

Look and Learn – A Computational Model of Gaze-Contingent Learning

Max Murakami

December 10, 2017

Contents

1	Introduction	1
1.1	Intrinsic Motivation	2
1.2	Contingency Learning in Infants	6
1.2.1	The Gaze-Contingent Paradigm	7
2	Methods	8
2.1	The Experiment	8
2.2	Data Processing	10
2.3	Model 1	12
2.3.1	The Prediction System	12
2.3.2	Intrinsic Salience	16
2.3.3	The Neural Circuit Model	16
2.3.4	Action Selection	19
2.3.5	Environment	19
2.3.6	Parameter and Simulation Details	19
2.4	Model 2	22
2.5	Model Fitting	23
3	Results	25
3.1	Behavioral Bias	25
3.2	Effect of Age	29
3.3	Individual Learning Progress	37
3.4	Learning Speed	41
3.4.1	Median Split	41
3.4.2	Bias Correlations	42
3.4.3	Outlier Analysis	52
3.4.4	Uncertainty Timing	53
3.4.5	Acquisition Time	65
3.4.6	Conclusion	75
3.5	Exploration Rate	76
3.5.1	Median Split	76
3.5.2	Bias Correlations	78
3.5.3	Exploration and Prediction Adaptation	87
4	Discussion	89

Zusammenfassung

Diese Dissertation beschäftigt sich mit der Frage, mit Hilfe welcher Prinzipien sich intrinsische Motivation formalisieren lässt und wie diese es uns ermöglichen, kausale Zusammenhänge unserer Umwelt zu erlernen. Konkret wird das Verhalten von sechs bis zehn Monate alten Säuglingen untersucht, die an einem Experiment des sogenannten blickkontingenten Paradigmas teilnehmen. Hierbei lernt der Säugling seine visuelle Umgebung mit Hilfe seines Blicks zu beeinflussen. Das Blickverhalten, das der Säugling während des Lernvorgangs an den Tag legt, wird quantitativ ausgewertet und mit Hilfe von Computermodellen nachgebildet. Auf diese Weise wird ein direkter Bezug zwischen dem Verhalten des Säuglings und mathematisch formulierten Theorien geschaffen, die Aussagen über mögliche zu Grunde liegende formale Prinzipien ermöglichen. Die Haupterkenntnis dieser Studie ist, dass sich typische Verhaltensmuster der Säuglingsprobanden, insbesondere funktional-ausgerichtete Blickpräferenzen, auf den Prozess der Vorhersageoptimierung des internen Weltmodells per aktiver Informationsmaximierung zurückführen lassen.

Die Dissertation umfasst vier Kapitel. Die Einleitung umreißt die wissenschaftliche Fragestellung und gibt einen Überblick über den Inhalt der Studie. Es folgt eine Einführung in das Thema der intrinsischen Motivation, aus einer historischen, einer interdisziplinären und einer Modellierungs-Perspektive. Entwicklungen auf der Ebene der Theorien werden ebenso abgehandelt wie wichtige Voraarbeiten und Modellierungsstudien, die einen direkten Bezug auf diese Dissertationsarbeit aufweisen. Darüber hinaus wird ein konziser Überblick über experimentelle entwicklungspsychologische Studien zum Kontingenzlernen in Säuglingen gegeben. Aufbauend auf den Entwicklungen der letzten Jahrzehnte wird übergegangen zum experimentellen Paradigma des blickkontingenten Lernens, das 2012 von Wang und Kollegen vorgestellt wurde [1]. Deren Studie wird im Detail diskutiert, da sie die Grundlage der empirischen Untersuchungen dieser Dissertation darstellt.

Der Methodenteil beschreibt im Detail die Materialien, Daten und Modelle, auf denen diese Studie basiert. Das Experiment wird erläutert ebenso wie die Datenextraktion und -verarbeitung. Die Computermodelle werden sowohl auf der konzeptionellen als auch auf der mathematischen Ebene definiert. Das Kapitel schließt mit einer Erläuterung des algorithmischen Modell-Fittings.

Das dritte Kapitel präsentiert die Ergebnisse. Hier werden die Analysen, die statistischen Auswertungen sowie deren Interpretationen vorgestellt. Nach einer qualitativen Betrachtung und quantitativen Erfassung der Blickpräferenz wird der Einfluss des Alters auf die Daten untersucht und inwiefern die Modelle Aussagen zu diesem Faktor treffen. Des Weiteren wird die Dynamik des Lernverlaufs individueller Modellprobanden quantitativ charakterisiert und schließlich der Einfluss der Lerngeschwindigkeit und

der Erkundungsbereitschaft auf das Lernverhalten der Säuglinge auf unterschiedlichen Ebenen analysiert.

Im letzten Kapitel werden die Befunde der Studie zusammengefasst und vor dem Hintergrund existierender Theorien und Arbeiten kritisch diskutiert. Mögliche Schwächen des vorgestellten Forschungsansatzes, der Daten und der Analysen werden dabei ebenfalls angesprochen. Die Dissertation schließt mit einem Ausblick und Vorschlägen, wie der vorliegende Forschungsansatz verbessert und die Realität detailgetreuer modelliert werden kann.

Intrinsische Motivation ist ein Konzept, das in der Psychologie der 1950er Jahre aufkam [2, 3, 4, 5] und sich in den letzten Jahren zu einem aktiven Forschungsthema in den Bereichen der Robotik, der künstlichen Intelligenz, des maschinellen Lernens und der theoretischen Neurowissenschaften entwickelt hat [6]. Während Psychologen von intrinsischer Motivation sprechen, wenn etwas getan wird, weil es inhärent interessant oder angenehm ist [7], existieren in der modernen algorithmisch-geprägten Literatur Definitionen, die intrinsische Motivation mit Lernfähigkeit assoziieren und zugleich abgrenzen von der Überlebens- und Fortpflanzungsfähigkeit eines Organismus [8]. Intrinsische Motivation wird in Verbindung gebracht mit Merkmalen menschlicher Zivilisation wie künstlerische Kreativität und wissenschaftlichem Entdeckungsdrang [9, 10]. Roboter und andere künstliche Systemen mit intrinsischer Motivation auszustatten wird heutzutage als ein notwendiger Schritt zu wahrer künstlicher Intelligenz angesehen [11].

Motivation beschreibt *warum* wir tun, was wir tun [7], setzt dabei aber Autonomie voraus, d.h. ein gewisses Maß an Selbstkontrolle [12]. Frühe Theorien der Motivation werden heutzutage als Theorien der extrinsischen Motivation bezeichnet, da sie sich auf körperliche Bedürfnisse beschränken. Ein Beispiel ist Hulls einflussreiche Theorie der Triebe, die in den 1940er und 50er Jahren formuliert wurde [13, 14, 15]. Der Begriff *intrinsische Motivation* geht auf Harlow aus dem Jahr 1950 zurück. Harlow berichtete damals von Affen, die sich in einer Art und Weise verhielten, die sich nicht mit den damaligen Theorien der Motivation erklären ließ [2]. Berlyne schlug daraufhin vor, intrinsische Motivation mittels bestimmter Stimulus-Eigenschaften wie Neuartigkeit und Vorhersagbarkeit zu erklären [4, 16, 17]. Da sich diese Eigenschaften auf das Wissen des jeweiligen Subjekts beziehen, formulierte er damit die *wissensbasierte* Sichtweise der intrinsischen Motivation [18]. Demgegenüber steht die *kompetenzbasierte* Sichtweise, die von White formuliert wurde und besagt, dass die Steigerung der Kompetenz mit der Umwelt zu interagieren im Zentrum der intrinsischen Motivation steht [5].

Die meisten Modelle der intrinsischen Motivation sind wissensbasiert, wie beispielsweise Schmidhubers bahnbrechende Arbeit von 1991, in der er Neugier als Drang zur Vorhersageoptimierung mittels Informationsmaximierung implementierte [19]. Jedoch existieren auch kompetenzbasierte Modelle, die üblicherweise das Erlernen und Perfektionieren bestimmter Fertigkeiten thematisieren, so wie Bartos Intrinsically Motivated Reinforcement Learning von 2004 [20, 21]. Zwar ist man sich darüber einig, dass mit beiden Ansätzen sowohl das Wissen als auch die Kontrolle des Agenten über die Umwelt zunimmt, welcher dieser Aspekte jedoch das fundamentale Prinzip ist, wird kontrovers diskutiert [8, 22, 23].

Nach neurowissenschaftlichem Erkenntnisstand basiert die intrinsische Motivation im Gehirn auf wissensbasierten Signalen, wobei dem Neurotransmitter *Dopamin* eine zentrale Rolle zugesprochen wird, da er sowohl unerwartete als auch neuartige Stimuseigenschaften zu kodieren scheint [24, 25, 26, 27]. Kompetenzbasierte Signale wurden hingegen noch nicht gefunden [8]. Basierend auf diesen Erkenntnissen formulierten Redgrave und Gurney ihre Hypothese des *Repetition Bias*, wonach es einen kausalen Zusammenhang zwischen intrinsisch motiviertem Kontingenzlernen und der Bildung von Verhaltensmustern gibt [28, 29, 30].

In einer abschließenden Betrachtung ist anzumerken, dass aus einer evolutionären Perspektive die Grenze zwischen extrinsischer und intrinsischer Motivation verschwimmt [22]. Barto plädiert dafür, diese strikte Unterscheidung aufzuheben und stattdessen von einem Kontinuum zu sprechen, wobei sich am Ende des Spektrums extrinsische Motivation eher direkt und unmittelbar auf evolutionären Erfolg auswirkt, auf der anderen Seite der Einfluss intrinsischer Motivation entsprechend eher indirekt und weniger offensichtlich ist [22].

Untersuchungen des *Kontingenzlernens* in Säuglingen stellen wegen derer langsamem motorischen Entwicklung eine besondere Herausforderung dar, weshalb sie sich traditionell auf Reflexe und basales motorisches Verhalten beschränken [31, 32]. Beispielsweise wurde der Saugreflex benutzt um zu zeigen, dass bereits Neugeborene ihr Saugverhalten ändern, wenn sie damit Eigenschaften von Stimuli beeinflussen [33]. In einem anderen Fall fanden Rochat und Striano anhand des Suchreflexes heraus, dass Neugeborene zwischen Berührungen durch sich selbst und Berührungen durch Objekte unterscheiden können [34]. In der sogenannten Mobile-Aufgabe zeigten Rovee-Collier und Kollegen, dass zwei Monate alte Säuglinge die Kontingenz zwischen ihrem Strampeln und der Bewegung eines mit ihren Beinen verbundenen Mobiles erlernen [31, 35]. Weiterhin konnte Kenward zeigen, dass zehn Monate alte Säuglinge visuelle Stimuli antizipieren, die sie per Knopfdruck auslösen, womit sie die Formierung von Erwartungen und Vorhersagen demonstrierten [36].

