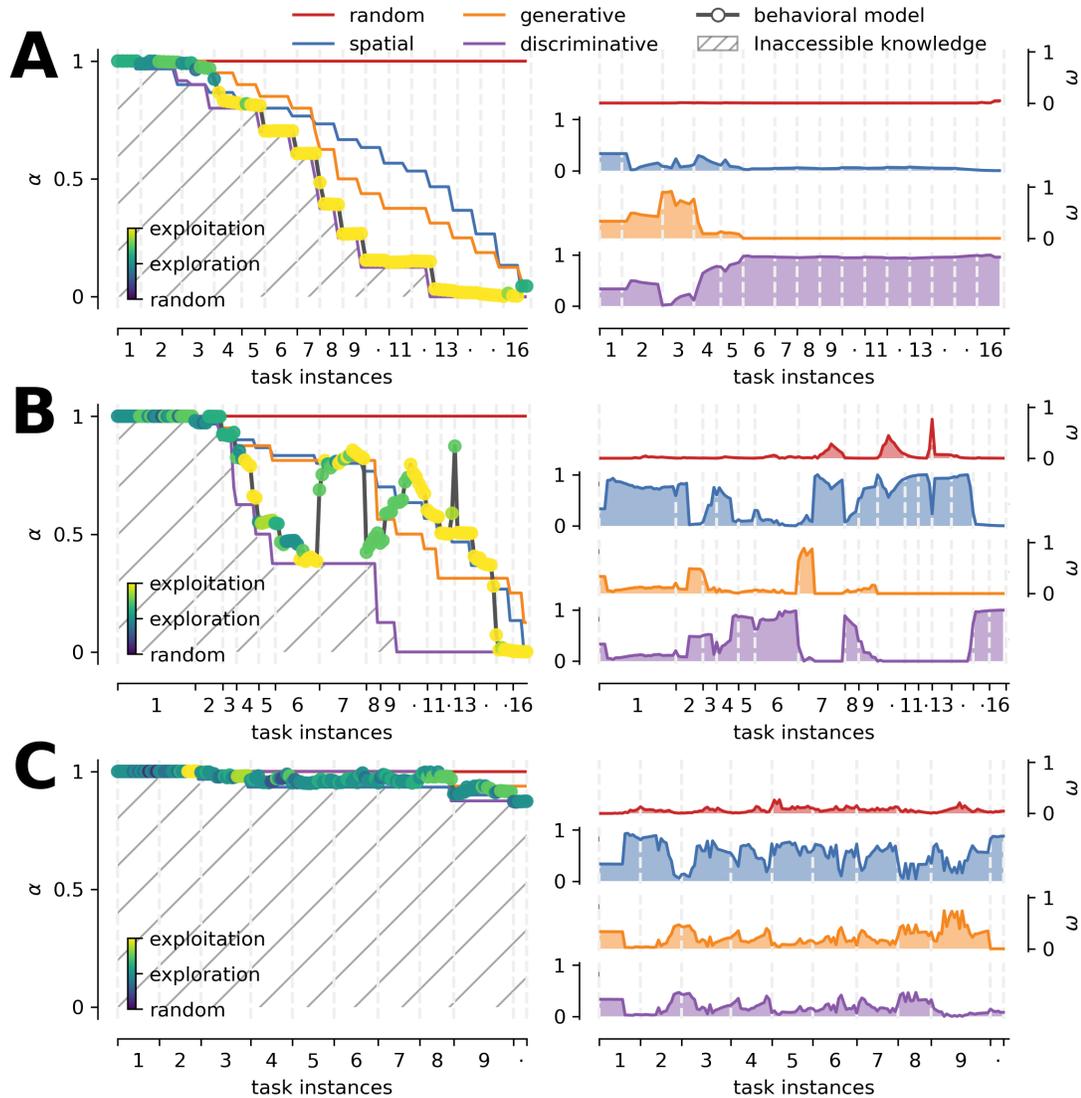


Figure 6.4. **Study Cases.** Three representative individual cases according to task performance. (A) Highly proficient learner. (B) Less proficient learner. (C) Non-learner. The left panel of each case shows the deployment of basal strategies in terms of α during the overall duration of the experiment (x-axis). Colored dots represent different degrees of expertise used to explore/exploit the interface knowledge (Eq. 5.19). The right panels of each case shows the evolution of the corresponding weights for each strategy across task instances. The average of the basal strategy weights yield α^* for the behavioral modeling.

panel. Few mistakes are committed, and the learning curve $(1 - \alpha^*)$ approximates a sigmoidal shape. A less proficient learner such as the one shown in Fig. 6.4B, initially goes through a similar process, but also through periods of difficulty in solving the task (e.g. from task instances 7 to 9 in Fig. 6.4B). The corresponding learning curve presents a sigmoid shape interrupted by intervals of evidence for topological and random strategies, which is caused by explorations and mistakes. These intervals have different durations and extensions, finally disappearing when the discriminative strategy regains dominance. A non-learner, as the case shown in Fig. 6.4C, typically has a similar starting strategy distribution as learners. However, instead of converging to the discriminative strategy, the participant falls back mainly to the topological strategy. As they do not appear to learn most of the concepts, they must search for positive feedback in less efficient ways during the whole task.

6.4. Predicting Participants' Choices

In order to better understand the actions of the participants, we tracked the emission probabilities of exploration (Eq. 5.20) and exploitation (Eq. 5.21), comparing them with the actual performance of the participant in the task. Fig. 6.5 leftmost panel shows these comparisons for learners and non-learners. In the case of learners, the model identifies complementary curves for exploration and exploitation (exploration + exploitation ≈ 1). Learners tend to be more explorative at first, but their performance falls between exploration and exploitation. From task instance 4 onwards, the performance of learners is similar to what the model predicts as exploitation. After that, exploration behavior tends to disappear and exploitation increases. This is congruent with Fig. 6.3A left panel, where learners begin to use the discriminative strategy more consistently from task instance 4. In contrast, non-learners do not have complementary exploration and exploitation curves. This is because of the random behavior during the search of positive feedback. Although the actual behavior of non-learners is similar to the exploitation curve, the performance

level remains under 50%. This is probably because they only map the easiest levels of the BDT (level 1 and 4), while exploring the rest of the tree in each task instance.

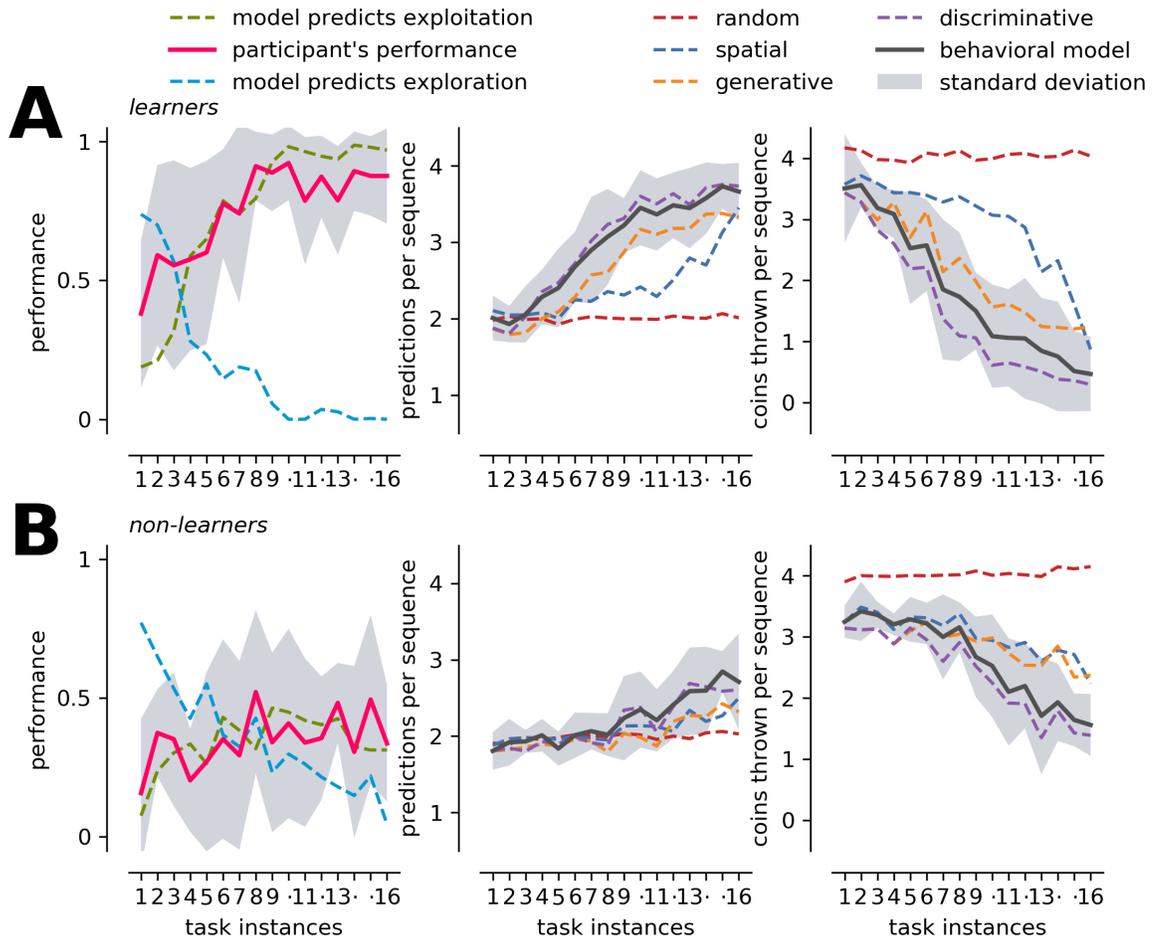


Figure 6.5. Participants performance, framework prediction, and 50/50 guesses averaged across tasks. (A) presents results for the learners group: the leftmost panel shows the participant's averaged performance (y-axis), compared with exploration/exploitation performance as predicted by the behavioral modeling; the center panel present the average number of predicted choices (y-axis) in a sequence for behavioral model and individual strategies; the rightmost panel present the number of 50/50 guesses (y-axis left) for center panel. The x-axis for all panels is the average across the sixteen different tasks. (B) presents the equivalent measures of (A) for non-learners. The corresponding standard deviation of participants' performance and behavioral modeling is shown in gray. Standard deviation for the rest of the curves is not presented to facilitate visualization.

To compare the performance of our behavioral model, we use a Mixture of Experts (Eq. 5.18) to predict the participant's most likely next choice. Prediction outputs two values: a next possible action and the number of 50/50 guesses (see Methods), normalized by the experts' weights. The same concepts apply to individual strategies, so that the prediction capabilities can be compared by the prediction output of each strategy. Fig. 6.5 center and rightmost panels compare choice prediction capabilities and the number of 50/50 guesses for both learners and non-learners. At the beginning, when the knowledge of the BDT is still poor, we expect a high 50/50 guess rate with a consequently low prediction accuracy. In both groups, prediction accuracy starts at around 25% (two choices). This is because, on average, it is not difficult to guess the first and four level icons at random, regardless of it being a participant or an observer. From then on, learners progressively improve their performance, reaching values close to 90% halfway through the experiment, and near 100% towards the end of the task. The number of 50/50 guesses follows the previous trend, stabilizing near zero around task instance 10. In the case of non learners, the number of 50/50 guesses is quite high, only decreasing around task instance 9 and converging to two 50/50 guesses choices. The initial prediction capabilities of two choices is maintained throughout the course of the task, and only rises to about three choices towards the end of the task.

In both groups the leading single strategy is the discriminative, which is followed closely by the behavioral model. This is consistent with the results presented in Fig. 6.3, where the weight of the discriminative strategy follows the critical moments where it differs from the rest of the strategies (task instance 4 for learners and 9 for non-learners). This suggests that the core rules of the BDT mapping are learned through this mechanism. However, the discriminative strategy is not able to fully explain the behavior of the participants in all the stages of expertise acquisition (and the other strategies do not explain it either). This may be due to the fact that before the critical point of the discriminative strategy, participants are not as systematic as the strategies propose (or it is necessary to define different search strategies to explain such behavior). In the case of learners, the behavioral model predicts more exploratory behavior (Fig. 6.5 leftmost panel) before that

point. In other words, the modeling framework predicts that there will be a greater amount of exploration, independent of the exact way in which it would take place.