Da sich die Augenkoordination bereits relativ früh entwickelt [37, 38], etablierte sich die Analyse des Blickverhaltens als informatives, flexibel einsetzbares Mittel in der Säuglingsforschung [39, 40, 41, 42, 43, 44, 45, 46, 47]. Dank technischer Neuerungen ermöglicht die *Blickerfassungsmethode* (eye tracking) automatisierte und präzise Auswertungen des Blickverhaltens in Echtzeit und gehört dadurch zum Repertoire moderner Säuglingsforschung [48, 46, 49, 50, 51]. Durch die Blickerfassungsmethode kann eine neue Klasse von Experimenten realisiert werden, die sogenannten *blickkontingenten* Experimente, die zwar schon seit geraumer Zeit mit Erwachsenen durchgeführt werden [52, 53], aber erst seit wenigen Jahren in der Säuglingsforschung angewendet werden [54, 55, 56, 57, 58].

Ein spezieller Fall ist das *blickkontingente Paradigma* (gaze-contingent paradigm), das von Wang und Kollegen 2012 vorgestellt wurde [1] und die Grundlage für die empirischen Untersuchungen in dieser Dissertation darstellt. In der Originalversion dieses Experiments ist eine rote Scheibe auf dem Bildschirm zu sehen, die bei Fixation darauf das Erscheinen eines Tierbilds auslöst und somit als optischer Schalter fungiert.

In der Originalstudie wurde gezeigt, dass sechs und acht Monate alte Säuglinge lernen das Tierbild zu antizipieren, nachdem sie es auslösen, und folglich die Kontingenz zwischen ihrem Blick und der visuellen Antwort erfolgreich erwerben. Mit Hilfe der erweiterten Zwei-Scheiben-Version demonstrieren die Autoren, dass die Säuglinge tatsächlich das Erscheinen des Tierbilds mit ihrem Blick bezwecken, da sie den optischen Schalter gegenüber einer identischen Scheibe ohne Schalterfunktion präferieren. Interessanterweise konnte ein Großteil der erwachsenen Probanden, die das Experiment ebenfalls absolvierten, den Mechanismus nicht erklären; diejenigen, die es konnten, zeigten ein vergleichbares Blickverhalten wie die Säuglinge, was für einen Erwerb von Einsicht und Erwartungen in den Säuglingen spricht.

Das Experiment, das im Rahmen dieser Dissertation ausgewertet wurde, basiert auf der Zwei-Scheiben-Version des blickkontingenten Paradigmas von Wang et al. Im Gegensatz zur Originalstudie wurden jedoch zum einen sechs, acht und auch zehn Monate alte Säuglinge getestet, zum anderen wurde der Test für jeden Probanden zwei Mal wiederholt, das erste Mal nach 15 bis 19 Minuten, das zweite Mal nach einer Woche. Außerdem gab es eine experimentelle Kontrollbedingung, in der Probanden keine Tierbilder per Blick auslösen konnten, sondern lediglich ein Video sahen vom Bildschirm, das bei einem früheren Test eines Säuglings in der aktiven Bedingung aufgezeichnet wurde. Im Rahmen dieser Studie wurden ausschließlich die Daten der aktiven, blickkontingenten Gruppe ausgewertet und modelliert.

Die Rohdaten des Eye-Trackers wurden sowohl in der zeitlichen als auch der räumlichen Domäne gefiltert bzw. ausgewertet. Zum einen wurden Zeitabschnitte ausgeschlossen, in denen die Probanden nicht aktiv am Experiment teilnahmen. Zum anderen wurde mittels einer Post-Hoc-Schätzung sichergestellt, dass nur Tests in die Analyse aufgenommen wurden, in denen die Blickkontingenz mit einer hohen Zuverlässigkeit gewährleistet war und die Lernaufgabe nicht durch Kalibrierungsungenauigkeiten stark erschwert wurde.

Modell 1 basiert auf einer Studie, in der die Gültigkeit der Repetition-Bias-Hypothese überprüft werden sollte [59]. Da diese Hypothese Aussagen über Verhaltensanpassung in intrinsisch motivierten Lernsituationen trifft, ist das Modell relevant für diese Studie und wurde übernommen und an das blickkontingente Lernexperiment angepasst. Es umfasst zum einen ein Modell der neuronalen Informationsverarbeitung in relevanten Hirnregionen, insbesondere den Basalganglien und assoziierten Bereichen. Zum anderen enthält es ein Vorhersagesystem, das den Wissensstand des Probanden repräsentiert, sich den Erfahrungen des Probanden anpasst, maßgeblich das Verhalten zu Gunsten der Informationsmaximierung beeinflusst und die Ausschüttung von Dopamin als sensorischem Vorhersagefehler veranlasst. Letzteres moduliert die synaptische Plastizität zwischen Hirnrinde und Basalganglien und trägt somit zur Verhaltensanpassung bei. Modell 2 ist eine abstrahierte Version von Modell 1, in der der biologisch motivierte Teil weggelassen wurde. Somit besteht es einzig aus dem Vorhersagesystem, das das Wissen und die intrinsische Motivation des Probanden repräsentiert und sich somit auf die abstrakten Prinzipien der intrinsisch motivierten Verhaltenssteuerung beschränkt. Beide Modelle werden mit Hilfe des CMA-ES-Algorithmus gefittet, wobei der Unterschied zwischen dem simulierten und dem experimentell beobachteten Verhalten pro Proband in einer

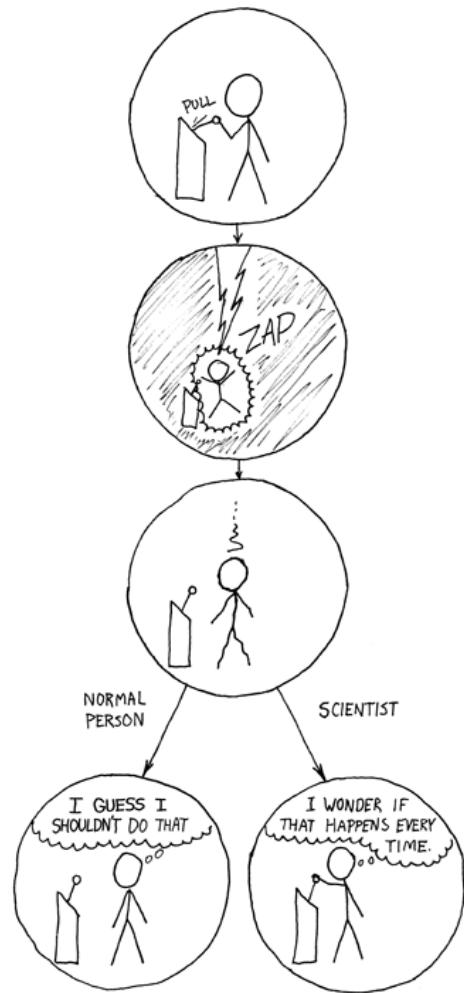
stochastischen Optimierung minimiert wird [60].

Die Simulationsergebnisse zeigen, dass beide Modelle die auf die funktionale Scheibe ausgerichtete Blickpräferenz der Säuglinge reproduzieren können. Zudem replizieren sie den experimentellen Befund, dass die Blickpräferenz der acht und zehn Monate alten Säuglinge im Gegensatz zu den sechs Monate alten stark ausgeprägt ist. Die Datenlage bezüglich der Modelle lässt allerdings keine statistisch haltbare Erklärung zu.

Auf der Ebene der einzelnen Probanden konnten zeitlich unterschiedlich ausgeprägte Lernverläufe beobachtet werden, die zu qualitativ unterschiedlichen Verhaltensmustern führen. Dies lieferte einen ersten Hinweis darauf, dass die Lerngeschwindigkeit, also die Adoptionsrate des Vorhersagesystems, maßgeblich die Entstehung der Blickpräferenz beeinflusst. Verschiedene Analysen führten zu folgendem Ergebnis: Je langsamer ein Proband lernt, desto später erwirbt er die Kontingenz, desto stärker ausgeprägt ist seine Blickpräferenz. Der Hauptbefund lässt sich somit wie folgt formulieren: *Die funktional-ausgerichtete Blickpräferenz ist eine Konsequenz der andauernden Vorhersageoptimierung.* Allgemeiner ausgedrückt: *Fortwährende Informationsmaximierung führt zur Bildung von Verhaltenspräferenzen.* Dieses Ergebnis steht im Einklang mit der Repetition-Bias-Hypothese, wonach Verhaltenspräferenzen auftreten müssen, damit kontingente, noch nicht erworbene Konsequenzen des eigenen Handelns zuverlässig erlernt werden können [28, 29, 30]. Dieser Sichtweise zufolge ist die Verhaltenspräferenz eine vorübergehende Phase der Vorhersageoptimierung, die solange anhält, bis die Vorhersagen des internen Modells den Erfahrungen bezüglich der Interaktion mit der Umwelt entsprechen.

Des Weiteren ergaben die Modellanalysen, dass langsame Lerner weniger zum Explorieren neigen als schnelle Lerner. Dieses Phänomen lässt sich gut mit Hilfe des *exploration-exploitation-Dilemmas* verstehen, das in der Literatur des verstärkenden Lernens seit längerem bekannt ist [61]. Dieses Dilemma besagt, dass ein Agent, der ein Belohnungssignal maximieren will, sich zu jeder Zeit entscheiden muss, ob er sein Wissen ausnutzen sollte um eine bekannte Menge an Belohnung zu erhalten (exploitation), oder ob er lieber noch unbekannte Optionen ausprobieren sollte, um eine potentiell größere Menge an Belohnung zu erhalten (exploration). Im Falle des intrinsisch motivierten Kontingenzlernens ist die Belohnung der erwartete Informationsgehalt, den der Agent nach Ausführung einer Aktion erhält, basierend auf seinem Wissensstand. Der Repetition-Bias-Hypothese zufolge ist die erwartete Belohnung gering am Anfang und am Ende des Kontingenzlernens. Zu diesen Zeiten ist das Explorieren also wichtig, um Aktionen mit unerwarteten Konsequenzen zu entdecken. Sobald eine solche gefunden wurde, herrscht die Exploitation-Phase vor, in der stetig das Vorhersagemodell bezüglich dieser Aktion verbessert wird. Je länger diese Phase also dauert, wie es bei langsamen Lernern der Fall ist, desto weniger ist der Agent auf Exploration angewiesen. Auf der anderen Seite haben schnelle Lerner ein größeres Bedürfnis zum Explorieren, da die Exploitation-Phase kürzer ist und schneller wieder nach neuen erlernbaren Kontingenzen gesucht wird.

Acknowledgements



When intrinsic motivation prevails (bottom right). From [62].

1 Introduction

How do we come to understand the world? Specifically, how do we learn about causality and the consequences of our actions? What drives us to interact with our environment, and why do we seek to manipulate the world, incessantly and in ever novel ways? What are the underlying principles of curious behavior?

These fundamental questions have driven my research, and this thesis reports on my efforts to help answer them. In particular, I studied the looking behavior of 6- to 10-month-old infants while they learned to manipulate their environment in an experiment following the so-called *gaze-contingent paradigm*. I analyzed the behavioral patterns, which they exhibit during the experiment, and sought to explain how these patterns develop from a theoretical perspective. To this end, I simulated two computational models of action selection, which are based on the notion of *intrinsic motivation*, a concept that is used to explain curious behavior, among other things. One model describes in detail processes within parts of the brain that may be responsible for the observed behavior, while the other model is purely abstract and contains no references to biological implementations. However, the models share the assumption that we act to improve our understanding of the world in an optimal way, i.e., we seek to maximize the information we gain based on our beliefs and expectations in order to improve our predictions. The models reproduce the observations, offering explanations for the infants' behavior in general as well as for the differences between infants. Ultimately, my work indicates that intrinsic motivation in terms of prediction optimization might indeed underlie the behavioral patterns that infants exhibit during gaze-contingent learning.