6.5. Mouse-Tracking

In addition to the previous behavioral results, mouse-tracking data was obtained for learners and non-learners groups. Fig. 6.6 and Fig. 6.7 present the averaged x and y pixel coordinates, as well as the distance covered by the mouse in the 800 ms prior to response

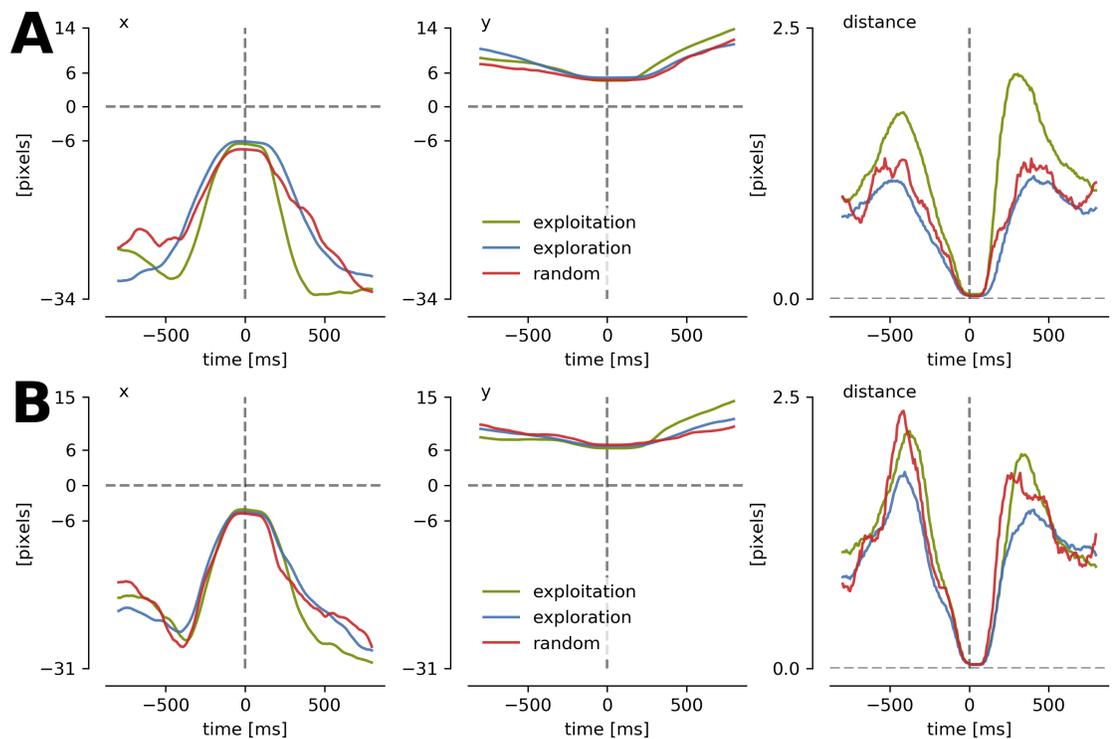


Figure 6.6. **Mouse-Tracking averaged data before and after clicking an icon separated by the most likely observation predicted by the model for learners and non-learners groups.** (A) presents results for the learners group: the leftmost and center panels show the participant's averaged x and y pixel coordinates (y -axis) before and after clicking and icon in the interval -800 to 800 milliseconds (x -axis); the rightmost panels present the average distance (y -axis) covered by the mouse in the same time interval mentioned above. (B) presents the equivalent measures of (A) for non-learners group.

(click) onset and ending 800 ms after, for the learners and non-learners groups. Note that, as expected from the design, movements occur mainly on the x-axis in both groups, and click responses are performed on average ($x=6$, $y=5$) pixels away from the center of the icons (upper right quadrant of the icon). Using the covered distance as our exploratory variable, we ran a three-way mixed ANOVA. We analyze movements clustered by the most likely observation predicted by the model (Fig. 6.6), and clustered by the action predicted by the model (Fig. 6.7) independently.

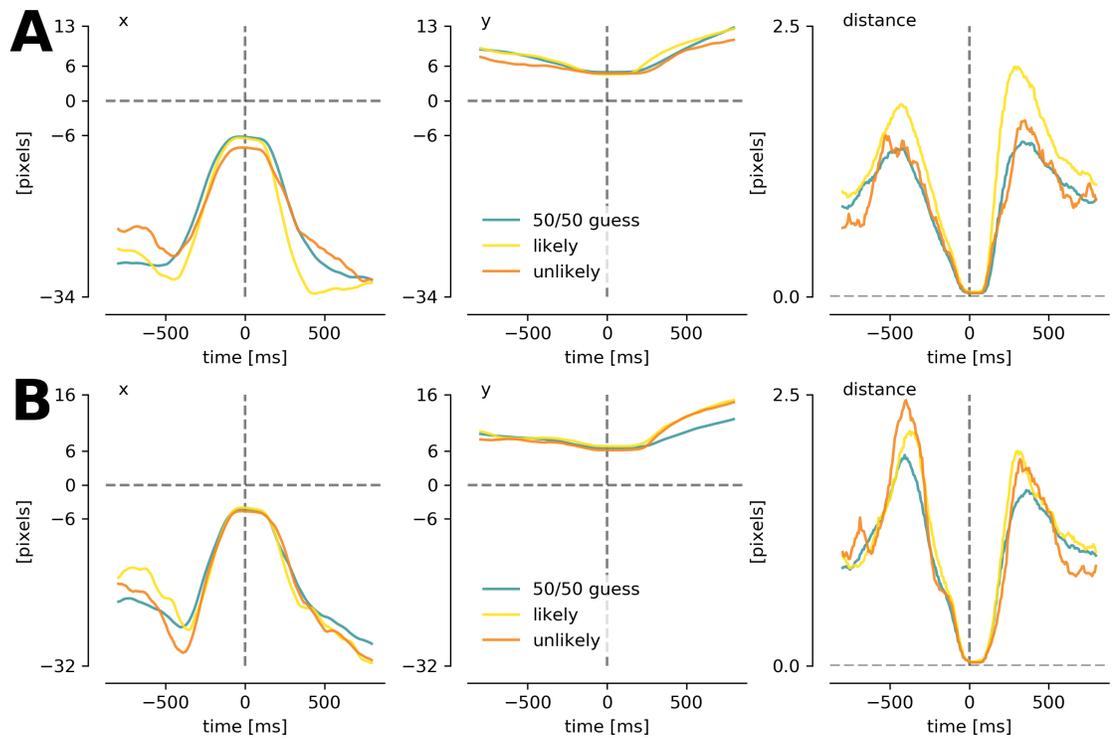


Figure 6.7. Mouse-Tracking averaged data before and after clicking an icon separated by the action predicted by the model for learners and non-learners groups. (A) presents results for the learners group: the leftmost and center panels show the participant's averaged x and y pixel coordinates (y-axis) before and after clicking and icon in the interval -800 to 800 milliseconds (x-axis); the rightmost panels present the average distance (y-axis) covered by the mouse in the same time interval mentioned above. (B) presents the equivalent measures of (A) for non-learners group.

In movements clustered by the most likely observation, we find that cluster condition (exploitation, exploration, and random) was significantly different ($F(2,40)=20.68$, $p<0.001$, $\eta_G^2=0.148$; Mauchly's Test: $W=0.76$, $p=0.07$) accompanied by a triple interaction ($F(2,40)=4.01$, $p=0.026$, $\eta_G^2=0.012$; Mauchly's Test: $W=0.90$, $p=0.4$). For this ANOVA, the dependent variable was not Normally distributed ($W=0.96655$, $p=0.002$). Both results were also confirmed by permutation analysis (cluster condition $p<0.001$, Interaction $p=0.025$). In order to better understand the interactions, we performed a paired post hoc test which only test differences between cluster conditions. Results are presented in Tab. 6.1, showing that in the learners case, exploitation is always different from exploration and random. Also, only before clicking an icon, the random condition is different from exploration. In non-learners case these differences almost disappear, and are only limited to the movements performed before the click between exploitation and exploration

Table 6.1. **Statistics for mouse movements clustered by the most likely observation predicted by the behavioral modeling.** Upper panel presents the ANOVA model details. Permutation ANOVA p-values are presented as Perm p. Effect sizes are show in the final column using Generalized Eta Squared (η_G^2). Lower panel presents Holm-Sidak post hoc p-values for cluster condition grouped by the remaining factor levels. Numbers in black indicate significant values.

Effect	DFn	DFd	SSn	SSd	F	p	Perm p	η_G^2
Group	1	20	1.47	11.77	2.49	0.130	0.114	0.061
Condition	2	40	3.95	3.82	20.68	0.000	0.000	0.148
Distance	1	20	0.01	5.68	0.03	0.860	0.081	0.000
Group:Condition	2	40	0.52	3.82	2.70	0.079	0.917	0.022
Group:Distance	1	20	0.38	5.68	1.34	0.261	0.278	0.016
Condition:Distance	2	40	0.14	1.41	1.94	0.157	0.176	0.006
Group:Condition:Distance	2	40	0.28	1.41	4.01	0.026	0.025	0.012

		Learners		Non-learners	
		Exploitation	Exploration	Exploitation	Exploration
<i>before click</i>	Exploration	0.00017	–	0.0072	–
	Random	0.04398	0.04398	0.6097	0.6097
<i>after click</i>	Exploration	0.00015	–	0.17	–
	Random	0.00094	0.1613	0.87	0.33

conditions. It is important to remark that the effect sizes presented here are rather small, suggesting minor behavioral differences in the distance before/after clicking an icon.

In general terms, learners have clear trends of movements under the most likely choice cluster scheme. Cluster conditions are characterized by progressively faster movements according to the following order: exploration, random and exploitation. The most significant difference is between exploitation and the rest of the conditions, which indicates the high level of specialization achieved by learners during the task when performing mouse movements. This is not surprising, since practice facilitates the execution of the task, as also seen in Fig. 6.1 as a decrease in the reaction times. On the other hand, non-learners cover similar distance between conditions. The lack of difference between random and exploitation suggest that they move more reactively towards the selection. This could be due to the lack of elaborate planning when non-learners made decisions.

Table 6.2. Statistics for mouse movements clustered by behavioral modeling prediction. Upper panel presents the ANOVA model details. Permutation ANOVA p-values are presented as Perm p. Effect sizes are show in the final column using Generalized Eta Squared (η_G^2). Lower panel presents Holm-Sidak post hoc p-values for cluster condition grouped by the remaining factor levels. Numbers in black indicate significant values.

Effect	DFn	DFd	SSn	SSd	F	p	Perm p	η_G^2
Group	1	20	0.50	15.44	0.65	0.428	0.457	0.018
Condition	2	40	1.33	5.022	5.29	0.009	0.025	0.047
Distance	1	20	0.03	4.99	0.13	0.860	0.183	0.001
Group:Condition	2	40	0.46	5.022	1.81	0.079	0.811	0.017
Group:Distance	1	20	0.99	4.99	3.95	0.261	0.043	0.035
Condition:Distance	2	40	0.17	1.431	2.44	0.157	0.133	0.006
Group:Condition:Distance	2	40	0.06	1.431	0.79	0.461	0.411	0.002

		Learners		Non-learners	
		50/50 guess	Likely	50/50 guess	Likely
<i>before click</i>	Likely	0.011	–	1	–
	Unlikely	0.478	0.478	0.47	1
<i>after click</i>	Likely	0.0036	–	1	–
	Unlikely	0.452	0.0494	1	1

In movements clustered by action prediction, we find that cluster condition (likely, unlikely, and 50/50 guesses) was again significant ($F(2,40)=5.29$, $p=0.009$, $\eta_G^2=0.047$; Mauchly's Test: $W=0.81$, $p=0.14$), but this time was not accompanied by any interaction. For this ANOVA, the dependent variables were not Normally distributed ($W=0.96655$, $p=0.002$). Nonetheless, when confirming the results through permutation analysis, we confirm a cluster condition effect, which also present a double interaction between groups and covered distance ($p=0.043$). As it is shown in Tab. 6.2, ANOVA results presented a $p=0.06$ close to that obtained by permutation ($p=0.043$), suggesting that this discrepancy it is most likely explained by lack of statistical power. We performed the same post hoc analyses as before (Tab. 6.2 lower panel), finding differences only for the group of learners between likely and 50/50 guess conditions. Despite these significant results, it is worth noting that the effect sizes are small ($\eta_G^2 < 0.1$), therefore, even if the effects are significant, the differences between conditions should be considered with caution.

Under this clustering scheme, and only for the learners group, the likely choice condition has significant differences compared to unlikely and 50/50 guess conditions. These effects are similar as those previously found for exploitation choices, because, from the point of view of the model, greater certainty in prediction is a consequence of knowledge. For non-learners, once again, no differences were found between conditions. This emphasizes the lack of future planning in non-learners' actions.

6.6. Electrophysiological Results

6.6.1. Checkerboard Probes

Checkerboard Probes were grouped into two groups with different attentional load: the clicked side (attended condition), and the discarded side (unattended condition). Left-side and right-side probes grand averages ERPs are presented in Fig. 6.8 and Fig. 6.9 for learners and non-learners groups respectively. These figures consider regions of interest

(ROIs) encompassing parieto-occipital areas, separated into left (channels P3, P7, PO3, and O1) and right (channels P4, P8, PO4, and O2) hemispheres.