This thesis is structured in four parts. After these opening words, I review basic concepts that my research is based on. I introduce and discuss the notion of intrinsic motivation from a historical and interdisciplinary perspective. My accounts of this topic also cover modelling approaches and theories related to my work. The following section briefly reviews past and current developments in the study of contingency learning in infants. It then elaborates on the gaze-contingent paradigm, which forms the empirical basis of my investigations. In the *Methods* chapter, I detail the experiment and how the data are extracted. I then present parsimonious descriptions of the models and how they are fitted to the experimental data. I report on my investigations and analyses in the *Results* chapter. This part includes considerations about the model fitting, age differences, the dynamics of contingency learning, and how learning speed and exploration impact looking behavior. Finally, the *Discussion* chapter summarizes the main findings and reviews them in light of existing theories and related work. It addresses notable observations of the analyses and shortcomings of the modelling approach. The thesis concludes by giving an outlook and offering suggestions for how to enhance the modelling approach presented here.

1.1 Intrinsic Motivation

The human organism is inherently active, and there is perhaps no place where this is more evident than in little children. They pick things up, shake them, smell them, taste them, throw them across the room, and keep asking, "What is this?" They are unendingly curious, and they want to see the effects of their actions. Children are intrinsically motivated to learn, to undertake challenges, and to solve problems. Adults are also intrinsically motivated to do a variety of things. They spend large amounts of time painting pictures, building furniture, playing sports, whittling wood, climbing mountains, and doing countless other things for which there are not obvious or appreciable external rewards. The rewards are inherent in the activity, and even though there may be secondary gains, the primary motivators are the spontaneous, internal experiences that accompany the behavior.

This quote by Deci and Ryan [63] nicely captures the phenomena associated with *intrinsic motivation*, a concept that has started to receive attention by psychologists and ethologists in the 1950s [2, 3, 4, 5] and has nowadays become an established research topic in the fields of robotics, artificial intelligence, and machine learning [6]. While classical definitions refer intrinsic motivation to "doing something because it is inherently interesting or enjoyable" [7], computational accounts describe intrinsic motivation as motivation that is "directly related not to an organism's survival and reproduction but rather to its ability to learn" [8]. Intrinsic motivation is linked to hallmarks of human civilization such as artistic creativity and scientific discovery [9, 10]. The idea that robots could eventually behave in such intrinsically motivated ways seems alien to us, given the current state of humanoid robots, yet endowing artificial systems with basic forms of intrinsic motivation is these days considered to be vital for achieving true artificial intelligence [11].

Motivation is about *why* we do what we do, or "[t]o be motivated means *to be moved* to do something" [7]. It is about processes that influence our behavior in terms of arousal, strength, and direction. The factors influencing our behavior are three-fold [64]:

- Irresistible influences of the environment: The strength and direction of these influences are mostly independent of our internal state. Reflexive behavior falls into this category.
- Internal factors such as urges, drives, needs, wants, plans.
- External situations or objects if they act as incentives or goals. Even though they are external factors, they control behavior based on value generated by our internal state.

Motivation refers to those factor that are linked to our internal state, i.e., the latter two of the listed items. Apart from internal states, motivation also requires autonomy, i.e., the capability of self-control [12].

Conventional theories of motivation are based on the principle of *physiological homeostasis* [65], which describes processes maintaining a general equilibrium of the bodily conditions despite perturbations by the environment. Drawing on this principle, Hull introduced his influential drive theories in the 1940s and 50s [13, 14, 15]. According to him, behavior is directed towards reducing physiological deficits (drive reduction) and thereby maintaining the physiological equilibrium. He distinguishes between primary drives (avoidance of pain, hunger, thirst, and sex), whose reduction is inherently rewarding and thereby reinforced, and derivative drives, which become motivationally significant by association with primary drives. Due to their strong focus on bodily needs, theories such as this represent accounts of *extrinsic motivation*.

But not all behavior can be explained by drives and extrinsic motivations. Harlow introduced the term *intrinsic motivation* in 1950 when he reported that rhesus monkeys spontaneously engage in object manipulation and attempt to solve intricate puzzles for extended periods of time without being rewarded [2]. While these kinds of behavior continued to be studied in animals as well as in humans, it became clear that they are not based on extrinsic motivators; in fact, behavior related to curiosity and play seems to elicit intrinsic reward that is on par with the primary reward predicted by the drive theories [5, 66]. Furthermore, this kind of behavior is particularly prominent in infants, who are unlikely to associate exploratory behavior with drive reduction before they are born.

Berlyne suggested formulating intrinsic motivation in the context of optimal level theories rather than stretching the concept of drives. Specifically, he hypothesized that animals are attracted by optimal levels of stimulus properties, i.e., novelty, surprising features, and complexity [4, 16, 17]. Because these properties depend on what the animal knows, Berlyne’s approach is an example of *knowledge-based* views of intrinsic motivation [18].

On the other hand, White proposed in 1959 to describe behavior directed to manipulation, exploration, and activity in terms of *effectance motivation*, which he defined as ”an intrinsic need to deal with the environment” [5]. In this context, he introduced the concept of *competence* as effective interaction with the environment. He argued that competence is ultimately a means of achieving *autonomy*, i.e., of mastering the environment. This *competence-based* view of intrinsic motivation focusses on the acquisition of skills rather than information [18].

Most computational models of intrinsic motivations so far are knowledge-based [8]. Most notably, the first of these models was devised by Schmidhuber in 1991 and implements curiosity as prediction optimization given an adaptive world model [19]. Oudeyer and colleagues have adapted and extended this approach to robotics [18]. In contrast to these *prediction-based* accounts of intrinsic motivation, *novelty-based* approaches focus on detecting and attending to stimuli that the agent has not encountered before [67, 22]. Models based on the latter comprise [68] and [69].

The first competence-based model was presented by Barto (and colleagues) in 2004, who introduced the concept of intrinsic motivation in robotics and machine learning [20, 21]. Other notable competence-based models have extended Barto’s ideas on autonomous skill learning [70, 71] and formulated goal formation during skill acqui-

tion [72].

The knowledge view and competence view represent two contrasting schools of thought. Favoring the competence view, Mirolli and Baldassarre point out that the *mechanisms* implementing intrinsic motivation might be knowledge-based or competence-based; however, in their view the ultimate *function* of intrinsic motivation is to allow cumulative acquisition of skills rather than knowledge [8]. Along the same lines, Barto argues that even though both predictions and environmental control improve during intrinsically motivated learning, "[t]he utility of prediction arises solely through its role in facilitating control" [22]. Schmidhuber, the most famous proponent of the knowledge view, offers an alternative perspective by interpreting control as the result of behavior aimed at improving predictive models [23].

But which of these approaches is realized in the brain? A likely site for action selection are the *basal ganglia* [73, 74], a group of subcortical nuclei that are also implicated in habit formation and reinforcement learning [75, 76, 77, 78]. Redgrave and colleagues hypothesized that behavioral adaptation takes place in the interface between cortex and basal ganglia and is modulated by the neurotransmitter *dopamine* [79]. Dopamine has been shown to encode the reward prediction error and thereby guide biological reinforcement learning [80, 77, 81]. Apart from reward-related situations, more recent accounts indicate that dopamine is also released in response to unpredicted stimuli [24, 25] as well as novel situations [26, 27]. These factors testify to knowledge-based intrinsic motivation signals in the brain, whereas evidence of biological competence-based learning signals is still lacking [8].

Based on these biological findings, Redgrave and Gurney formulated the *repetition bias hypothesis*, which might explain skill learning in biological agents [28, 29, 30]. Suppose the agent encounters an unpredicted event in the environment. The resulting intrinsic learning signal (dopamine) then leads to a repetition bias, i.e., a behavioral bias to reproduce the actions executed prior to the event. The agent thereby focusses on learning about the event and identify the actions that caused it. As the agent becomes adept at reproducing and predicting the event, the intrinsic learning signal diminishes and the repetition bias is abolished. This hypothesis links contingency learning to knowledge-based intrinsic motivation and is therefore highly relevant to this thesis, which addresses behavioral bias formation during a contingency learning task. Consequently, the models my work is based on are derived from work by Redgrave and Gurney [59] (see Sections 2.3 and 2.4).

Finally, from an evolutionary perspective distinguishing between intrinsic and extrinsic motivation is not straightforward, as discussed by Barto [22]. In a view shared among current theoreticians, intrinsic motivation is linked to "activities that often do not confer obvious evolutionary benefit" [82], where *obvious* indicates utility-related ambiguity of intrinsic and extrinsic motivation. Furthermore, intrinsic motivation "facilitate[s] the cumulative and virtually open-ended acquisition of knowledge and skills that can later be used to accomplish fitness-enhancing goals" [82], which emphasizes the evolutionary relevance of intrinsic motivation. Despite the difficulty in defining concrete evolutionary benefits of intrinsic motivation, Barto argues that "extrinsically rewarding stimuli or events are those that have a relatively immediate and direct relationship to evolutionary

success," whereas "intrinsically rewarding activities, on the other hand, bear a much more distal relationship to evolutionary success" [22]. He concludes that "there is no clean distinction between these types of reward signals; instead, there is a continuum ranging from clearly extrinsic to clearly intrinsic" [22].

On a concluding note, Barto compares this motivation duality to research, where basic research is distinguished from applied research. While basic research is exclusively aimed at furthering our understanding of the world and is therefore intrinsically motivated, applied research is directed toward predefined goals and can thus be seen as extrinsically motivated. Pinpointing the utility of intrinsic motivation is then like justifying basic research in social debates. Just like our intrinsic motivation to control the environment contributed to the evolutionary triumph of the human, basic research generates vast practical benefits, even though they are not a direct consequence of basic research. To conclude this discussion, let us recite Bush, who wrote a compelling exposition of basic research utility in a 1945 report to the United States president [83]:

Basic research is performed without thought of practical ends. It results in general knowledge and an understanding of nature and its laws. This general knowledge provides the means of answering a large number of important practical problems, though it may not give a complete specific answer to any one of them. The function of applied research is to provide such complete answers. The scientist doing basic research may not be at all interested in the practical applications of his work, yet the further progress of industrial development would eventually stagnate if basic scientific research were long neglected. . . . Basic research leads to new knowledge. It provides scientific capital. It creates the fund from which the practical applications of knowledge must be drawn.

1.2 Contingency Learning in Infants

Infants' motor skills develop slowly, which constrains their interaction with the environment [31, 32]. However, they possess reflexes and simple motor behaviors, which are used to study how and when infants learn contingencies between their own actions and sensory events.

Using the sucking reflex, it has been shown that young infants and even newborns learn to vary their sucking behavior if sensory stimuli are contingent on their sucking [84, 33, 85].

Based on the rooting response, Rochat and Striano were able to show that newborns differentiate between being touched by themselves and by external objects [34]. In the former case, the newborns may have learned the contingency between their limb movements and resulting double touch experiences, i.e., touch sensations on the limb and the touched body part, and thereby learned to predict these experiences. On the other hand, the single touch by external objects is unpredicted and may therefore be more salient.