In both groups and for both conditions, the P1 and N1 effects were identified bilaterally over occipito-parietal areas, being more visible over the contralateral hemisphere respect to where the stimulus was presented. Components peaks latency were detected at 65 ms for the P1, and 120 ms for the N1. Also, the effect is more prominent with stimuli presented on the left. No significant differences were found between conditions. This may be

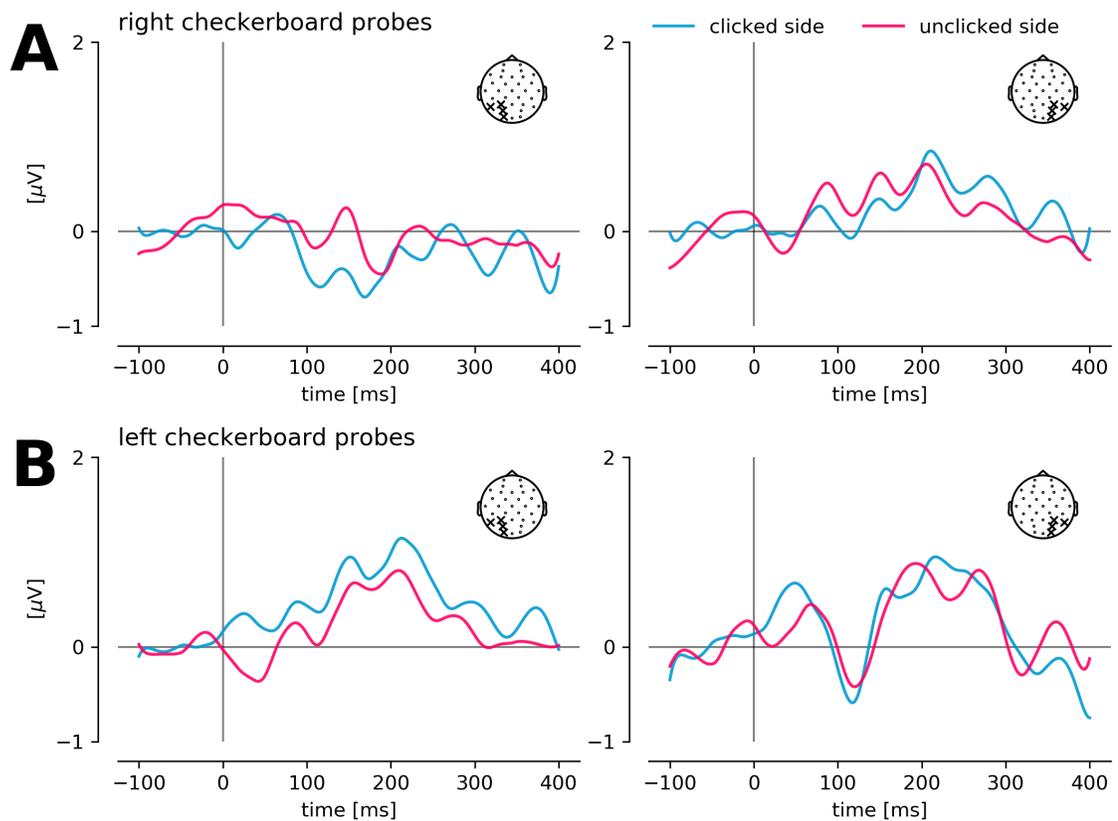


Figure 6.8. Checkerboard probes ERP components at selected ROIs for learners group. Figures (A) and (B) presents the grand averaged ERP waveforms (y-axis) for right and left checkerboard probes respectively, in the -100 to 400 ms time interval (x-axis). Left panels present the average ERP at the ROI defined by electrodes P3, P7, PO3, and O1 (left parieto-occipital region). Right panels present the average ERP at the ROI defined by electrodes P4, P8, PO4, and O2 (right parieto-occipital region).

due by the large amount of noise presented in the averaged ERPs (especially noticeable in right checkerboard probes), independent of the high number of samples obtained per condition (on average around 230 probes). Specifically, this noise can be caused by saccadic eye movements related to the high visual demands of scrutinizing the interface elements. Particularly, icons automatically draw the attention of participants, thus the mapping of the visual cortex could be different between probes with the consequence of less defined visual ERPs.

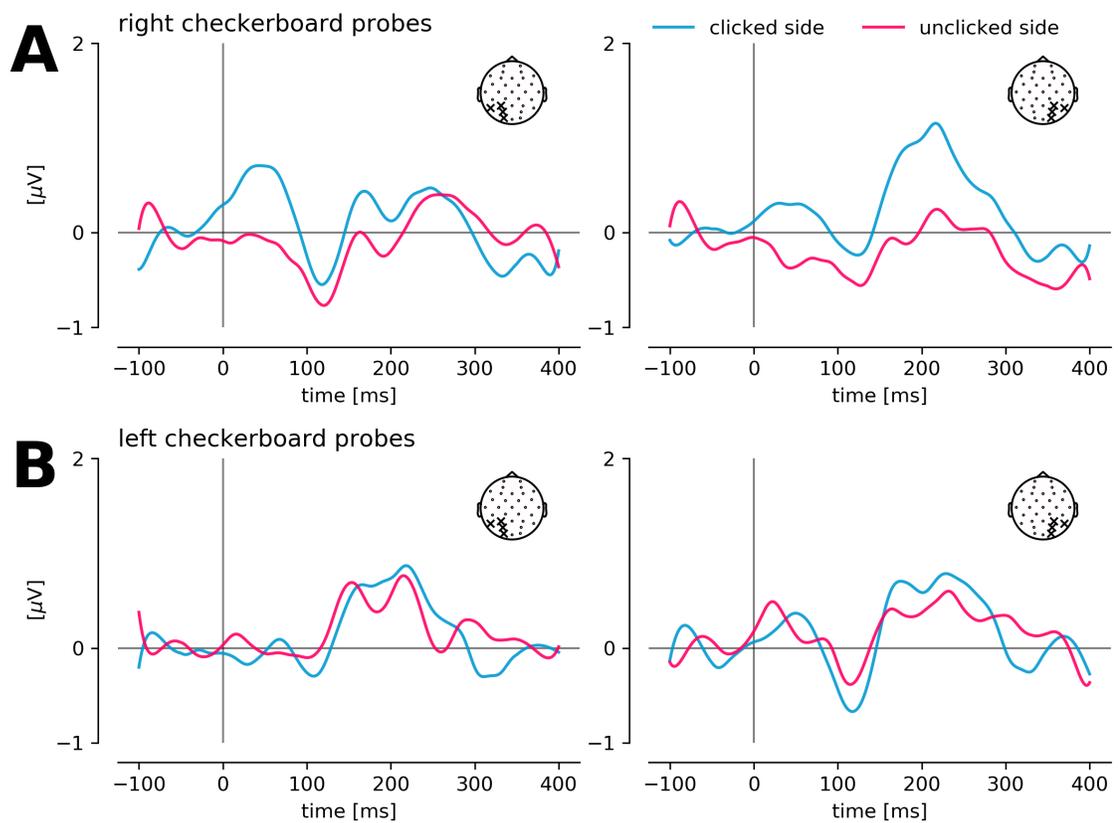


Figure 6.9. Checkerboard probes ERP components at selected ROIs for non-learners group. Figures (A) and (B) presents the grand averaged ERP waveforms (y-axis) for right and left checkerboard probes respectively, in the -100 to 400 ms time interval (x-axis). Left panels present the average ERP at the ROI defined by electrodes P3, P7, PO3, and O1 (left parieto-occipital region). Right panels present the average ERP at the ROI defined by electrodes P4, P8, PO4, and O2 (right parieto-occipital region).

Finally, it cannot be ruled out that the proposed division is not the best to observe attentional differences. For example, a participant may observe more than one side (left or right) before making a decision, resulting in a miss-assignment in the attended/unattended conditions. This is not only the case with the proposed attentional separation, but also with other more complex schemes, for instance, by using the behavioral modeling prediction capabilities. While it is possible to use other techniques of attentional assignment (such as collapsing the probes by its BDT level), in practice, the small number of checkerboard samples does not allow to elaborate into more complex partitions. We believe that these kind of probes could be useful, but it would be necessary to find a better role within the design to take advantage of them.

6.6.2. Binary Choice ERP Components

Not all of the participants' choices generate explicit feedback about the sequence of actions. This happens between levels 1 to 3 of the BDT, where each icon selection leads to a new binary choice without providing any details of accuracy. In the following sections, we analyze the evoked EEG activity of binary choices separated by the most likely observation predicted by the model (section 6.6.2.1), and separated by the action predicted by the model (section 6.6.2.2) independently.

6.6.2.1. Binary Choice Separated by Observation Prediction

Exploitation, exploration, and random observations grand averages ERPs are presented in Fig. 6.10 and Fig. 6.11 for learners and non-learners groups respectively. These figures consider ROIs encompassing parieto-occipital areas, separated into left (electrodes CP5, P3, P7, PO3, and O1) and right (electrodes CP6, P4, P8, PO4, and O2) hemispheres; and a fronto-central area (electrodes AF3, F3, FC1, Fz, FC2, F4, and AF4). In these ROIs, we identify (in temporal order) the ERN, P1, N1, P2, P3, and LPC components. ERPs are time-locked to the response (click) stimulus, therefore the latencies of visual ERPs (such as P1, N1, and P2) are shifted about 60 ms. P1/N1 effects can be seen in the left

and right parieto-occipital ROIs at latencies about 155 ms for P1 and 190 ms for N1, P2 centro-frontal peak observed around 265 ms after the response, and P3 peak at 400 ms (i.e. 340 ms from the next binary choice onset). LPC is observed from 500 ms onwards.

The only significant spatiotemporal difference was found in learners for the LPC component. The exploration condition presented a more positive deflection than exploitation condition, and random condition presented a more positive deflection than both, exploration and exploitation. This led to a significant difference between the three conditions ($p < 0.0397$, cluster permutation test) between 537 to 763 ms after the icon selection (Fig. 6.10A). The topographical distribution of this difference involved the right centro-parietal region, where Pz and CP6 electrodes showed the strongest differences (see Tab. 6.3 for more details). In the case of non-learners, and between learners and non-learners groups, no significant differences were found.

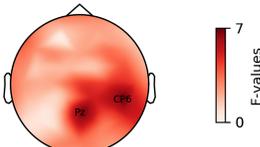
In general terms, there was no early attentional facilitation between conditions in both groups. This is expected, since the selections were not subdivided by side (left/right). It is important to note that the previous subdivision is not necessary, because the early attentional facilitation of the clicked side does not add extra information about the decision-making process.

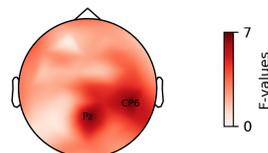
Regarding the significant differences found in learners, these denote how the use of attentional resources changes at different levels of the development of expertise. Random condition, which is infrequent in this group, presented the highest amplitude in late ERP components. Usually, random choices were performed in the early stages of the task. In consequence, participants were subjected to a greater cognitive load in order to understand the rules of the BDT. From there on, the general patterns of behavior of learners can be extracted from Fig. 6.5A rightmost panel. In this figure, exploration and exploitation curves follow complementary paths, going from exploration to exploitation behavior as task unfolds. Moreover, as seen in Fig. 6.3A, exploration was performed using the topological strategy, while exploitation was performed using the discriminative strategy. This suggests that by using the discriminative strategy to solve the task, learners optimize the

use of attentional resources, reducing the cognitive load of conducting a search guided by exploration.

In the case of non-learners no significant differences were found. Interestingly, although no significant, random condition presented a more negative deflection than the other conditions from P3 onwards (Fig. 6.11A). This suggests that exploitation and exploration choices gather more attentional resources than random choices when planning future actions. Even, as the LPC tends to disappear in the random condition, this could indicate little or no engagement with the task. Such differentiation is critical, because in learners the random condition presented significant differences in the opposite direction. As the random condition has a direct relationship with the random strategy, this implies a clear distinction between groups which can also be used as a potential indicator of task performance and strategic use of an interface.

Table 6.3. Summary of significant spatiotemporal statistics for binary choice EEG signals separated by observation prediction. Non-parametric permutation clustering F-test corrected p-values are presented in the last column for $p < 0.05$ clusters. The time-averaged test statistic F-map of all permutations is also presented to facilitate the visualization of the spatial significance of the electrodes. ERP components column shows the ERPs found in the significant time interval of the cluster.