Leg kicking is at the heart of the mobile task by Rovee-Collier and colleagues [31, 35]. They showed that 2-month-old infants are able to learn the contingency between their leg kicking and the movement of a mobile, which is connected with their legs. Furthermore, Bahrick and Watson provided evidence that 5-month-old infants recognize the contingency between their leg kicking and a live recording of their legs implying implicit knowledge about their self at that age [86].

Studying slightly older infants, Kenward demonstrated that 10-month-olds learn to anticipate visual stimuli whose appearance is contingent on the press of a button [36], which indicates the formation of expectations and predictions.

These reflexes and simple motor behaviors offer a way to study infants' learning abilities on a coarse scale. However, infant motor behavior with more fine-tuned control would be desirable to gain more informative measures of infant learning capabilities. Fortunately, accurate eye coordination develops comparatively early [37, 38], which is why analyzing looking behavior is a standard technique in infant studies [39, 40, 41, 42, 43, 44, 45, 46, 47].

A more recent development has been *eye tracking*, which is automatic tracking of subjects' gaze position using high-frequency sampling cameras mostly operating in the infrared light spectrum. Eye tracking enabled researchers to overcome limitations of conventional analysis of looking behavior such as low resolution off-line coding by humans and is now being widely used in infant studies [48, 46, 49, 50, 51].

Eye tracking has made possible a whole new category of experiments called *gaze-contingent eye tracking*, i.e., eye tracking experiments with stimuli contingent on the subject's gaze position. Even though this class of experiments has been used extensively with adult subjects [52, 53], it has been applied to infant studies only very recently [54, 55, 56, 57, 58].

A particular instance of gaze-contingent eye tracking has been introduced by Wang and colleagues in 2012, the *gaze-contingent paradigm* [1]. Because the experiment modelled in this thesis is based on their work, let us review their study in more detail.

1.2.1 The Gaze-Contingent Paradigm

The idea of the gaze-contingent paradigm is to give infants control of their visual environment by presenting to them an optical switch. Specifically, on the screen they were facing there was a red disc on a white background (one disc version). By fixating the red disc, they triggered a "bing" sound and the appearance of an animal picture after 0.6 s, which disappeared again after 1.5 s. Each trigger led to the presentation of a novel animal picture. These events in their environment were thus contingent on their own behavior, i.e., fixating the disc.

Both the 6- and the 8-month-old subjects frequently triggered new pictures, the 8-month-olds significantly more so than the 6-month-olds. Also, 48% of the saccades from the disc to the picture area were classified as anticipatory gaze shifts as they started within 0.2 s of the stimulus onset [42]. Finally, a significant decrease of the reaction time was observed within the first few triggers. The authors concluded that the infants were indeed able to learn the contingency within few trials.

But did the subjects fixate the disc because they wanted to trigger an animal picture, or did they do it rather because the disc was the only salient object on the screen when no animal picture was present? To answer this question, Wang and colleagues conducted experiment 2, the two disc version. Here, two identical red discs were presented, one on the left and one on the right of the screen; the area where animal pictures appeared was right in between. One of the discs was *functioning*, i.e., fixating it triggered animal pictures like in the first experiment. The other one was *non-functioning* such that fixating it had no effect. The side of functioning and non-functioning discs was counterbalanced across infants. Additionally, the latency between trigger onset and stimulus onset was reduced to 0.45 s, and the picture did not disappear at once but slowly faded out within 17 s. Analyzing the infants' gaze behavior indicated a behavioral preference for the functioning disc over the non-functioning one, so their gaze was not purely driven by visual salience but was related to the discs' functionality. This implies that the infants did indeed acquire the contingency.

Finally, Wang and colleagues conducted experiment 2 with adults to assess how much the infants' gaze behavior during the experiment is influenced by cognitive aspects such as expectation and insight. Importantly, in order to create a similar situation as for the infants, the adult subjects were not instructed how to behave during the experiment. Surprisingly, a survey after the experiment revealed that most of the adult subjects (64%) did not acquire the contingency, i.e., they could not explain how the experiment works. However, those who did learn the contingency showed looking behavior similar to the infants', which indicates that the infants gained insight and formed expectations while they participated in the gaze-contingent learning task.

How exactly does this knowledge formation influence the infants' gaze behavior? Specifically, what is the relation between contingency learning and the behavioral preference observed during the experiments? How can these processes be formulated in terms of intrinsic motivation? And which underlying principle can explain the infants' behavior? These are the questions I address in this thesis. Let us now conclude the review of previous work and turn to the description of the methods.

2 Methods

2.1 The Experiment

The experiment is a modified version of the two disc version of the gaze-contingent paradigm introduced by Wang et al. [1] (see Section 1.2.1). 6-, 8-, and 10-month-old subjects were facing a white screen with two identical red discs (see Fig. 2.1a), one of them functioning, the other one non-functioning. Whenever the subject fixated the functioning disc, an animal picture appeared in the center of the screen within about 80 ms and faded out within 17 seconds. Fixations of the non-functioning disc had no effect. This test (session) was conducted until either 5 minutes passed, 30 animal pictures appeared in the center, or the subject showed clear signs of discomfort. After the initial test (T0), the test was repeated after about 19 minutes (T1) and one week (T2).

To test the effect of active control, some subjects faced a yoked control version of the experiment, in which there was no functioning disc. Instead, these subjects were watching a screen recording of a gaze-contingent test of a previous subject. If this yoked control version was presented, it was during T0 and/or T1. Subjects were then part of the active-active group (only gaze-contingent tests), active-yoked group (yoked control during T1), yoked-active group (yoked control during T0), or yoked-yoked group (yoked control during T0 and T1). Because this thesis focusses on behavior during gaze-contingent learning and not on the effect of inconsistent experiences, we shall consider only active-active (all tests) and active-yoked (only T0) subjects.

After data processing (see below), the final sample then consisted of 19 6-month-olds, 27 8-month-olds, and 25 10-month-old infants (71 in total). Table 2.1 lists the numbers of all analyzed subjects and tests.

Age in months	Subjects	Tests		
		T0	T1	T2
6	19	15	5	9
8	27	21	9	7
10	25	17	7	10
all	71	53	21	25

Table 2.1: Size of analyzed samples.

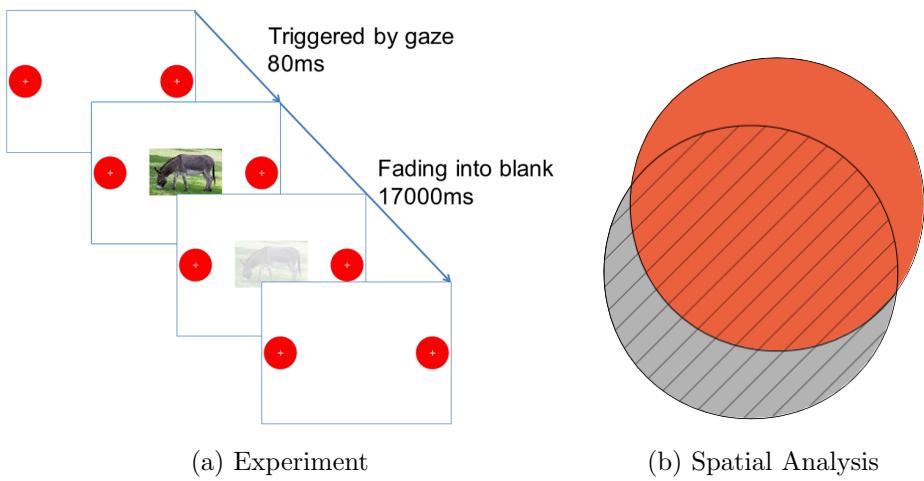


Figure 2.1: (a) Schematic of the stimulus set-up during the gaze-contingent learning task. Once the subject fixates the functioning disc, an animal picture appears within 80 ms and fades out within 17 seconds. Adapted from [1]. (b) Estimating trigger failures. In case of inaccurate eye tracking, there is a discrepancy between where the subject perceives the red disc (striped circle) and where the eye tracker defines it (red disc). Intended disc fixations fall within the striped circle and are missed by the tracker if they fall within the gray area (false negatives). If more than 10% of all intended functioning disc fixations are false negatives, the test data are discarded.

2.2 Data Processing

The raw eye tracking data were processed both in the temporal and the spatial domain. First, saccades were excluded from the data if their duration exceeded 200 ms, which indicated fixations outside of the screen area. This way, we excluded from analysis periods during which the subjects did not actively participate in the experiment.

Second, the spatial precision of the eye tracker and the resulting trigger accuracy were assessed to ensure that the analyzed tests were indeed highly gaze-contingent. For each test, the spatial fixation density was estimated using Gaussian kernels (using Scott's method with smoothing factor 0.1 [87], see Fig. 2.2). Based on the assumption that the salient white crosses in the centers of the red discs attract the most fixations within the left and right thirds of the screen, the two maxima within these regions were established as the *subjective disc* centers, i.e., those areas on the screen that fell within the discs from the subject's point of view. Ideally, the subjective discs and the actual discs overlap completely. In case of tracking inaccuracies, these discs may differ and cause the tracker to miss fixations aimed at the discs (false negatives) and erroneously record disc fixations that were really aimed next to the discs (false positives, see Fig. 2.1b). During data post-processing, fixations were relabelled according to the fitted subjective discs. The *trigger failure rate* was defined as the fraction of false negatives out of all fixations within the subjective functioning disc. To ensure that the data comprise only tests during which the subjects experienced a reliable gaze-contingency, all tests with a failure rate above 10% were excluded from the analysis (50.4% of all tests). Table 2.2 lists for how many subjects data of only specific sessions remained.

Data only for tests	Number of subjects
T0	36
T1	6
T2	7
T0, T1	4
T0, T2	7
T1, T2	5
T0, T1, T2	6
total	71

Table 2.2: Number of subjects grouped by session data. T0-only group also includes active-yoked subjects.

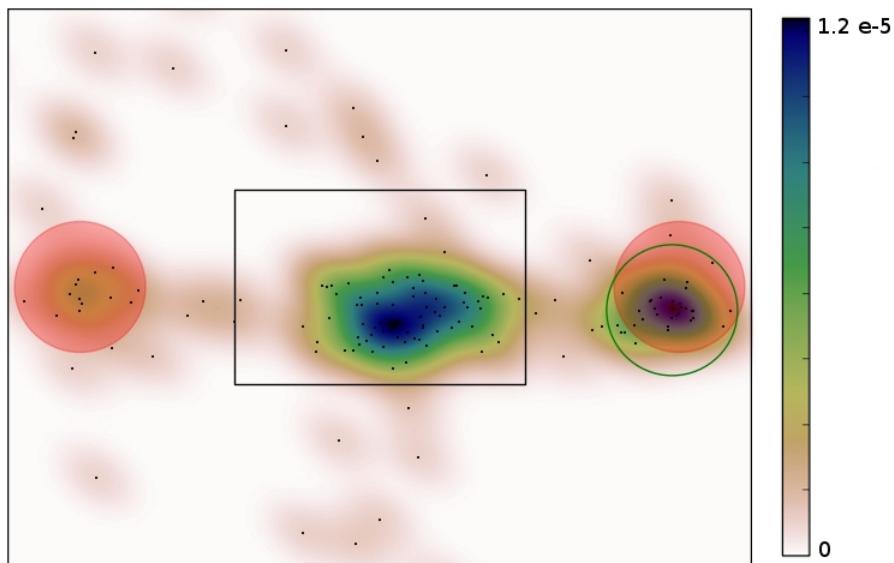


Figure 2.2: Spatial distribution of fixations during a gaze-contingent test. The black dots indicate recorded fixations and the color map the estimated fixation density. The red circles and the central rectangle mark the actual positions of the red discs and the central image area, respectively. The open circle on the right marks the fitted subjective functioning disc fitted around the local maximum in the right portion of the screen. 4 fixations out of 33 within the subjective disc are marked as trigger failures, yielding a failure rate of 12%. Adapted from [88].

2.3 Model 1

To model the subject's behavior and study the underlying processes, I adapted and extended a computational model created by Bolado-Gomez and Gurney [59]. The model is an instance of a prediction-based model of intrinsically motivated action selection (see Section 1.1). However, unlike other models, it includes explicit representations of the biological implementation (neural circuits, see Section 2.3.3). This is because the model was originally created to test the repetition bias hypothesis [28, 29, 30], which makes explicit assumptions about the underlying neural processes. It was then successfully used to model a contingency learning task in rats [89]. Because the repetition bias hypothesis is pertinent to my research question (see Section 1.1), I chose to adapt this model.

I will summarize the original model and detail my own extensions. Parameter values are listed in Section 2.3.6. I refer to [59, 90] for further details about the original model and to [91] for the derivation and fitting of the synaptic plasticity rule.

The core components of the model are a prediction system and a neural circuit model, which is realized by a neural network representing the basal ganglia and associated structures (see Fig. 2.3¹). On a functional level, the model determines which action out of a predefined set of actions to execute in a given context. Processing takes place within parallel action channels. I defined four actions:

- fixate the left disc,
- fixate the right disc,
- fixate the central image area,
- fixate the white background.

2.3.1 The Prediction System

Prediction Adaptation

The prediction system represents the agent's knowledge about the behavior of the environment. Specifically, it generates predictions about future sensory states following specific actions and captures learning by adapting internal belief variables (see Fig. 2.4A): Let $y(t)$ be a scalar sensory feature at time t and $y^*(t)$ its prediction. Then prediction adaptation follows

$$y^*(t + \delta t) = y(t) - \tau_p(y(t) - y^*(t)), \quad (2.1)$$

where τ_p is the prediction adaptation time constant with $0 < \tau_p < 1$ and $\delta t = 0.01$ seconds is one time step.

Here y represents the appearance of an animal picture, so y is binary. In this case, the prediction y^* can be interpreted as the agent's estimated probability that an animal

¹Fig. 2.2 has been discussed in the previous section

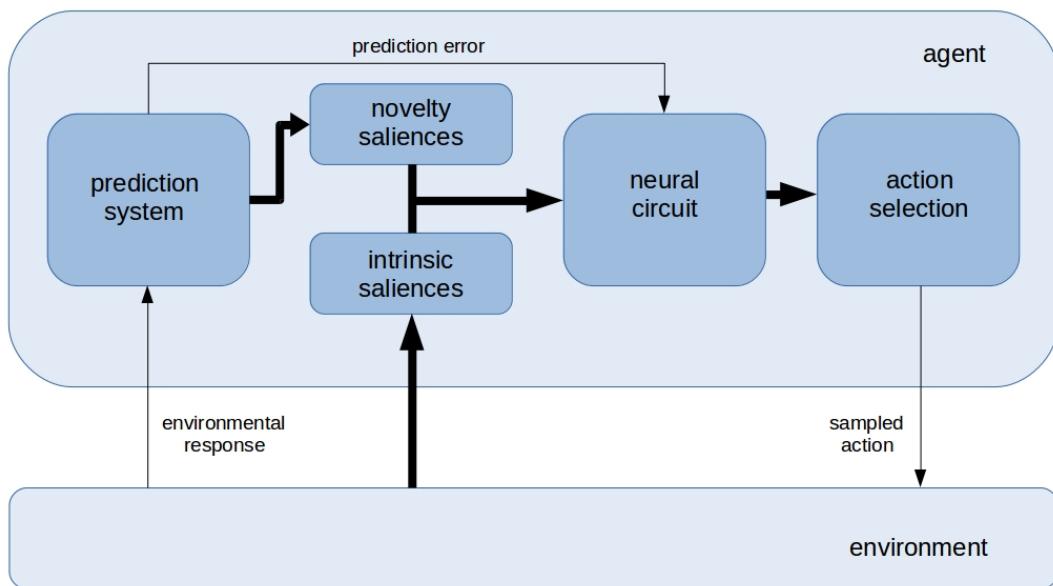


Figure 2.3: Schematic of Model 1. The prediction system stores and processes the internal beliefs about the environment's behavior. It assigns to each possible action a novelty salience, which is combined with intrinsic salience and fed into the neural circuit model. Depending on its output, the current action is chosen and executed. Based on the environment's response, the prediction system is updated and the next iteration begins. Thin arrows denote propagation of scalar values, thick arrows indicate information flow within each parallel action channel.

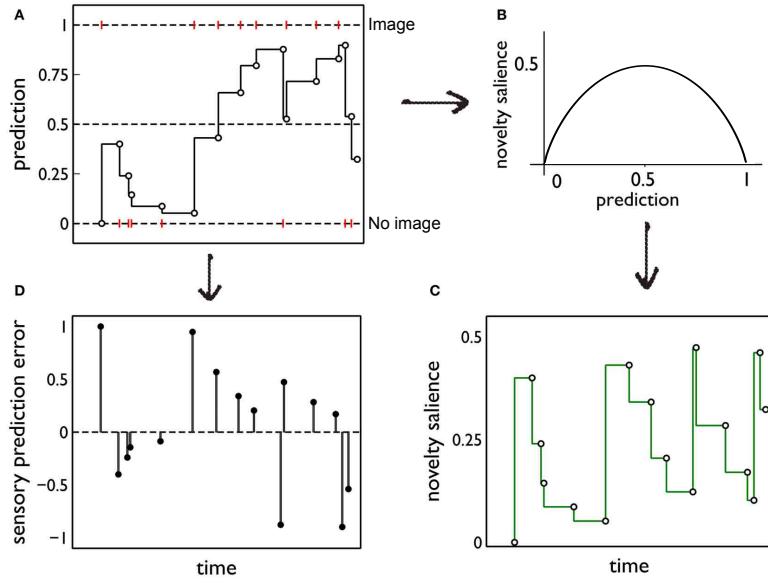


Figure 2.4: Processing within the prediction system. **(A)** Cartoon of a prediction trajectory. Every time the considered action is executed (circles), the prediction variable is updated based on the sensory outcome (red bars). **(B)** Mapping from prediction onto novelty salience. The binary entropy function indicates the expected information gain based on the prediction variable, so uncertain outcomes yield higher saliences. The function is scaled for model consistency. **(C)** The novelty salience corresponding to the prediction variable in **(A)**. Based on the mapping function in **(B)**, this salience reflects the evolution of the expected information gain for this action. **(D)** The sensory prediction error corresponding to the prediction variable in **(A)**. Adapted from [59].

picture will appear, i.e., the degree of belief that this event will happen. Importantly, y^* exists for each considered action, so the agent expects different outcomes after different actions. Specifically, the prediction system stores and updates four variables y_i^* with action index i , one for each action. However, because only one action (fixate the functioning disc) generates a non-zero response y , only the corresponding prediction variable will evolve over time². Thus I refer to this specific prediction whenever I mention prediction adaptation in the following sections and will omit the action index.

All prediction variables are initialized at 0 based on the assumption that initially the subjects did not expect to trigger animal pictures by looking anywhere on the screen. Between test times T0 and T1, the prediction variables are retained, reflecting memory retention during the short break between these tests. However, the prediction variables at the beginning of T2 ($y_{2,b}^*$) are reduced compared to the end of T1 ($y_{1,e}^*$) to take into account forgetting during the seven days between these tests:

$$y_{2,b}^* = k_{ret} y_{1,e}^*, \quad (2.2)$$

where k_{ret} is an individual memory retention factor with $0 < k_{ret} < 1$.

Novelty Salience

For each action, the prediction system maps the internal prediction y^* onto the action's novelty salience s_n . In contrast to the action's intrinsic salience (see below), this novelty salience represents the top-down, experience- and knowledge-based factor of action selection.

Novelty saliences are calculated based on the assumption that actions with uncertain outcomes are more salient than actions with predictable outcomes. Therefore, the mapping function $s_n(y^*)$ has a maximum at $y^* = 0.5$ (maximal uncertainty) and minima at $y^* = 0$ and $y^* = 1$ (minimal uncertainty). Based on these constraints, the original model assumed a piecewise linear dependence between novelty salience and prediction. I replaced this piecewise linear function by the binary entropy function, which is defined as the expected amount of information generated by a Bernoulli process with outcome probabilities p and $1 - p$ and random variable X (see Fig. 2.4B):

$$H_b(X) = -p \log_2 p - (1 - p) \log_2(1 - p). \quad (2.3)$$

Here X is the binary sensory state y , and p is the agent's estimated outcome probability y^* . To retain the original scaling, a prefactor is added, resulting in:

$$s_n(y^*) = -\frac{1}{2} \left(y^* \log_2 y^* + (1 - y^*) \log_2(1 - y^*) \right). \quad (2.4)$$

Ultimately, the novelty salience reflects the expected information gain per action based on the agent's current beliefs. This approach to implement intrinsic motivation, i.e., intrinsic motivation based on information maximization, has been proposed by Schmidhuber and others [19, 92, 93].

²This would be different for the yoked condition. However, that case can be ignored as I'm only modelling the active condition.

Sensory Prediction Error

Apart from salience generation, the prediction system also generates prediction error signals (see Fig. 2.4D), which are directly fed into the neural circuit model. They represent dopaminergic signals, which have been implicated in prediction errors of rewards [80] as well as of sensory events [94]. In the basal ganglia model, they affect synaptic learning in the cortico-striatal interface (see below).

The sensory prediction error e is evaluated whenever an action has been executed and is given by:

$$e(t) = y_i(t) - y_i^*(t), \quad (2.5)$$

where i is the index of the executed action.

2.3.2 Intrinsic Salience

In contrast to novelty salience (see above), intrinsic salience summarizes the bottom-up, sensation-based aspects of action selection. For each action there is an intrinsic salience value s_i , which may change over time and determine the probability to execute this action in the absence of knowledge and expectations.

In the case of fixations, the intrinsic salience per action directly reflects the visual salience of the corresponding fixation area. Therefore, s_i of the central image area has the same initial value as the white background, is set to a high value once an image is triggered and decreases linearly ($s_i(t + \delta t) = s_i(t) - k_{im}$) thereafter until s_i of the background is reached after 17 seconds to simulate the decreasing image contrast while the image is fading out.

Additionally, habituation is taken into account by decreasing the intrinsic salience only of action i whenever action i has been executed: $s_i(t + \delta t) = \tau_h s_i(t)$ with habituation time constant $\tau_h < 1$. Between the tests at T0 and T1 the intrinsic salience values are retained, whereas they are reinitialized at the start of T2 based on the assumption that the subjects dishabituate during the week-long break between T1 and T2 but not during the short break between T0 and T1.