Group	Electrodes	ERP components	Average F-map	Time (ms)	p-value
Learners	Pz, CP6	LPC		537 – 763	0.0397



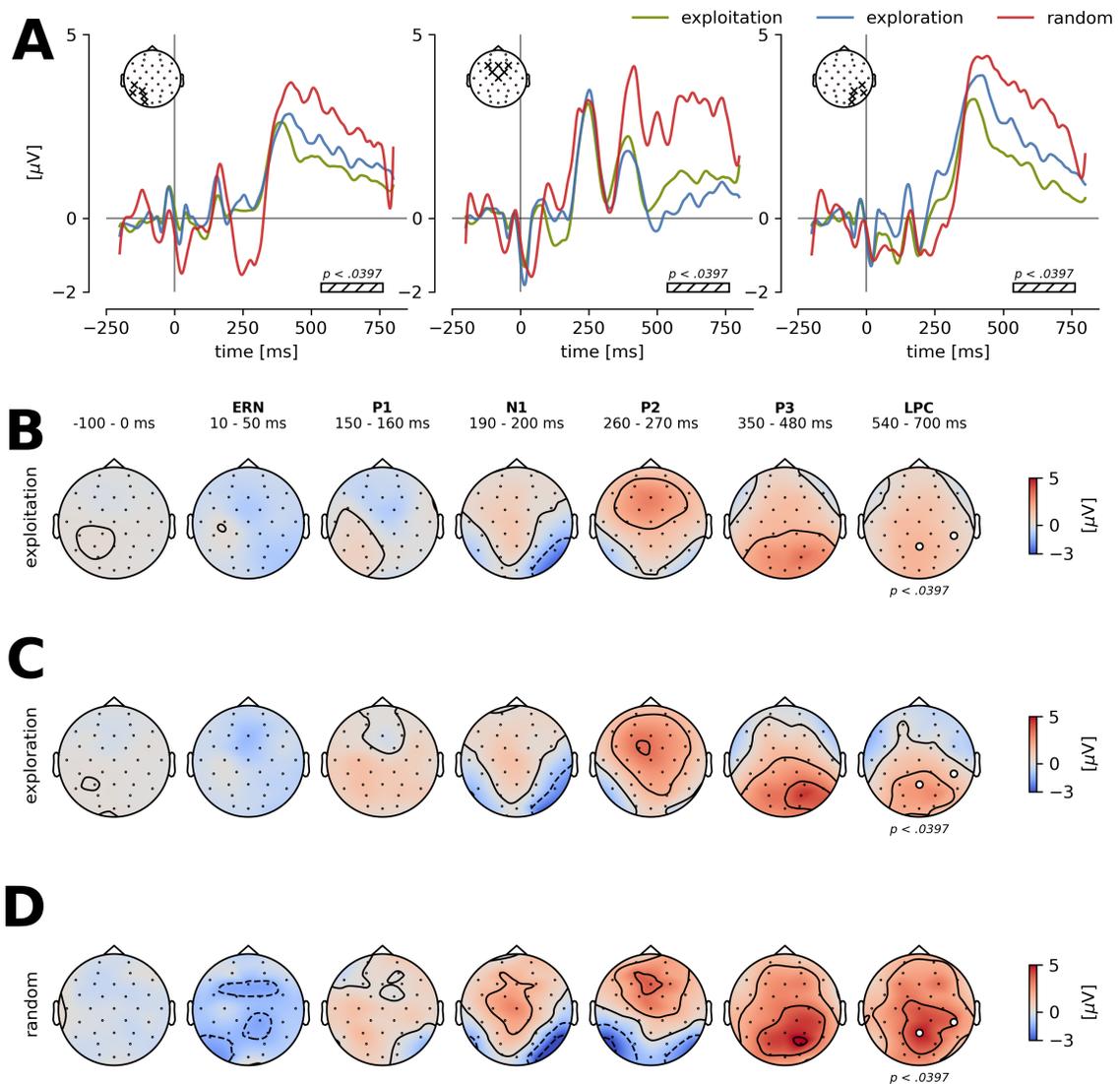


Figure 6.10. **Learners binary choice related ERPs separated by the most likely observation predicted by the model at selected ROIs and its corresponding scalp maps.** Figure (A) presents three ROIs (see text for details) of the grand averaged ERP waveforms (y-axis) for exploitation, exploration, and random conditions displayed from 200 ms prior to response (click) onset and ending 800 ms after (x-axis), baseline corrected using the -200 – 0 ms interval. Figures (B), (C) and (D) presents a scalp maps averaged in the displayed temporal windows for exploitation, exploration, and random conditions respectively. These scalp maps show topologies for baseline, ERN, P1, N1, P2, P3, and LPC components.

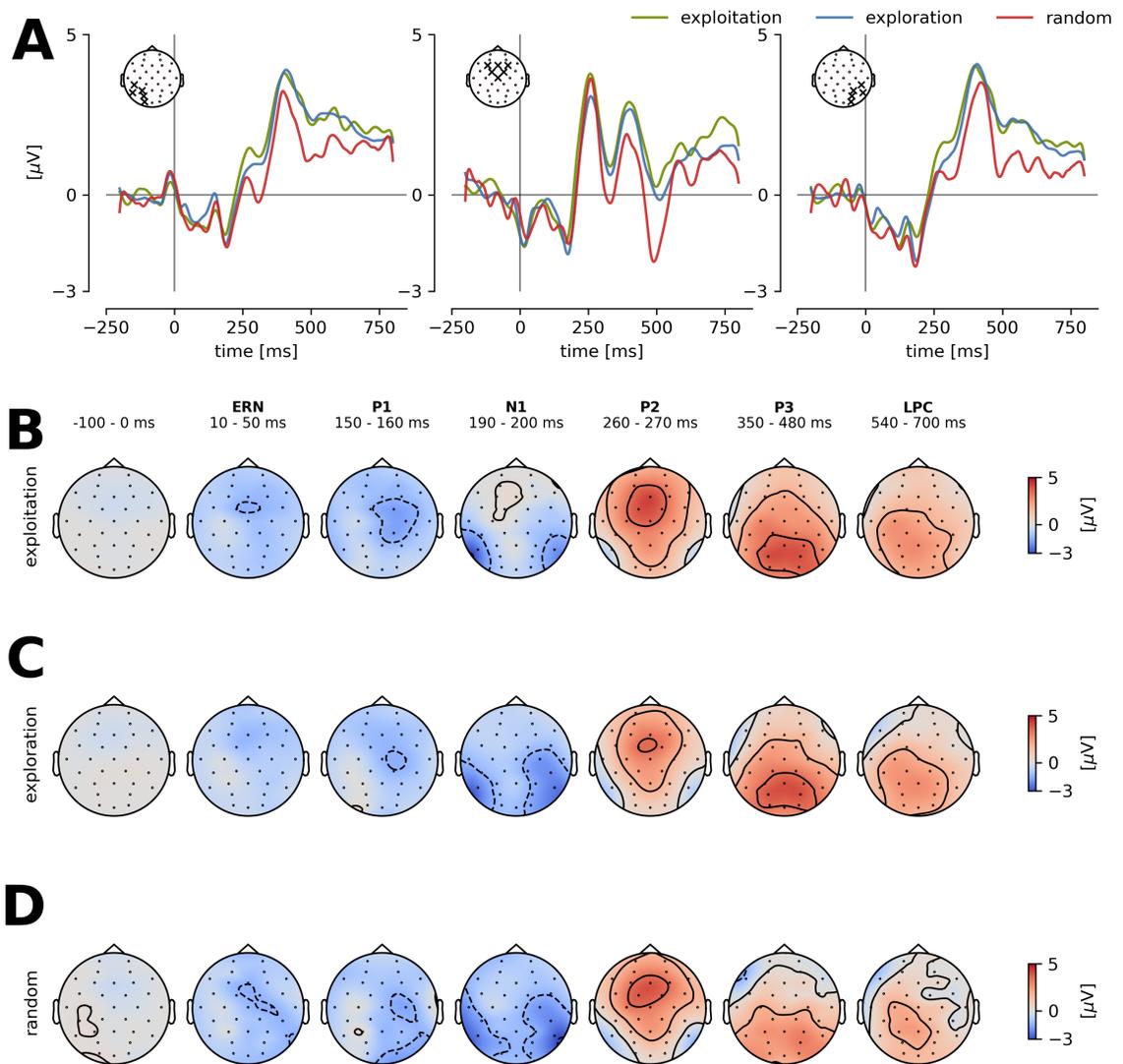


Figure 6.11. **Non-learners binary choice related ERPs separated by the most likely observation predicted by the model at selected ROIs and its corresponding scalp maps.** Figure (A) presents three ROIs (see text for details) of the grand averaged ERP waveforms (y-axis) for exploitation, exploration, and random conditions displayed from 200 ms prior to response (click) onset and ending 800 ms after (x-axis), baseline corrected using the -200 – 0 ms interval. Figures (B), (C) and (D) presents a scalp maps averaged in the displayed temporal windows for exploitation, exploration, and random conditions respectively. These scalp maps show topologies for baseline, ERN, P1, N1, P2, P3, and LPC components.

6.6.2.2. Binary Choice Separated by Action Prediction

Likely, 50/50 guess, and unlikely icon selections grand averages ERPs are presented in Fig. 6.12 and Fig. 6.13 for learners and non-learners groups respectively. These figures consider the same scheme of ROIs defined in the previous section, identifying also the same set of ERP components.

No significant spatiotemporal differences were found for learners and non-learners groups, nor between groups. As in the case of binary choices separated by observation prediction, there was no early attentional facilitation between conditions. However, although no significant, it is interesting to note that in late ERP components (mainly in P3 and more attenuated in LPC), different trends were observed in amplitude: unlikely choice condition presented a more positive deflection than 50/50 guess and likely choice conditions, and 50/50 guess condition presented a more positive deflection than likely choice condition. These trends were similar between groups, and in general terms, these effects may be explained by changes in P3 due to collection of information, and changes in LPC due to the post-retrieval effect of strategic processing.

Specifically, this arrangement in amplitude suggest that unlikely choices gather more attentional resources, which may be due to two causes: a) a random choice in a learning stage; b) a mismatch between response and outcome. Option (a) is a simile to what happens in learners in random choices. However, non-learners' random choices were characterized as the condition with less amplitude (see the previous section), thus the interpretation would be closer to alternative (b). In this sense, the participants thought they had the information to answer correctly (according to a systematic strategy), but they commit an unforced error with an unexpected outcome. This would cause adjustments, such that upon receiving feedback, the wrong choice could be bounded more easily.

In the case of 50/50 guesses (the intermediate amplitude condition) the domain of action is open to both icons of the binary choice. This means that the participant is in the middle of a search process with short-term consequences. If the current selection does

not lead to positive feedback, it does not imply that the other does. Thus, the information presented at selection time is still insufficient to assign the icon a more important role in future planning (e.g. a concept-icon relationship). Importantly, it is worth noting that participants used the topological strategy to perform 50/50 guesses (as detailed in section 5.3.1). As such, this condition behaves similarly to exploration condition seen in the previous section.

Finally, likely choice condition presented the lower amplitude in P3 and LPC components. These choices occur for two reasons: a) because the participant already learned the relationship; or b) because he or she has narrowed the domain of action enough so that there are no more paths to continue the search process. In the case of learners, as seen also in the previous section for exploitation choices, they used the discriminative strategy early on the task. Thus, likely choices in this group were largely performed by using option (a). However, in non-learners, the use of the discriminative strategy occurred later in the task (see Fig. 6.3B left panel). For the most part, they used the topological strategy as their main search tool (option (b)). This suggests that narrowing the domain of action by using topological cues (in order to generate likely choices) also decreases the cognitive load. Moreover, this would indicate that, as non-learners solved instances of the task, implicit knowledge of concept-icon relationships was generated from the choices that reached positive feedback. Ultimately, this may have been the reason why the non-learners used the discriminative strategy towards the end of the task.

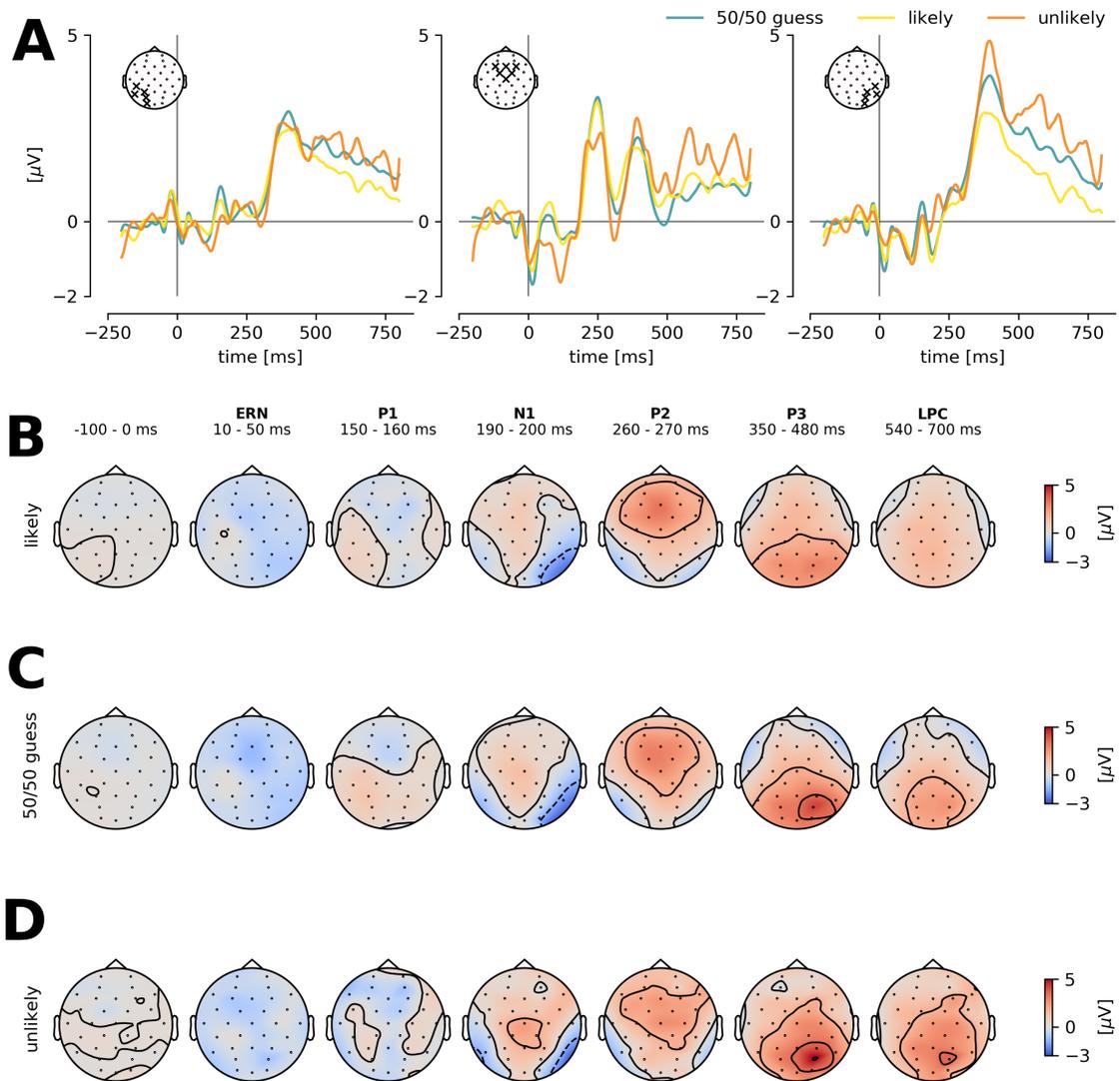


Figure 6.12. **Learners binary choice related ERPs separated by the most likely observation predicted by the model at selected ROIs and its corresponding scalp maps.** Figure (A) presents three ROIs (see text for details) of the grand averaged ERP waveforms (y-axis) for likely, 50/50 guess, and unlikely conditions displayed from 200 ms prior to response (click) onset and ending 800 ms after (x-axis), baseline corrected using the -200 – 0 ms interval. Figures (B), (C) and (D) presents a scalp maps averaged in the displayed temporal windows for likely, 50/50 guess, and unlikely conditions respectively. These scalp maps show topologies for baseline, ERN, P1, N1, P2, P3, and LPC components.

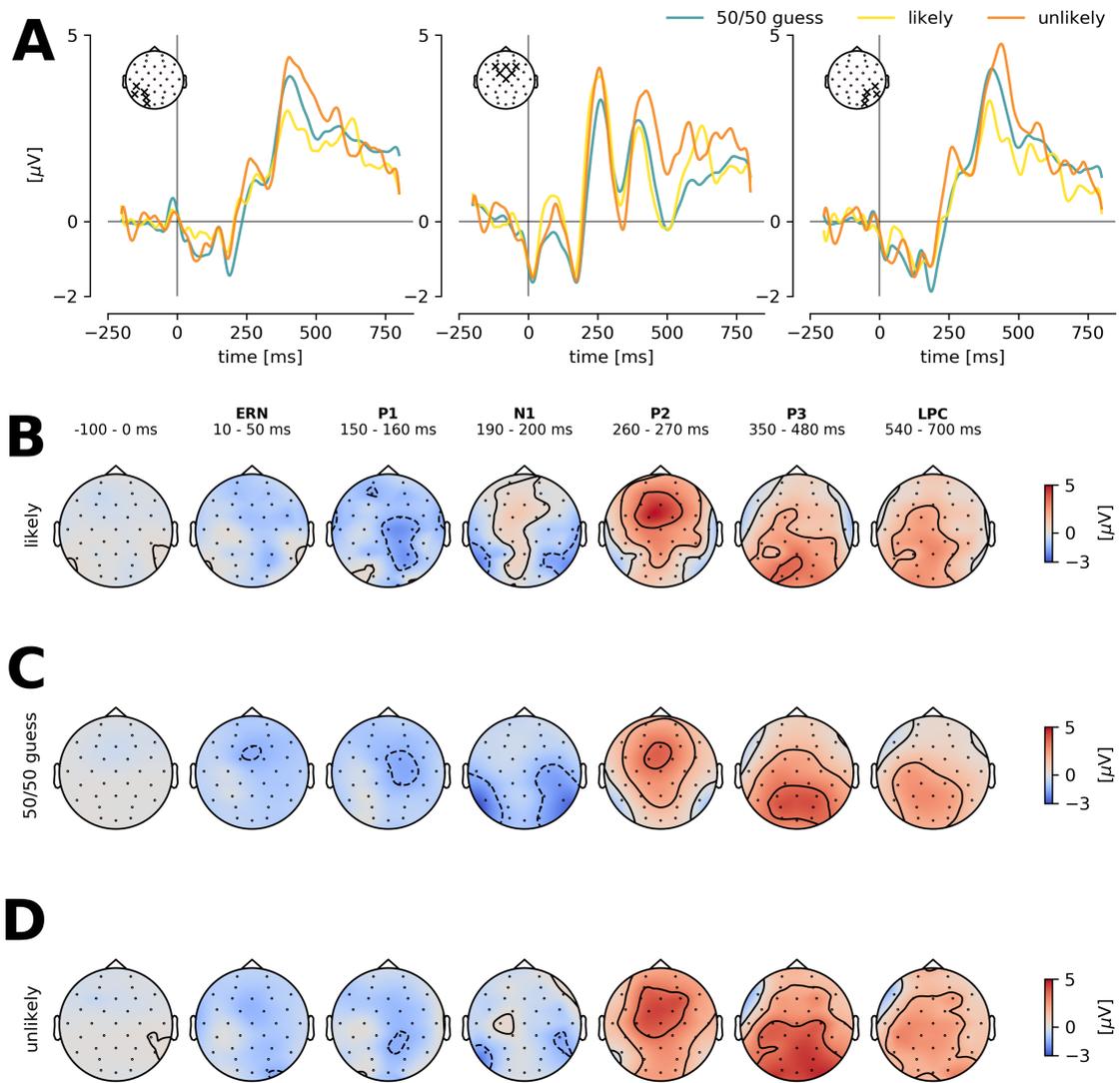


Figure 6.13. **Non-learners binary choice related ERPs separated by the most likely observation predicted by the model at selected ROIs and its corresponding scalp maps.** Figure (A) presents three ROIs (see text for details) of the grand averaged ERP waveforms (y-axis) for likely, 50/50 guess, and unlikely conditions displayed from 200 ms prior to response (click) onset and ending 800 ms after (x-axis), baseline corrected using the -200 – 0 ms interval. Figures (B), (C) and (D) presents a scalp maps averaged in the displayed temporal windows for likely, 50/50 guess, and unlikely conditions respectively. These scalp maps show topologies for baseline, ERN, P1, N1, P2, P3, and LPC components.

6.6.3. Feedback ERP Components

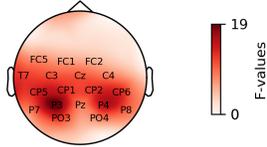
The last choice of the participant generates explicit feedback about whether the path of four selections were right or wrong. Correct and incorrect feedback grand averages ERPs are presented in Fig. 6.14 and Fig. 6.15 for learners and non-learners groups respectively. These figures consider the same scheme of ROIs defined in the previous sections, identifying (in temporal order) the ERN, P2, feedback-related P3, and LPC. ERPs are time-locked to the feedback stimulus, therefore ERN peak is observed around 55 ms (130 ms from the 4 level click), P2 peak is observed around 145 ms, FRN negative-peak is observed around 270 ms, and P3 peak at 330 ms; LPC is observed from 400 ms onwards.

The only significant spatiotemporal difference was found in learners for the FRN, P3, and LPC components. The negative feedback condition presented a more positive deflection, leading to a significant difference between conditions ($p < 0.0002$, cluster permutation test) between 265 to 800 ms after feedback presentation (Fig. 6.14A). The topographical distribution of this difference involved the parieto-occipital region, where P3 and P4 electrodes showed the strongest differences (see Tab. 6.4 for more details). In the case of non-learners, and between learners and non-learners groups, no significant differences were found.

Different groups usually tend to get a specific type of feedback than the other: learners get more positive than negative feedback, and the opposite occurs in non-learners. For learners, positive outcomes start to be expected as they learn the structure of the BDT, therefore positive feedback ceases to be informative and consequently having less salience in the FRN and feedback-related P3. By contrast, negative feedback has information to make adjustments in order to correct wrong choices, as seen in feedback-guided learning. Moreover, this difference is maintained in the LPC, suggesting that negative feedback has a major impact on planning. This may also be due to the fact that learners care about their mistakes and they are committed to solve the task. On the other hand, non-learners do not seem to differentiate between feedbacks. As non-learners use the topological strategy as their main search tool, negative feedback does not provide critical information to guide the

decision-making process. Indeed, they expect that their actions lead to negative feedback, therefore it does not tell them to change their strategy, only to keep trying. Nevertheless, positive feedback seems to evoke, although no significant, higher ERPs amplitudes. Perhaps because positive feedback triggers the signal to interrupt the search (in order to repeat the sequence), it has greater involvement in the decision-making process.

Table 6.4. Summary of significant spatiotemporal statistics for feedback EEG signals. Non-parametric permutation clustering F-test corrected p-values are presented in the last column for $p < 0.05$ clusters. The time-averaged test statistic F-map of all permutations is also presented to facilitate the visualization of the spatial significance of the electrodes. ERP components column shows the ERPs found in the significant time interval of the cluster.

Group	Electrodes	ERP components	Average F-map	Time (ms)	p-value
Learners	FC1, FC5, T7, C3, CP1, CP5, P7, P3, Pz, PO3, PO4, P4, P8, CP6, CP2, C4, FC2, Cz	FRN, P3, LPC		265 – 800	0.0002

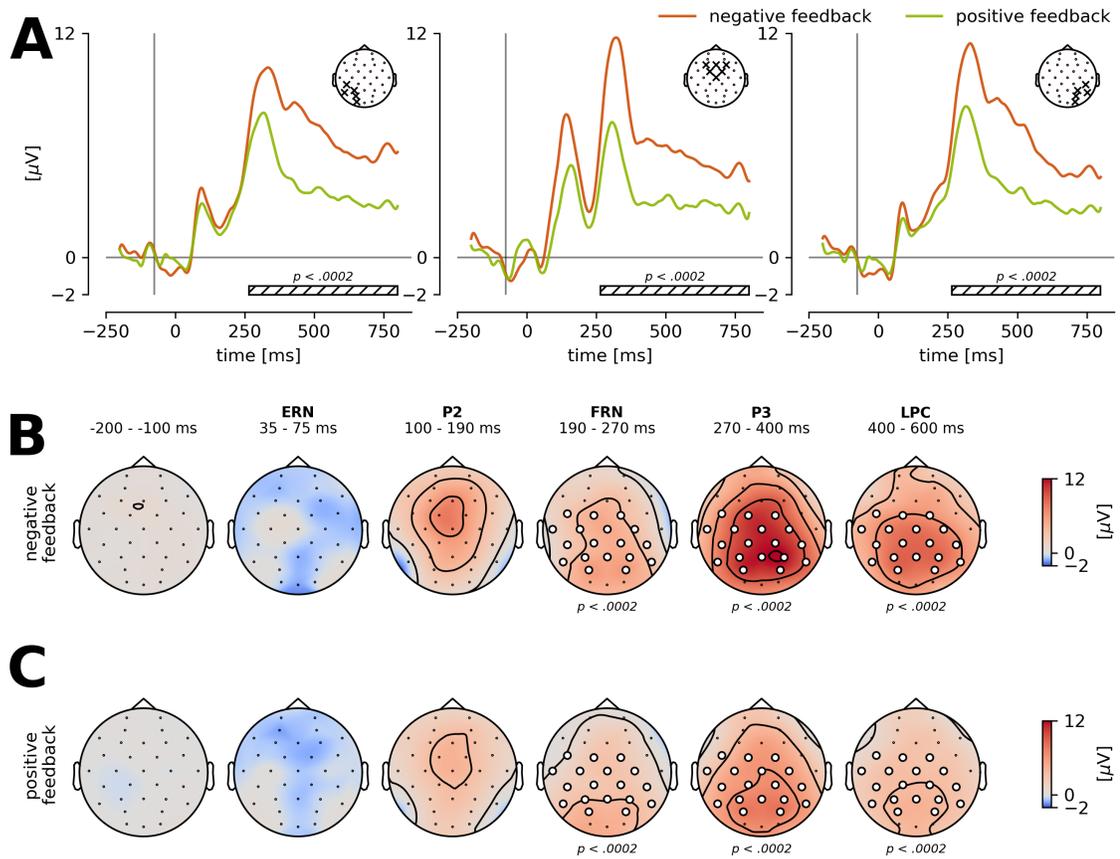


Figure 6.14. **Learners feedback related ERPs at selected ROIs and its corresponding scalp maps.** Figure (A) presents three ROIs (see text for details) of the grand averaged ERP waveforms (y-axis) for positive and negative feedback conditions displayed from 200 ms prior to feedback onset and ending 800 ms after (x-axis), baseline corrected using the -200 – -100 ms interval. Figures (B) and (C) present scalp maps averaged in the displayed temporal windows for positive and negative feedback conditions respectively. These scalp maps show topologies for baseline, ERN, P2, FRN, P3, and LPC components. Spatiotemporal significant differences are presented in (A) for time (hatched boxes), and in the scalp maps for the electrode space (white dots) with their respective corrected cluster p-value.