Finally, the intrinsic saliences are added to the corresponding novelty saliences and fed into the neural circuit model.

2.3.3 The Neural Circuit Model

Structure and Dynamics

The neural circuit model incorporates a set of parallel neural networks (channels) representing processing loops through cortex, thalamus, basal ganglia, and brainstem. Each channel corresponds to one action.

The network units represent neuron populations, which process the saliences of the corresponding action. They are instantiated by leaky integrator units, whose activation

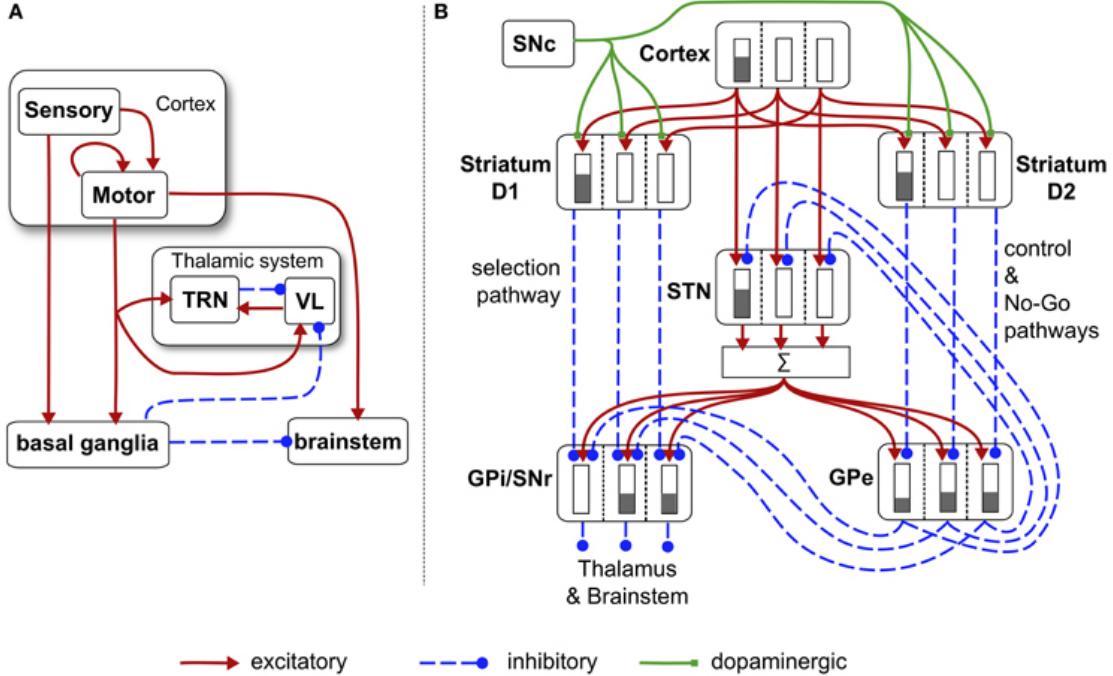


Figure 2.5: (A) Schematic of the overall neural circuit model indicating the sensory and motor cortex, the thalamic system comprising the thalamic reticular nucleus (TRN) and the ventrolateral thalamus (VL), and the basal ganglia and brainstem. Each connection here includes four parallel action channels. (B) Schematic of the basal ganglia network model indicating subthalamic nucleus (STN), globus pallidus internal and external segment (GPi and GPe), substantia nigra pars reticulata and pars compacta (SNr and SNC). Compartments represent action channels with population activity level (gray bars). Note that four action channels are present in my model implementation. Adapted from [59].

variable a follows:

$$\tau_m \frac{da}{dt} = -a(t) + I(t). \quad (2.6)$$

Here τ_m is the characteristic membrane time constant (40 ms), and I is the summed, weighted input.

The corresponding normalized firing rate r is computed using a piecewise linear squashing function with threshold ϵ :

$$r(a) = \begin{cases} 0 & a \leq \epsilon \\ a - \epsilon & \epsilon < a \leq 1 + \epsilon \\ 1 & a > 1 + \epsilon. \end{cases} \quad (2.7)$$

Overall, the network model has a feedforward structure: input signals arrive in the cortex and ultimately lead to motor output via the brainstem (see Fig. 2.5). The salience

generators activate cortical units. These units excite the basal ganglia via cortico-striatal projections onto both neuron populations expressing dopamine receptor type 1 (D1-striatum) and those expressing dopamine receptor type 2 (D2-striatum). These cortico-striatal projections are subject to dopamine-dependent synaptic plasticity (see below). The cortex also excites the subthalamic nucleus (STN), which sums its input across channels and excites both the external and internal segments of the globus pallidus (GPe and GPi) as well as the substantia nigra pars reticulata (SNr). In line with the focusing model of action selection [95], the D1-striatum and D2-striatum form inhibitory pathways that converge onto the output nuclei, i.e., the GPi and SNr. Activity in the D1-striatum leads to selection of a specific action whereas activity in the D2-striatum inhibits the other action channels. Finally, the GPi and SNr project onto the thalamus and brainstem, where the final selection and execution of the action takes place.

Synaptic Learning

The cortico-striatal projections are the only plastic connections in the model and receive modulatory dopamine signals from the substantia nigra pars compacta (SNc), whose activity is determined by the prediction system (see Section 2.3.1). The learning rule implemented in the cortico-striatal weights is derived from Gurney's dopamine-dependent plasticity rule, which is based on biological data [96]. In terms of rate-coded neurons, the expected weight change $\langle dw/dt \rangle$ takes the form:

$$\left\langle \frac{dw}{dt} \right\rangle = A_3 \tau^+ \tau^y y (y - \theta_{BCM}) x \quad (2.8)$$

$$\theta_{BCM} = \langle y^2 \rangle C_{BCM} \quad (2.9)$$

$$C_{BCM} = \frac{-(A_-^{D1/D2} \tau^- + A_+^{D1/D2} \tau^+)}{A_3 \tau^+ \tau^y} \quad (2.10)$$

where x and y are the pre- and postsynaptic firing rates, τ^+ and τ^- time constants associated with post-pre and pre-post firing, τ_y a time constant associated with spike triplets, A_3 a plasticity factor for triplet timing, and $A_+^{D1/D2}$ and $A_-^{D1/D2}$ factors associated with post-pre and pre-post firing for D1-type and D2-type striatal neurons, respectively.

The dopamine dependence is incorporated in the $A_+^{D1/D2}$ and $A_-^{D1/D2}$ factors. These factors were fitted to *in vitro* data [96] using plasticity coefficients $A_\pm^{D1/D2(\text{hi})}$ and $A_\pm^{D1/D2(\text{lo})}$ signifying extreme ("high" and "low") levels of dopamine [97]. Values for intermediate levels of dopamine d are interpolated by a blending function $\alpha(d)$:

$$\alpha(d) = \frac{4d}{1 + 4d}. \quad (2.11)$$

For instance, the dopamine-dependent contribution of positive spike pair timing in D1-type neurons is given by

$$A_+^{D1}(d) = \alpha(d) A_+^{D1(\text{hi})} + (1 - \alpha(d)) A_+^{D1(\text{lo})}. \quad (2.12)$$

The factors $A_{-}^{D1}(d)$, $A_{+}^{D2}(d)$, and $A_{-}^{D2}(d)$ are defined similarly.

Finally, the plastic weights are modified according to

$$w(t + \Delta t) = w(t) + \eta \left\langle \frac{dw}{dt} \right\rangle \Delta t, \quad (2.13)$$

where η is the synaptic learning rate.

2.3.4 Action Selection

Action selection is implemented in a stochastic fashion to allow for exploratory behavior. Based on the output activities $r_{o,i}$ of the neural circuit model, the action probability distribution $p(i)$ is calculated using a softmax function:

$$p(i) = \frac{\exp\left(\frac{r_{o,i}}{k_{\text{exp}}}\right)}{\sum_j \exp\left(\frac{r_{o,j}}{k_{\text{exp}}}\right)}, \quad (2.14)$$

where i is the action index and $k_{\text{exp}} > 0$ is the individual exploration rate. The action is selected by sampling from this probability distribution.

2.3.5 Environment

The environment model simulates the experiment itself, i.e., the responses of the environment to each of the subject's action. Because only fixations on the functioning disc trigger a response, $y = 0$ whenever another area is fixated. Taking into account the finite failure probabilities p_f of the eye tracking system (see Section 2.2), $y = 0$ with probability p_f and $y = 1$ with probability $1 - p_f$ after each fixation on the functioning disc. Also, successful image triggers modify the intrinsic salience of the central image area (see Section 2.3.2).

2.3.6 Parameter and Simulation Details

The parameters listed in Table 2.3 are treated as free model parameters that are fitted to each subject (see Section 2.5). The parameter values listed in Tables 2.5 and 2.6 are the original values used in [59].

Simulated time steps cover 500 ms, which is about the average time period between fixations of the subjects. Simulations of each test were run until either 5 simulated minutes passed or 30 images were triggered by the agent.

Parameter	Description
τ_p	prediction adaptation time constant
k_{ret}	memory retention factor
η	synaptic learning rate
k_{exp}	exploration rate

Table 2.3: Free parameters subject to fitting.

Parameter	Value	Description
$\tau_{h,im}$	0.81	habituation time constant of central image area
τ_h	0.77	habituation time constant of other areas
$s_{i,im}$	0.49	intrinsic salience of new animal picture
		initial intrinsic salience of
$s_{i,L}(0)$	0.1	left disc
$s_{i,R}(0)$	0.1	right disc
$s_{i,W}(0)$	0.01	white background
$s_{i,C}(0)$	0.01	central image area

Table 2.4: Parameters of intrinsic saliences.

Parameter	Value	Description
τ_m	40 ms	membrane time constant
ϵ_S	0	sensory cortex activation threshold
ϵ_M	0	motor cortex activation threshold
ϵ_{TRN}	0	TRN activation threshold
ϵ_{VL}	0	VL activation threshold
ϵ_{BS}	0	brainstem activation threshold
ϵ_{D1}	0.1	D1 cells activation threshold
ϵ_{D2}	0.1	D2 cells activation threshold
ϵ_{STN}	-0.25	STN activation threshold
ϵ_{GPe}	-0.2	GPe activation threshold
ϵ_{GPi}	-0.075	GPi activation threshold
ϵ_{SNC}	-0.198	SNC activation threshold
$w_{S,M}$	1.0	weight from sensory to motor cortex
$w_{VL,M}$	1.0	weight from VL to motor cortex
$w_{M,TRN}$	1.0	weight from motor cortex to TRN
$w_{VL,TRN}$	1.0	weight from VL to TRN
$w_{M,VL}$	1.2	weight from motor cortex to VL
$w_{GPi,VL}$	1.0	weight from GPi to VL
$w_{TRN,VL}^w$	0.1	weight from TRN to VL (within)
$w_{TRN,VL}^b$	0.3	weight from TRN to VL (between)
$w_{S,D1/D2}^0$	0.5	initial weights from sensory cortex to striatum
$w_{M,D1/D2}^0$	0.5	initial weights from motor cortex to striatum
$w_{S,STN}$	0.3	weight from sensory cortex to STN
$w_{M,STN}$	0.5	weight from motor cortex to STN
$w_{GPe,STN}$	0.2	weight from GPe to STN
$w_{STN,GPe}$	0.3	weight from STN to GPe
$w_{D2,GPe}$	0.9	weight from D2-striatum to GPe
$w_{STN,GPi}$	0.33	weight from STN to GPi
$w_{D1,GPi}$	0.75	weight from D1-striatum to GPi
$w_{GPe,GPi}$	0.4	weight from GPe to GPi
$w_{SC,SNC}$	4.0	weight from SC to SNC
$w_{M,BS}$	1.0	weight from motor cortex to brainstem
$w_{GPi,BS}^+$	0.7	additive weight from GPi to brainstem
$w_{GPi,BS}^x$	2.0	multiplicative weight from GPi to brainstem

Table 2.5: Parameters of the neural circuit model.