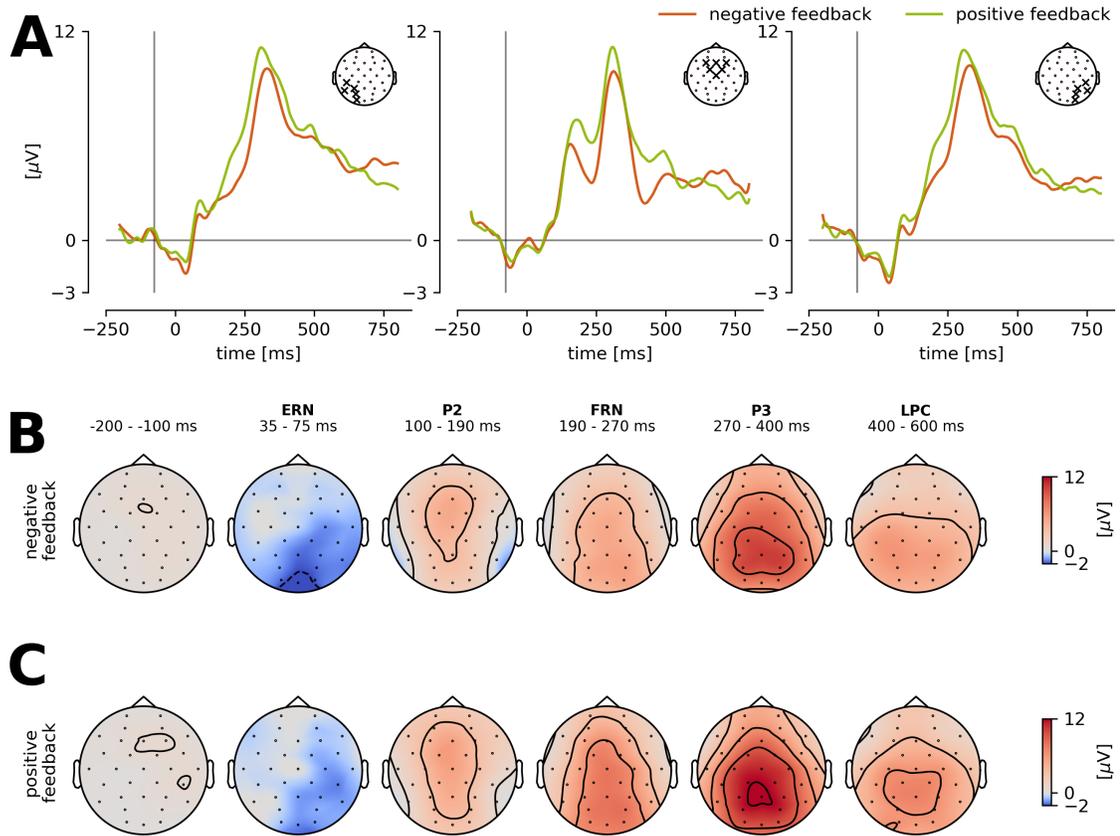


Figure 6.15. **Non-learners feedback related ERPs at selected ROIs and its corresponding scalp maps.** Figure (A) presents three ROIs (see text for details) of the grand averaged ERP waveforms (y-axis) for positive and negative feedback conditions displayed from 200 ms prior to feedback onset and ending 800 ms after (x-axis), baseline corrected using the -200 – -100 ms interval. Figures (B) and (C) present scalp maps averaged in the displayed temporal windows for positive and negative feedback conditions respectively. These scalp maps show topologies for baseline, ERN, P2, FRN, P3, and LPC components.

7. DISCUSSION

We have presented here a modeling framework which, on the basis of stereotyped, well-defined search strategies, is able to track the development of expertise in a sequential decision-making task. We have tested the framework using a novel BDT task that has the potential to provide a well-defined computer interface scenario for the study of expertise acquisition in real world situations. By following individual learning processes, when confronted with users that are able to learn the task, our model quickly reaches a point where it can predict the users' most likely next choice. This suggest that the framework presented here could play a role in the development of adaptive interfaces where sequential choices are required to reach a desired outcome.

We organize the following discussion around four main issues: the modeling framework choice and its consequences for both the tracking of the learning process and for adaptive interface design; the structure of the sequential decision making task; the criteria for temporal and spatial information encoding; the electrophysiological analysis and the issue of individual differences among users.

7.1. Modeling Framework

As outlined in the introduction, a host of approaches exist to deal with the problem of modeling decision making behavior in general (Sukthankar, Geib, Bui, Pynadath, & Goldman, 2014), and sequential decision making in particular (M. M. Walsh & Anderson, 2014). Here we have chosen to follow an HMM-based approach that, in line with type-based methods, relies on a set of pre-defined strategies to model a potentially infinite set of behaviors (Albrecht et al., 2016). This approach reduces the dimensionality of the problem while at the same time aims to provide some insight into the learning and expertise acquisition process.

While the approach presented here is indeed inspired by the HMM structure, we do not use traditional algorithms to determine the parameters of the HMM (e.g. BaumWelch, see (Rabiner, 1989) for more details). This is because, in our task, emission and transition probabilities are dynamic consequences of the participant's actions. Therefore, we can ask the strategies for the emission probabilities, because they directly operationalize possible outcomes (as a consequence of their formal definition). Transition probabilities are, in contrast, a more open problem because they define how the HMM is connected (i.e. its structure), which can itself change as learning happens. To deal with this issue, we define transition probabilities in terms of the comparative changes of emission probabilities between strategies. This is because, as discussed above, strategies formalize actual domains of action so that their comparison allows us to determine towards which strategy it is better to transit in order to best explain the current choice made by the participant. Finally, in the case of weight optimization, we use a formulation which is equivalent to finding the Viterbi path in a HMM (also described in (Rabiner, 1989)).

As our results show, the modeling approach presented here was able to consistently follow the performance of individual users as they became familiarized with the task. In those that were able to discover the underlying icon-concept mapping and fully solve the task (learners), the model was able to match their performance as early as the fourth instance of the task. More importantly for the question of how learning proceeds in a sequential decision making situation, the use of pre-defined strategies suggests that learners and non-learners are likely to use different approaches to the problem: while learners dwell on topological, brute-search behavior only during a short while and then rapidly start using concept-mapping knowledge (mostly in discriminative terms), non-learners persist in topological searches and only well into the task do they start showing behavior of consistent icon-concept mappings.

However, to the extent that the pre-defined strategies represent one possible set of high-level behaviors among many, our approach cannot provide definite evidence that users actually use such strategies. In this sense, it may well be the case that a different modeling

approach (or a different subset of strategies) might capture equally well the user's behavior. This is, of course, a pervasive problem in behavioral modeling when users are free to tackle the problem at hand in whatever way they choose. Approaches that combine modeling with subjective reports such as (Mariano et al., 2015) are therefore an interesting extension and worth exploring further. Indeed, Beta Process Hidden Markov Models (BP-HMM) (Fox, Jordan, Sudderth, & Willsky, 2009) such as the one used by (Mariano et al., 2015), consider libraries of states. Here, one pattern of states, represented by an HMM, is active at current time. This approach captures the idea of multiple patterns of HMM structures, but the interpretation of such patterns in terms of the user's behavior is necessarily an exercise that has to be done a posteriori. Nevertheless, the pre-defined strategies that we chose here are informed by the structure of the problem and represent increasingly efficient ways to solve it, thus representing a valid alternative in the context of type-based approaches while being more sensitive to potential differences in the way users approach the task. Other modeling techniques, such as Interactive Partially Observable Markov Decision Processes (I-POMDP) (Gmytrasiewicz & Doshi, 2005), also rely on a similar approach by using predefined intuitions to solve the task (i.e. types of behaviors/strategies), thus allowing to capture the learning of users more efficiently and faster.

A final point regarding the use of a subset of high-level behaviors is worth noting: for instance, as shown in Fig. 6.5 it could be argued that considering the discriminative strategy is enough to yield comparable results in terms of tracking the behavior of those who learn the task. This is expected given that by the fourth iteration, users should have been exposed to most of the discriminative knowledge necessary to fully map the problem. Yet, if we chose this approach, we would have no insight into potentially relevant strategies that unfold throughout the learning processes.

In addition to the previous considerations, an a posteriori validation of the approach is its capacity to predict the behavior of individual users. When it comes to studying the agent's interaction with a user, prediction inference is usually performed over the users' goals (Oh, Meneguzzi, & Sycara, 2011; Ramírez & Geffner, 2011). Such models identify

the goal and the necessary output actions to fulfill that specific goal. In contrast, we specify the goal and our inference is about the way in which the user perform the actions in a search process. The strategies defined here represent policies of actions that could be optimal during different stages of learning. In this sense, prediction performance is not in and by itself, the primary measure of the model's capacity. For instance, even with poor prediction performance (e.g. Fig. 6.5B center panel), the behavioral model can follow the actual performance of the participants based on the exploitation score. In addition, the number of 50/50 guesses can be considered an estimation of how good the prediction could be. As such, this approach (and extensions of it in terms, for instance, of a different set of strategies) could represent a viable approach to inform the construction of flexible interfaces that can adapt to the different moments of the learning curve of their users.

7.2. Task Structure

There are several types of sequential tasks studied in the literature (Ruh N. & Mareschal, 2010; Diuk et al., 2013; Friedel et al., 2014; Huys et al., 2015). In this context, decision trees present an interesting scenario, because they naturally embody the most basic structures of a sequential decision-making situation (W.-T. T. Fu & Anderson, 2006) while allowing for a clear description of the task structure. Here we chose a BDT design to instantiates a simple case of sequential action in Human-Computer Interactions (HCI).

An important aspect of our BDT design is the feedback structure. Specifically, participants have to reach the end of each branch before obtaining information about the appropriateness of previous decisions. In this kind of limited feedback scenario, a *credit assignment problem* appears (W.-T. Fu & Anderson, 2008; M. Walsh & Anderson, 2011; M. M. Walsh & Anderson, 2014). This means that participants have to learn how to use the consequences of their actions to assign value to different parts of the sequential choice. In our task, participants face two main types of credit assignment problems: on the one hand, negative feedback is not enough in and by itself to discover at which point of the BDT wrong decisions were made. On the other hand, positive feedback can only be used by the

participant to generate specific icons-concepts mappings through successive exposures to different instances of the task.

One could argue that immediate feedback interactions (such as interfaces characterized by labeled icons or explicit icon-concept mappings) is the dominant mode of HCI interaction and should therefore be the target for behavioral modeling. This would have the advantage of sidestepping the credit assignment problem. However, limited feedback scenarios such as the one used here have the advantage of making the task more difficult (M. Walsh & Anderson, 2011), therefore revealing the learning process more clearly. This has obvious advantages in terms of making the process of expertise acquisition by different users explicit and thus available for modeling efforts. Moreover, given the nature of our behavioral model, which deals with relationships among strategies, immediate feedback scenarios could be considered as a particular case for the framework. When interfaces are built on the basis of labeled icons, only two strategies are required: topological and discriminative: the topological strategy is necessary to map the spatial arrangement throughout the interface; the discriminative strategy, on the other hand, allows one to relate complementary or neighboring icons to future (alternative) task instances. The generative strategy, in contrast, is unnecessary because the primary icon-concept mapping is explicit from the start.

7.3. Temporal and Spatial Encoding Criteria

Related to the credit assignment problem, another important issue arises when modeling behavior, especially in sequential decision making situations. This issue, which has received much less attention from modeling studies, pertains the criteria that one sets in order to determine which observations count as relevant information (Behrens, Woolrich, Walton, & Rushworth, 2007). Such criteria can be of temporal nature (for instance, how much of past experience we consider when planning future actions) or spatial (for example, how much of the BDT's structure is remembered and used for making decisions).

In our work, the function-parameter τ determines the influence of past outcomes on the behavioral modeling. Here we have chosen a continuous Gaussian kernel (Eq. 5.10) with a peak at the current time. This choice aims to capture underlying short term memory processes whereby current items have a higher probability of influencing behavior than those encountered previously (Cowan, 2008). Whether by interference of novel task information or due to temporal decay, current information cannot not persist indefinitely and therefore it is necessary to consider the dynamics of its causal influence on ongoing decision making (Barrouillet, De Paepe, & Langerock, 2012; Lewandowsky & Oberauer, 2009). The choice of the Gaussian kernel could be a target for improvement, eventually considering the possibility of adapting it to individual users' profiles. Nevertheless, similar exponential metrics, including Gaussian kernels have been used to model complex memory processes and therefore represents a viable first choice (Brown, Neath, & Chater, 2007).

In addition to the temporal problem, a spatial navigation problem arises that is related to the localization of feedback (Madl, Chen, Montaldi, & Trapp, 2015). To localize the correct path, as well as to propagate credit to previous choices, it is necessary for the model to encode some type of memory of the user's actions (W.-T. Fu & Anderson, 2008). In our work, each strategy has two types of rules: a global memory associated to learning the overall BDT structure according to that particular strategy across task instances, and a local memory, which is associated with narrowing the domain of action leading to correct feedback in the current instance of the task. Importantly, such encoding does not forget landmarks that have been encountered by the participant and in this sense does not represent actual memory processes. However, this is in principle not necessary because we are dealing with a highly restricted situation in which a full mapping of the task can happen within one experimental session. Indeed, we use a task repetition criterion to ensure learning and make forgetting more difficult (Corazzini, Thinus-Blanc, Nesa, Geminiani, & Peruch, 2008). Eventually, when dealing with more complex search scenarios, taking into account that the user might forget a previously encountered location might be necessary (Baumann, Skilleter, & Mattingley, 2011; Liverence & Scholl, 2015). Nevertheless, we

account for apparent reductions in the user's knowledge of the BDT structure using the random model to penalize actions that, according to the model's encoding of the user's previous actions, represent mistakes in landmark choices.

7.4. Electrophysiological Analysis

Classical electrophysiological analysis can be framed in the context of Behavioral and Cognitive sciences. In Behavioral Sciences, a common way to study the concept of iteration is the immediate reward/feedback trial experimental design (Pool, Brosch, Delplanque, & Sander, 2014; Chica, Martín-Arévalo, Botta, & Lupiáñez, 2014). It usually takes place in a laboratory environment and the participant performs a sequence of trials, each followed immediately by a reward or some kind of feedback. Even when no feedback is given to the experimental subject, behavior is measured and interpreted mostly independently in each trial. This paradigm is extensively used in Cognitive Neuroscience, which is strongly influenced by the attentional orienting paradigm (Chica et al., 2014). In such studies, cue guides attention towards the target stimulus, an action is usually performed and then there is an immediate consequence of the decision as an explicit or implicit feedback.

However, when the iterations are recurrent over time, there are changes in performance that are not taken into account until the process is automatized with a consequent change in attentional demand (Goh, Gordon, Sullivan, & Winstein, 2014; Akizuki & Ohashi, 2013). Moreover, in more ecological environments we have to perform n -actions to actually get a feedback or reward, or to make clear what we intend to do. Formalizing more powerful indicators of decision-making processes, as the behavioral modeling used here, could improve the understanding of the dynamics of electrophysiological data. Being able to identify the different stages of the learning process could enhance, for instance, the separation of ERP averaging trials into more complex conditions or sub-conditions (as seen in the Electrophysiological Results section). This would make possible to correlate single

trial analysis of EEG records with behavioral indicators that cannot be formalized from classical indicators, among others.

Such aspects are also relevant to create efficient Brain-Computer Interface (BCI) designs (Nicolas-Alonso & Gomez-Gil, 2012; Ahn et al., 2014). An BCI that is aware of the electrophysiological distinctions, either between conditions or at different stages of the development of expertise, could potentially use that information to improve the user's experience or deploy more options in ongoing interactions. In this work we show clear ERP markers that are related to the decision-making process, such as the P300 and LPC effects. P3 ERP component is widely used in BCI designs (Powers, Bieliaieva, Wu, & Nam, 2015; Jeon & Shin, 2015). However, it is not used as a descriptor of learning and expertise development, but rather as a way to distinguish the saliency of different objects. Here, P3 effect not only describes differences between learners and non-learners, but also changes in the automation of the learning process, in the use of different strategies, and in how feedback is evaluated by the participant.

7.5. Individual Differences among Users

One of our main results, which is related to the structure of the task, is the notorious difference among participants in terms of performance. Beyond the evident interest this has in terms of underlying psychological and brain mechanisms, from our perspective, this represents an opportunity to contrast the performance of participants of different skill levels when dealing with search problems (Sacchi & Burigo, 2008). Our task reveals that a standard group of participant is not uniform and can be separated in learners and non-learners, as a first level of analysis. These two groups differ not only in terms of whether they manage to complete the task, but also in terms of which strategies they prefer.

Such differentiation is relevant when we want to distinguish experts from non-experts. When dealing with a problem-solving situation, experts and non-experts differ on a number of dimensions. For instance, they differ in how they search for information (Hershler

& Hochstein, 2009; Barrick & Spilker, 2003), make decisions (Gorman, Abernethy, & Farrow, 2015; Connors, Burns, & Campitelli, 2011) or pay attention (Schriver, Morrow, Wickens, & Talleur, 2008), among others. Taking into account these differences is a critical step towards a more natural and truly adaptive HCI, as the user's needs could be focused more accurately according to their level of expertise. For example, being able to identify the strategy that is currently being deployed could enable the interface to display contents accordingly. Likewise, it could be in principle possible to speed the learning processes with unfamiliar interfaces by showing specific types of interface-hints depending on the user's history of interactions. In the case of non-experts, one could provide contextual cues to facilitate the development or discovery of strategies that have proven successful for experts. Alternatively, because the model is capable of detecting quite robustly when a user is not learning (i.e. when the random strategy dominates), the interface could choose to display challenging options or hints in order to "wake up" the user in such cases.

It is notorious that individual differences in behavior have rarely been the target for computational modeling or, in some cases, even treated as a nuisance (Karwowski & Cuevas, 2003). This is all the more surprising given the importance that recognizing individual differences can have for such disparate but fundamental domains as educational interventions (Detterman & Thompson, 1997; Melby-Lervag, Lyster, & Hulme, 2012; Phillips & Lowenstein, 2011) or human performance studies (Goel, Basner, Rao, & Dinges, 2013; Van Dongen, 2006; Parasuraman & Jiang, 2012). We surmise that the framework presented here could be used to tackle this issue because it is built upon modular strategies whose combination can capture a diversity of behaviors, even among learners (see for instance Fig. 6.4 A vs. B in which two different types of learners can be clearly distinguished). We believe that much more research is needed on the issue of modeling behavior in a way that takes into account the specifics of individuals, in addition to average cases or proof of concept approaches (Smith, Henning, Wade, & Fisher, 2014).

8. CONCLUSIONS AND FUTURE OUTLOOK

The work in this thesis was built upon Hidden Markov Models traditions in the context of decision-making, but extends them in two important directions: On the one hand, by using a small subset of stereotyped search strategies as the hidden states, our model is capable of adapting to the participant's actions as they are deployed online. In contrast to classical training-based approaches, this allows us to dynamically follow—and eventually predict—behavior as it unfolds. On the other hand, we use this approach to tackle a sequential decision-making scenario with limited feedback that represents a plausible model of standard human-computer interactions. Taken together, these two aspects support the idea that our framework could serve as a novel basis for the development of computational interfaces that dynamically react to the behavior of individual users.

Furthermore, this framework was able to capture at electrophysiological level the complex structure of the decision-making process, validating the behavioral modeling as a representative model of the participants' behavior. This aspect facilitates the study of more environmental-friendly experimental settings, which considers both learning and expertise development as crucial factors to understand the ongoing dynamics of the brain processes.

In addition to the presented results, we expect to develop the current work along several lines and questions. For instance, could this approach be used to integrate single-trial EEG data with the behavioral modeling? And could this help to improve user experience or speed learning processes with unfamiliar interfaces? We would also expect to model more complex decision-making environments could be into this framework. Ideally, one could take into account other indexes of behavior such as mouse movements, eye-movements and ongoing EEG to reveal more complex behavioral strategies.

REFERENCES

- Abbeel, P., & Ng, A. Y. (2004). Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the twenty-first international conference on machine learning* (p. 1).
- Acuña, D. E., & Schrater, P. (2010, 12). Structure learning in human sequential decision-making. *PLoS Comput Biol*, *6*(12), e1001003.
- Ahn, M., Lee, M., Choi, J., & Jun, S. C. (2014, 08). A review of brain-computer interface games and an opinion survey from researchers, developers and users. *Sensors (Basel, Switzerland)*, *14*(8), 14601–14633.
- Akizuki, K., & Ohashi, Y. (2013). Changes in practice schedule and functional task difficulty: a study using the probe reaction time technique. *J Phys Ther Sci*, *25*(7), 827-31.
- Akyurek, E. G., & Schubo, A. (2013, May). Electrophysiological correlates of early attentional feature selection and distractor filtering. *Biol Psychol*, *93*(2), 269–278.
- Alagoz, O., Hsu, H., Schaefer, A. J., & Roberts, M. S. (2009). Markov decision processes: a tool for sequential decision making under uncertainty. *Medical Decision Making*.
- Albrecht, S. V., Crandall, J. W., & Ramamoorthy, S. (2016). Belief and truth in hypothesised behaviours. *Artificial Intelligence*, *235*, 63–94.
- Anllo-Vento, L., & Hillyard, S. A. (1996, Feb). Selective attention to the color and direction of moving stimuli: electrophysiological correlates of hierarchical feature selection. *Percept Psychophys*, *58*(2), 191–206.
- Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior research methods*, *37*(3), 379–384.
- Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329–349.
- Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2011). Bayesian theory of mind: Modeling joint belief-desire attribution. In *Cogsci* (Vol. 11, pp. 2469–2474).

- Barrick, J. A., & Spilker, B. C. (2003). The relations between knowledge, search strategy, and performance in unaided and aided information search. *Organizational Behavior and Human Decision Processes*, 90(1), 1 - 18.
- Barrouillet, P., De Paepe, A., & Langerock, N. (2012, Feb). Time causes forgetting from working memory. *Psychon Bull Rev*, 19(1), 87–92.
- Baumann, O., Skilleter, A. J., & Mattingley, J. B. (2011, 05). Short-term memory maintenance of object locations during active navigation: Which working memory subsystem is essential? *PLoS ONE*, 6(5), e19707.
- Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007, 09). Learning the value of information in an uncertain world. *Nat Neurosci*, 10(9), 1214–1221.
- Botvinick, M., & Weinstein, A. (2014). Model-based hierarchical reinforcement learning and human action control. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 369(1655).
- Botvinick, M. M. (2012). Hierarchical reinforcement learning and decision making. *Current Opinion in Neurobiology*, 22(6), 956 - 962. (Decision making)
- Brown, G. D. A., Neath, I., & Chater, N. (2007, Jul). A temporal ratio model of memory. *Psychol Rev*, 114(3), 539–576.
- Bruner, J. S. (1973). Organization of early skilled action. *Child development*, 1–11.
- Chica, A. B., Martín-Arévalo, E., Botta, F., & Lupiáñez, J. (2014). The spatial orienting paradigm: How to design and interpret spatial attention experiments. *Neuroscience & Biobehavioral Reviews*, 40(0), 35 - 51.
- Coles, M. G. H., & Rugg, M. D. (1996). Event-related brain potentials: an introduction. In M. D. Rugg (Ed.), *Electrophysiology of mind—event-related brain potentials and cognition*. Oxford University Press.
- Connors, M. H., Burns, B. D., & Campitelli, G. (2011). Expertise in complex decision making: The role of search in chess 70 years after de groot. *Cognitive Science*, 35(8), 1567–1579.
- Corazzini, L. L., Thinus-Blanc, C., Nesa, M.-P., Geminiani, G. C., & Peruch, P. (2008,

- Oct). Differentiated forgetting rates of spatial knowledge in humans in the absence of repeated testing. *Memory*, *16*(7), 678–688.
- Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Prog Brain Res*, *169*, 323–338.
- Curran, T., Schacter, D. L., Johnson, M. K., & Spinks, R. (2001). Brain potentials reflect behavioral differences in true and false recognition. *Journal of Cognitive Neuroscience*, *13*, 201–216.
- Cushman, F., & Morris, A. (2015). Habitual control of goal selection in humans. *Proceedings of the National Academy of Sciences*, *112*(45), 13817–13822.
- Daw, N. D. (2013). Advanced reinforcement learning. *Neuroeconomics: Decision-Making and the Brain*, 299–320.
- Dayan, P., & Niv, Y. (2008, April). Reinforcement Learning: The Good, The Bad and The Ugly. *Current Opinion in Neurobiology*, *18*(2), 185–196.
- Detterman, D. K., & Thompson, L. A. (1997, Oct). What is so special about special education? *Am Psychol*, *52*(10), 1082–1090.
- Dezfouli, A., & Balleine, B. W. (2013, 12). Actions, action sequences and habits: Evidence that goal-directed and habitual action control are hierarchically organized. *PLoS Comput Biol*, *9*(12), e1003364.
- Diuk, C., Schapiro, A., Crdova, N., Ribas-Fernandes, J., Niv, Y., & Botvinick, M. (2013). Divide and conquer: Hierarchical reinforcement learning and task decomposition in humans. In G. Baldassarre & M. Mirolli (Eds.), *Computational and robotic models of the hierarchical organization of behavior* (p. 271–291). Springer Berlin Heidelberg.
- Donchin, E. (1981). Surprise! surprise? *Psychophysiology*, *18*(5), 493–513.
- Donchin, E., & Coles, M. (1988, 9). Is the p300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*, 357–374.
- Doya, K., & Samejima, K. (2002). Multiple model-based reinforcement learning. *Neural Computation*, *14*, 1347–1369.
- Duffin, E., Bland, A. R., Schaefer, A., & De Kamps, M. (2014). Differential effects

- of reward and punishment in decision making under uncertainty: a computational study. *Frontiers in Neuroscience*, 8(30).
- Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological review*, 87(6), 477.
- Fox, E., Jordan, M. I., Sudderth, E. B., & Willsky, A. S. (2009). Sharing features among dynamical systems with beta processes. In *Advances in neural information processing systems* (pp. 549–557).
- Friedel, E., Koch, S. P., Wendt, J., Heinz, A., Deserno, L., & Schlagenhauf, F. (2014). Devaluation and sequential decisions: linking goal-directed and model-based behaviour. *Frontiers in Human Neuroscience*, 8(587). doi: 10.3389/fnhum.2014.00587
- Friedel, E., Koch, S. P., Wendt, J., Heinz, A., Deserno, L., & Schlagenhauf, F. (2015). Devaluation and sequential decisions: linking goal-directed and model-based behavior. *Habits: plasticity, learning and freedom*.
- Friedman, D., & Johnson, R. (2000). Event-related potential (erp) studies of memory encoding and retrieval: a selective review. *Microscopy research and technique*, 51(1), 6–28.
- Fu, W.-T., & Anderson, J. (2008, May 19). Solving the credit assignment problem: explicit and implicit learning of action sequences with probabilistic outcomes. *Psychological Research*, 72(3), 321–330.
- Fu, W.-T. T., & Anderson, J. R. (2006, May). From recurrent choice to skill learning: a reinforcement-learning model. *Journal of experimental psychology. General*, 135(2), 184–206.
- Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science*, 4(6), 385-390.
- Gershman, S. J., Markman, A. B., & Otto, A. R. (2014). Retrospective revaluation in sequential decision making: A tale of two systems. *Journal of Experimental Psychology: General*, 143(1), 182.
- Ghezzi, C., Pezzè, M., Sama, M., & Tamburrelli, G. (2014). Mining behavior models from