Parameter	Value	Description
τ^+	38 ms	time constant for positive spike pair timing
τ^-	38 ms	time constant for negative spike pair timing
τ^y	45 ms	time constant for spike triplets
A_3	0.1	plasticity coefficients for spike triplets
$A_+^{D1(\text{lo})}$	-0.2	positive spike pair timing, D1-striatum, low dopamine level
$A_+^{D1(\text{hi})}$	1.27	positive spike pair timing, D1-striatum, high dopamine level
$A_-^{D1(\text{lo})}$	-0.8	negative spike pair timing, D1-striatum, low dopamine level
$A_-^{D1(\text{hi})}$	-0.15	negative spike pair timing, D1-striatum, high dopamine level
$A_+^{D2(\text{lo})}$	0.4	positive spike pair timing, D2-striatum, low dopamine level
$A_+^{D2(\text{hi})}$	0.25	positive spike pair timing, D2-striatum, high dopamine level
$A_-^{D2(\text{lo})}$	0.3	negative spike pair timing, D2-striatum, low dopamine level
$A_-^{D2(\text{hi})}$	-1.5	negative spike pair timing, D2-striatum, high dopamine level

Table 2.6: Parameters of the cortico-striatal plasticity.

2.4 Model 2

Model 2 is a simplified version of Model 1 (see Fig. 2.6). It lacks the neural circuit model and is therefore less biologically realistic but more focussed on the functionality of prediction adaptation.

Action selection in Model 2 depends on the summed novelty and intrinsic saliences instead of the neural circuit model output. The used parameter values are listed in Table 2.8. All other parts of Model 2 are identical to Model 1.

Parameter	Description
τ_p	prediction adaptation time constant
k_{ret}	memory retention factor
k_{exp}	exploration rate

Table 2.7: Free parameters subject to fitting.

Parameter	Value	Description
$s_{i,L}(0)$	0.1	initial intrinsic salience of left disc
$s_{i,R}(0)$	0.1	initial intrinsic salience of right disc
$s_{i,W}(0)$	0.01	initial intrinsic salience of white background
$s_{i,C}(0)$	0.01	initial intrinsic salience of central image area
$s_{i,\text{im}}$	0.5	intrinsic salience of new animal picture
τ_h	0.8	habituation time constant of all areas

Table 2.8: Parameters of intrinsic saliences.

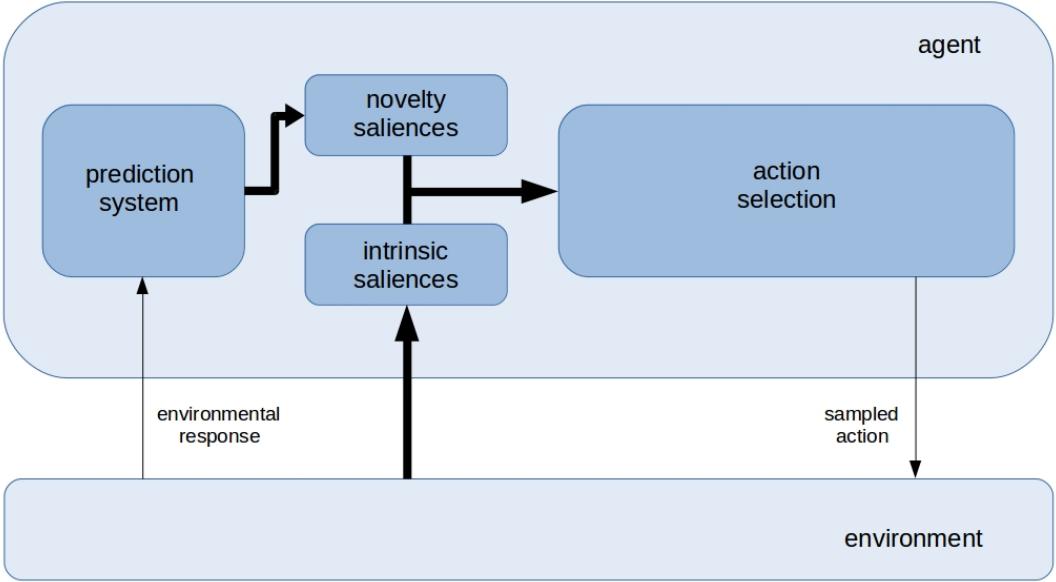


Figure 2.6: Schematic of Model 2, which is essentially Model 1 without the neural circuit model (see Fig. 2.3, same conventions).

2.5 Model Fitting

For both models, the free parameters were fitted to each subject using Covariance Matrix Adaptation–Evolution Strategy (CMA-ES) [60], a stochastic optimization algorithm.

Each sampled set of parameters was evaluated by simulating the model and extracting the bias (see Section 3.1) from the simulated agent’s behavior during the three tests. The sample’s fitness error f was then:

$$f = \sum_{T=0}^2 c_T (b_T^{\text{exp}} - b_T^{\text{sim}})^2 \quad (2.15)$$

$$c_T = \begin{cases} 0 & \text{if failure rate} > 0.1 \text{ during test } T, \\ 1 & \text{else,} \end{cases} \quad (2.16)$$

where b_T^{exp} is the empirical bias of the subject during test T , i.e., the bias extracted from the experimental data, and b_T^{sim} the simulated agent’s bias during test T , i.e., the bias extracted from the simulation data.

The fitting was terminated as soon as the best sample’s fitness error fell below 0.1 times the number of tests with a failure rate of at most 0.1. This sample’s parameters were then accepted as the fitting results for that subject. Otherwise, if the parameters or fitness error converged, or the conditioning number of the covariance matrix diverged, the parameters were reset and the step sized increased by 0.05 up to a maximum of 0.5.

After 100 search reinitializations the best set of parameters was selected as the fitting solution.

The sampling parameters are listed in Table 2.9. The population size λ was 99, and the other hyperparameters were set as recommended in [98].

Parameter	Initial Value	Standard Deviation	Description
τ_p	0.9	0.05	prediction adaptation time constant
k_{ret}	0.9	0.1	memory retention factor
η	100	10	synaptic learning rate (only Model 1)
k_{exp}	0.6	0.1	exploration rate

Table 2.9: Sampling parameters during fitting.

3 Results

Conventions

In this chapter's figures, samples are indicated by their means. Error bars indicate the standard errors of the means. In case of box plots, horizontal lines in boxes indicate first quartile, median, and third quartile. Whiskers cover the range of the data up to 1.5 times the interquartile range. Crosses indicate extreme outliers. The mean is marked by squares.

Unless stated otherwise, deviations of sample means are tested using Student's t-test. Significant deviations are graphically indicated by * ($p < 0.05$), ** ($p < 0.01$), or *** ($p < 0.001$). Significant deviations between two samples are indicated by bars. Asterisks without bars indicate a significant deviation of the sample mean from 0.

In case of correlation analysis, the following measures are reported:

- the adjusted R^2 of the regression,
- the Pearson coefficient of correlation,
- the p-value of a correlation t-test,
- the adjusted p-value using the Holm-Bonferroni method [99].

The regression line is plotted along with the corresponding 95% confidence bands.

Outliers of a given sample are defined as all data points lying outside of range

$$[Q1 - 1.5 \cdot IQR, Q3 + 1.5 \cdot IQR], \quad (3.1)$$

where $IQR = Q3 - Q1$ is the sample's interquartile range, and $Q1$ and $Q3$ are its first and third quartile, respectively.

3.1 Behavioral Bias

A first qualitative assessment of the subjects' gaze behaviors revealed a higher concentration of fixations on the functioning disc than on the non-functioning one during the active control sessions (see Fig. 3.1a), indicating a preference for the functioning over the non-functioning disc. During the yoked control sessions, no such preference was observed (see Fig. 3.1b), which implies that the behavioral bias is indeed tied to the functionality of the functioning disc.

To quantify this preference, I applied the notion of *gaze patterns* as introduced in [1] and extended in [88]: A (non-)functioning gaze pattern is defined as a fixation sequence

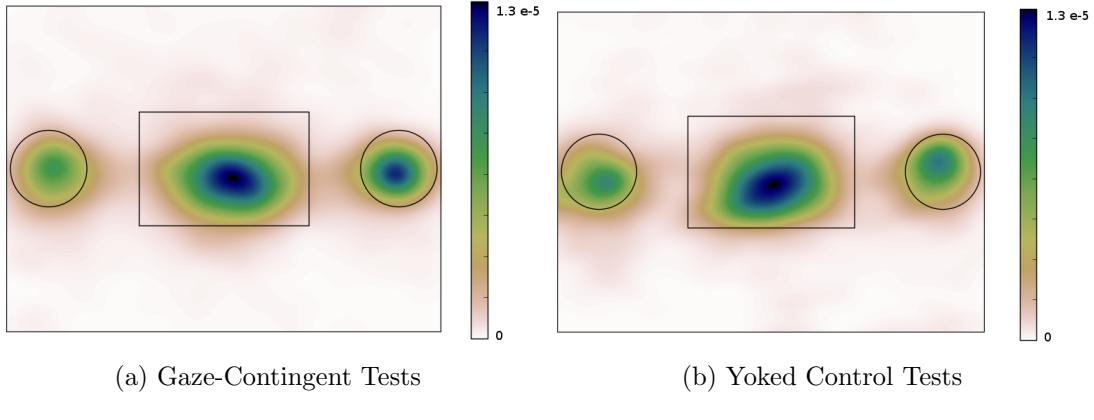


Figure 3.1: Aggregate fixation distribution during all gaze-contingent tests (a) and yoked control tests at T0 (b). The central rectangle indicates the location of the animal pictures, the circles the locations of the red discs. In (a), fixation locations during tests with the functioning disc on the left were mirrored along the central vertical axis such that the functioning disc is visualized on the right. The density estimation is described in Section 2.2. Adapted from [88].

starting in the central image area, leading to the (non-)functioning disc and back to the image area. Because the total experiment durations varied between subjects and tests, the gaze pattern counts were divided by the test durations to yield the normalized *gaze pattern rates* [88]. To take into account the different activity levels of subjects, I finally define the *functioning bias* as the difference between the functioning gaze pattern rate and the non-functioning gaze pattern rate. This scalar quantity is plotted in units $\frac{1}{\text{min}}$ and indicates whether a subject preferred the functioning disc ($\text{bias} > 0$), the non-functioning disc ($\text{bias} < 0$), or showed no preference ($\text{bias} \approx 0$). Thus for all bias samples, the deviations from 0 are analyzed.