- user-intensive web applications. In *Proceedings of the 36th international conference on software engineering* (pp. 277–287).
- Gmytrasiewicz, P. J., & Doshi, P. (2005). A framework for sequential planning in multi-agent settings. *Journal of Artificial Intelligence Research*, 24, 49–79.
- Goel, N., Basner, M., Rao, H., & Dinges, D. F. (2013). Chapter seven - circadian rhythms, sleep deprivation, and human performance. In M. U. Gillette (Ed.), *Chronobiology: Biological timing in health and disease* (Vol. 119, p. 155 - 190). Academic Press.
- Goh, H.-T., Gordon, J., Sullivan, K. J., & Winstein, C. J. (2014). Evaluation of attentional demands during motor learning: Validity of a dual-task probe paradigm. *Journal of Motor Behavior*, 46(2), 95-105. (PMID: 24447033)
- Gorman, A. D., Abernethy, B., & Farrow, D. (2015). Evidence of different underlying processes in pattern recall and decision-making. *The Quarterly Journal of Experimental Psychology*, 0(0), 1-19. (PMID: 25662705)
- Gramfort, A., Luessi, M., Larson, E., Engemann, D., Strohmeier, D., Brodbeck, C., ... Hämäläinen, M. (2013). Meg and eeg data analysis with mne-python. *Frontiers in Neuroscience*, 7, 267.
- Hershler, O., & Hochstein, S. (2009, Oct). The importance of being expert: top-down attentional control in visual search with photographs. *Atten Percept Psychophys*, 71(7), 1478-86.
- Huys, Q. J. M., Lally, N., Faulkner, P., Eshel, N., Seifritz, E., Gershman, S. J., ... Roiser, J. P. (2015). Interplay of approximate planning strategies. *Proceedings of the National Academy of Sciences*, 112(10), 3098-3103.
- Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991, February). Adaptive Mixtures of Local Experts. *Neural Computation*, 3(1), 79–87.
- Jasper, H. H. (1958). The ten twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, 10, 371–375.
- Jeon, H., & Shin, D. A. (2015). Experimental set up of p300 based brain computer interface using a bioamplifier and bci2000 system for patients with spinal cord injury. *Korean Journal of Spine*, 12(3), 119–123.

- Karwowski, W., & Cuevas, H. M. (2003). Considering the importance of individual differences in human factors research: No longer simply confounding noise. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 47(9), 1082-1086.
- Larson, M. J., Clayson, P. E., & Clawson, A. (2014). Making sense of all the conflict: A theoretical review and critique of conflict-related {ERPs}. *International Journal of Psychophysiology*, 93(3), 283 - 297.
- Laud, A. D. (2004). *Theory and application of reward shaping in reinforcement learning* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.
- Lewandowsky, S., & Oberauer, K. (2009, Nov). No evidence for temporal decay in working memory. *J Exp Psychol Learn Mem Cogn*, 35(6), 1545–1551.
- Liu, Y., Nelson, L. D., Bernat, E. M., & Gehring, W. J. (2014). Perceptual properties of feedback stimuli influence the feedback-related negativity in the flanker gambling task. *Psychophysiology*, 51(8), 782–788.
- Liverence, B. M., & Scholl, B. J. (2015, Jul). Object persistence enhances spatial navigation: A case study in smartphone vision science. *Psychol Sci*, 26(7), 955–963.
- Luck, S. (2005). *An introduction to the event-related potential technique (cognitive neuroscience)*. The MIT Press.
- Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.
- Madl, T., Chen, K., Montaldi, D., & Trapp, R. (2015). Computational cognitive models of spatial memory in navigation space: A review. *Neural Networks*, 65(0), 18 - 43.
- Maier, M. E., Yeung, N., & Steinhauser, M. (2011). Error-related brain activity and adjustments of selective attention following errors. *NeuroImage*, 56(4), 2339 - 2347.
- Manoel, E. d. J., Basso, L., Correa, U. C., & Tani, G. (2002). Modularity and hierarchical organization of action programs in human acquisition of graphic skills. *Neuroscience Letters*, 335(2), 83–86.
- Mariano, L., Poore, J., Krum, D., Schwartz, J., Coskren, W., & Jones, E. (2015). Modeling strategic use of human computer interfaces with novel hidden markov models. *Frontiers in Psychology*, 6, 919.

- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of eeg- and meg-data. *Journal of Neuroscience Methods*, *164*(1), 177 - 190.
- Marken, R. S. (1986). Perceptual organization of behavior: a hierarchical control model of coordinated action. *Journal of Experimental Psychology: Human Perception and Performance*, *12*(3), 267.
- Matsuzaka, Y., Picard, N., & Strick, P. L. (2007). Skill representation in the primary motor cortex after long-term practice. *Journal of neurophysiology*, *97*(2), 1819–1832.
- Melby-Lervag, M., Lyster, S.-A. H., & Hulme, C. (2012, Mar). Phonological skills and their role in learning to read: a meta-analytic review. *Psychol Bull*, *138*(2), 322–352.
- Miltner, W. H., Braun, C. H., & Coles, M. G. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a generic neural system for error detection. *Journal of cognitive neuroscience*, *9*(6), 788–798.
- Nicolas-Alonso, L. F., & Gomez-Gil, J. (2012). Brain computer interfaces, a review. *Sensors (Basel, Switzerland)*, *12*(2), 1211–1279.
- Oh, J., Meneguzzi, F., & Sycara, K. (2011). Probabilistic plan recognition for intelligent information agents. *Proc. ICAART*, 281–287.
- Otto, A. R., Skatova, A., Madlon-Kay, S., & Daw, N. D. (2014). Cognitive control predicts use of model-based reinforcement learning. *Journal of Cognitive Neuroscience*.
- Parasuraman, R., & Jiang, Y. (2012). Individual differences in cognition, affect, and performance: Behavioral, neuroimaging, and molecular genetic approaches. *NeuroImage*, *59*(1), 70 - 82. (Neuroergonomics: The human brain in action and at work)
- Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annu Rev Neurosci*, *35*, 73–89.
- Phillips, D. A., & Lowenstein, A. E. (2011). Early care, education, and child development. *Annual Review of Psychology*, *62*(1), 483-500. (PMID: 20822436)
- Polezzi, D., Lotto, L., Daum, I., Sartori, G., & Rumiati, R. (2008). Predicting outcomes of decisions in the brain. *Behavioural brain research*, *187*(1), 116–122.

- Polich, J. (2007). Updating p300: an integrative theory of p3a and p3b. *Clinical neurophysiology*, *118*(10), 2128–2148.
- Pool, E., Brosch, T., Delplanque, S., & Sander, D. (2014). Where is the chocolate? rapid spatial orienting toward stimuli associated with primary rewards. *Cognition*, *130*(3), 348 - 359.
- Posner, M. (1980). Orienting of attention. *The Quarterly journal of experimental psychology*, *32*(1), 3–25.
- Posner, M., Snyder, C., & Davidson, B. (1980). Attention and the detection of signals. *Journal of experimental psychology*, *109*(2), 160–174.
- Powers, J. C., Bieliaieva, K., Wu, S., & Nam, C. S. (2015, 09). The human factors and ergonomics of p300-based brain-computer interfaces. *Brain Sciences*, *5*(3), 318–354.
- Rabiner, L. (1989, February 06). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257–286.
- Rabinovich, M. I., Huerta, R., & Afraimovich, V. (2006). Dynamics of sequential decision making. *Physical review letters*, *97*(18), 188103.
- Ramírez, M., & Geffner, H. (2011). Goal recognition over pomdps: Inferring the intention of a pomdp agent. In *Twenty-second international joint conference on artificial intelligence*.
- Ranganath, C., & Paller, K. A. (2000). Neural correlates of memory retrieval and evaluation. *Cognitive Brain Research*, *9*(2), 209–222.
- Rosenbaum, D. A., Cohen, R. G., Jax, S. A., Weiss, D. J., & Van Der Wel, R. (2007). The problem of serial order in behavior: Lashleys legacy. *Human movement science*, *26*(4), 525–554.
- Ross, B., & Tremblay, K. (2009). Stimulus experience modifies auditory neuromagnetic responses in young and older listeners. *Hearing research*, *248*(1), 48–59.
- Ruh N., C. R. P., & Mareschal, D. (2010). Action selection in complex routinized sequential behaviors. *Journal of Experimental Psychology: Human Perception and Performance*, *106*, 99–114.

- Sacchi, S., & Burigo, M. (2008). Strategies in the information search process: Interaction among task structure, knowledge, and source. *The Journal of General Psychology*, *135*(3), 252-270. (PMID: 18649492)
- Sallet, J., Camille, N., & Procyk, E. (2013). Modulation of feedback-related negativity during trial-and-error exploration and encoding of behavioural shifts. *Frontiers in Neuroscience*, *7*(209).
- San Martín, R. (2012). Event-related potential studies of outcome processing and feedback-guided learning. *Frontiers in Human Neuroscience*, *6*(304).
- Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? *Methodology*.
- Schriver, A. T., Morrow, D. G., Wickens, C. D., & Talleur, D. A. (2008, Dec). Expertise differences in attentional strategies related to pilot decision making. *Hum Factors*, *50*(6), 864-78.
- Schuermann, B., Endrass, T., & Kathmann, N. (2012). Neural correlates of feedback processing in decision making under risk. *Frontiers in Human Neuroscience*, *6*(204).
- Schulte, P. J., Tsiatis, A. A., Laber, E. B., & Davidian, M. (2014). Q-and a-learning methods for estimating optimal dynamic treatment regimes. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *29*(4), 640.
- Sepahvand, N. M., Stöttinger, E., Danckert, J., & Anderson, B. (2014). Sequential decisions: a computational comparison of observational and reinforcement accounts. *PloS one*, *9*(4), e94308.
- Shalgi, S., & Deouell, L. Y. (2012). Is any awareness necessary for an ne? *Frontiers in Human Neuroscience*, *6*(124).
- Shteingart, H., & Loewenstein, Y. (2014). Reinforcement learning and human behavior. *Current Opinion in Neurobiology*, *25*(0), 93 - 98. (Theoretical and computational neuroscience)
- Sims, C. R., Neth, H., Jacobs, R. A., & Gray, W. D. (2013). Melioration as rational choice: Sequential decision making in uncertain environments. *Psychological review*, *120*(1), 139.

- Singer, P., Helic, D., Taraghi, B., & Strohmaier, M. (2014, 07). Detecting memory and structure in human navigation patterns using markov chain models of varying order. *PLOS ONE*, *9*(7), 1-21.
- Smith, T., Henning, R., Wade, M., & Fisher, T. (2014). *Variability in human performance*. Taylor & Francis.
- Solway, A., & Botvinick, M. M. (2012, January). Goal-directed decision making as probabilistic inference: a computational framework and potential neural correlates. *Psychological review*, *119*(1), 120–154.
- Spencer, K. M., Abad, E. V., & Donchin, E. (2000). On the search for the neurophysiological manifestation of recollective experience. *Psychophysiology*, *37*(4), 494–506.
- Sukthankar, G., Geib, C., Bui, H. H., Pynadath, D., & Goldman, R. P. (2014). *Plan, activity, and intent recognition: Theory and practice*. Newnes.
- Van Dongen, H. P. A. (2006). Shift work and inter-individual differences in sleep and sleepiness. *Chronobiology International*, *23*(6), 1139-1147.
- Visser, I., Raijmakers, M. E. J., & Molenaar, P. C. M. (2002, August). Fitting hidden markov models to psychological data. *Sci. Program.*, *10*(3), 185–199.
- Walsh, M., & Anderson, J. (2011). Learning from delayed feedback: neural responses in temporal credit assignment. *Cognitive, Affective, & Behavioral Neuroscience*, *11*(2), 131-143.
- Walsh, M. M., & Anderson, J. R. (2012). Learning from experience: event-related potential correlates of reward processing, neural adaptation, and behavioral choice. *Neuroscience & Biobehavioral Reviews*, *36*(8), 1870–1884.
- Walsh, M. M., & Anderson, J. R. (2014, 03). Navigating complex decision spaces: Problems and paradigms in sequential choice. *Psychological bulletin*, *140*(2), 466–486.
- Wolk, D. A., Schacter, D. L., Lygizos, M., Sen, N. M., Holcomb, P. J., Daffner, K. R., & Budson, A. E. (2006). Erp correlates of recognition memory: Effects of retention interval and false alarms. *Brain research*, *1096*(1), 148–162.