The experimental data indicate that the subjects¹ prefer the functioning disc at T0 and tend to increase this preference during repeated tests (see Fig. 3.2a). This functioning bias is highly significant when averaged over all tests.

The models were fitted to these data (see Section 2.5).² The model fits faithfully reproduce the experimental data (see Fig. 3.2b). Figure 3.2c shows the empirical functioning bias as well as the model fits including their estimations for the tests that were excluded from the analysis because of the failure rate estimation (see sections 2.1 and 2.2). Both models estimate a slightly higher bias during T0 for subjects missing in the empirical T0 data, which implies a systematic trend: If a subject would exhibit a high bias, his T0 test is more likely to be rejected due to its failure rate than for a subject with low bias, according to the models. Why this should be so is unclear. These model estimates (or predictions) are the basis for the following analyses.

Note that estimating the retention factor is problematic in cases where T2 data are not present. In this case, the fitting algorithm would be able to estimate an arbitrary

¹Only gaze-contingent tests are considered henceforth.

²Only individual subjects were fitted.

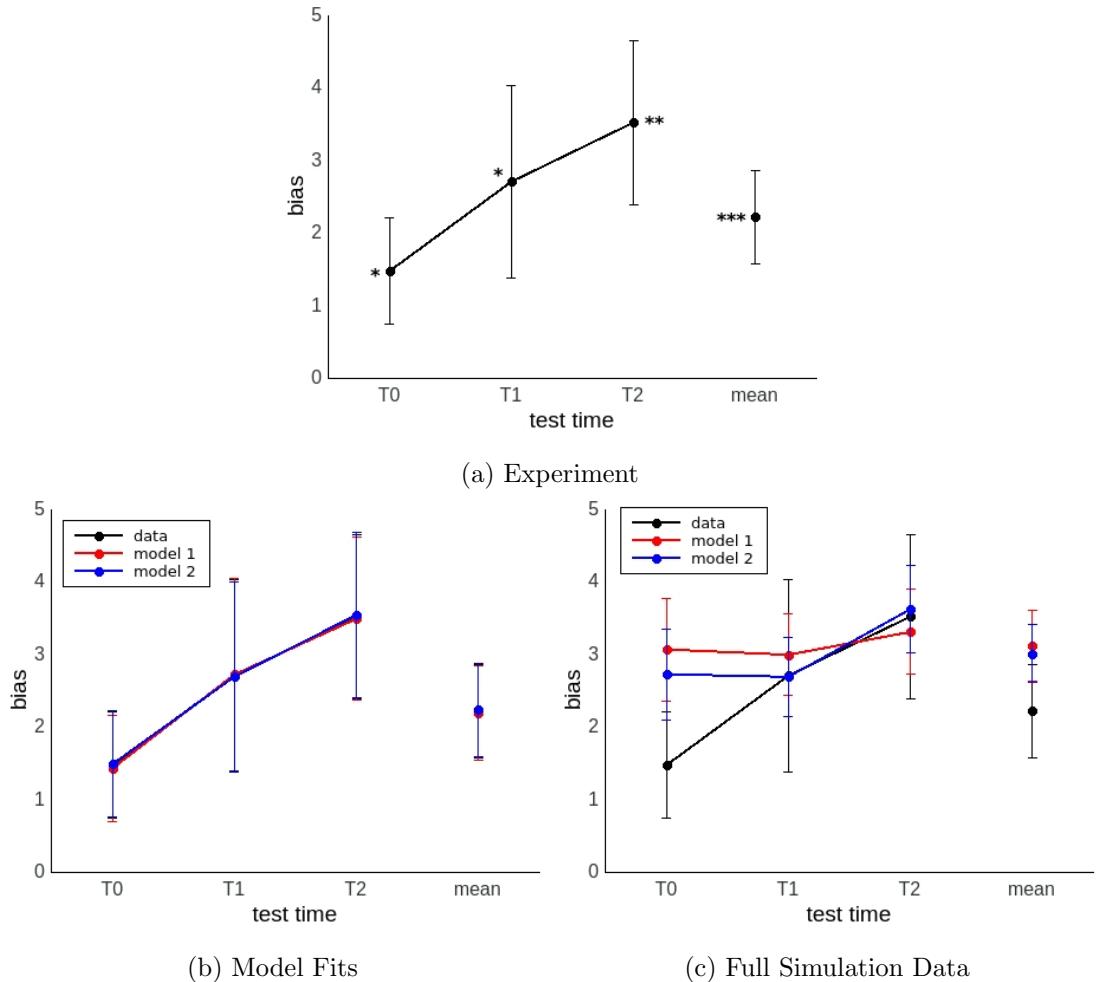


Figure 3.2: (a) Functioning bias of all subjects tested in the gaze-contingent condition. (b) Experimental data and their model fits. (c) Experimental and full simulation data.

value for the retention factor. However, in practice it estimates a value around the initial search value (see Table 2.9). Due to these difficulties, I disregard the estimated retention in the following analyses.

Histograms of the achieved fitness errors are plotted in Figure 3.3.

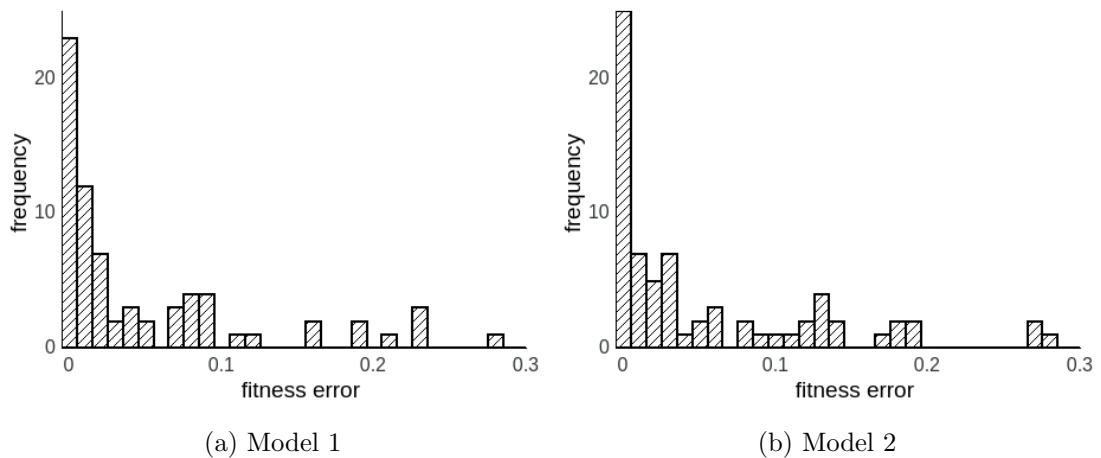


Figure 3.3: Histograms of subjects' fitness errors for model 1 (a) and model 2 (b).

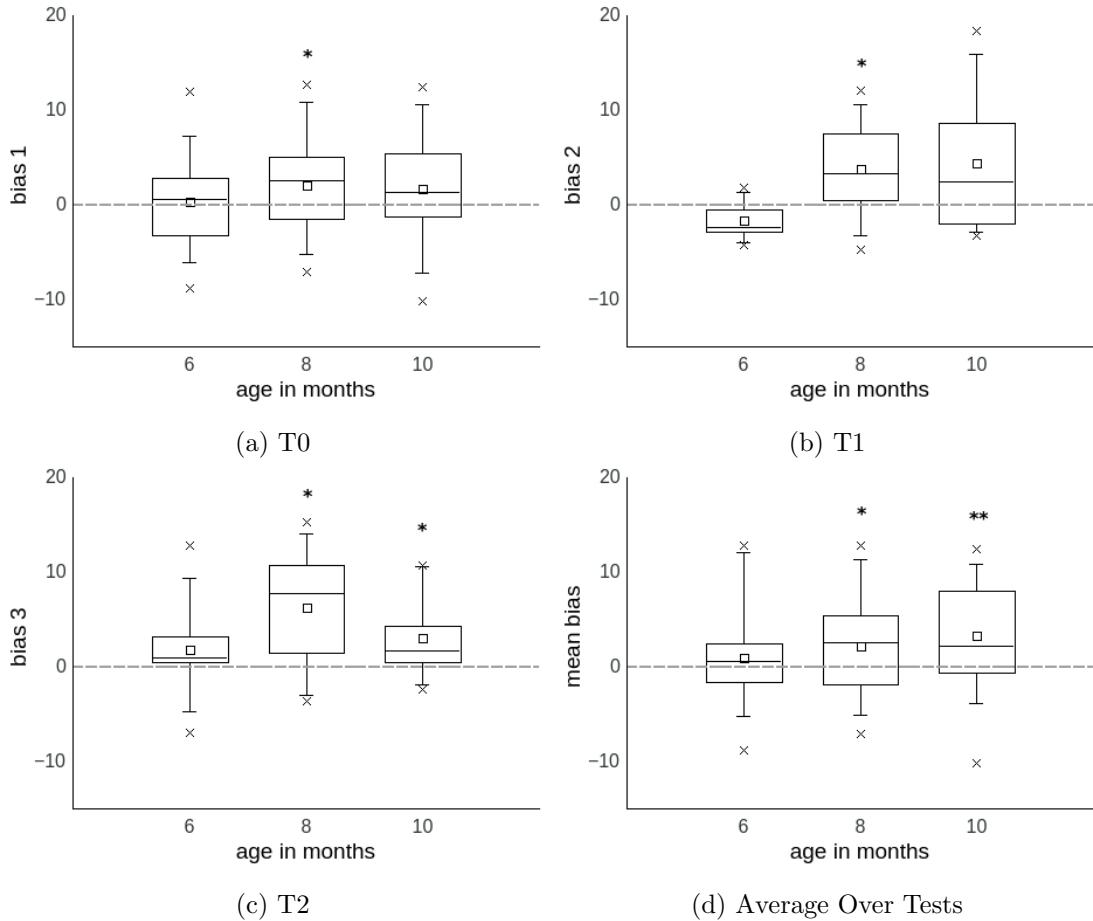


Figure 3.4: Empirical functioning bias of different age groups during individual tests (a-c) and averaged over tests (d). Dashed lines indicate level of unbiased behavior.

3.2 Effect of Age

Analysing the behavior of different age groups reveals that this functioning bias is mostly due to the 8-month-old subjects, who show a consistent bias during all tests (see Fig. 3.4). The 6-month-olds don't show a bias during any test, whereas the 10-month-olds do during T2 and when averaged over tests. Both models estimate a similarly biased behavior of the 8- and 10-month-old subjects and a functioning bias of the 6-month-olds when averaged over tests (see Figs. 3.5 and 3.6).

To study this qualitative difference of the 6-month-olds on the one hand and the 8- and 10-month-olds on the other hand, I grouped the latter and compared the two resulting groups. Empirically, the 6-month-olds show no functioning bias, whereas the older subjects consistently prefer the functioning disc during all tests (see Fig. 3.7). The behavior of the older subjects is reproduced by both models (see Figs. 3.8 and 3.9). Both models agree on a significantly stronger bias of the older subjects compared to the

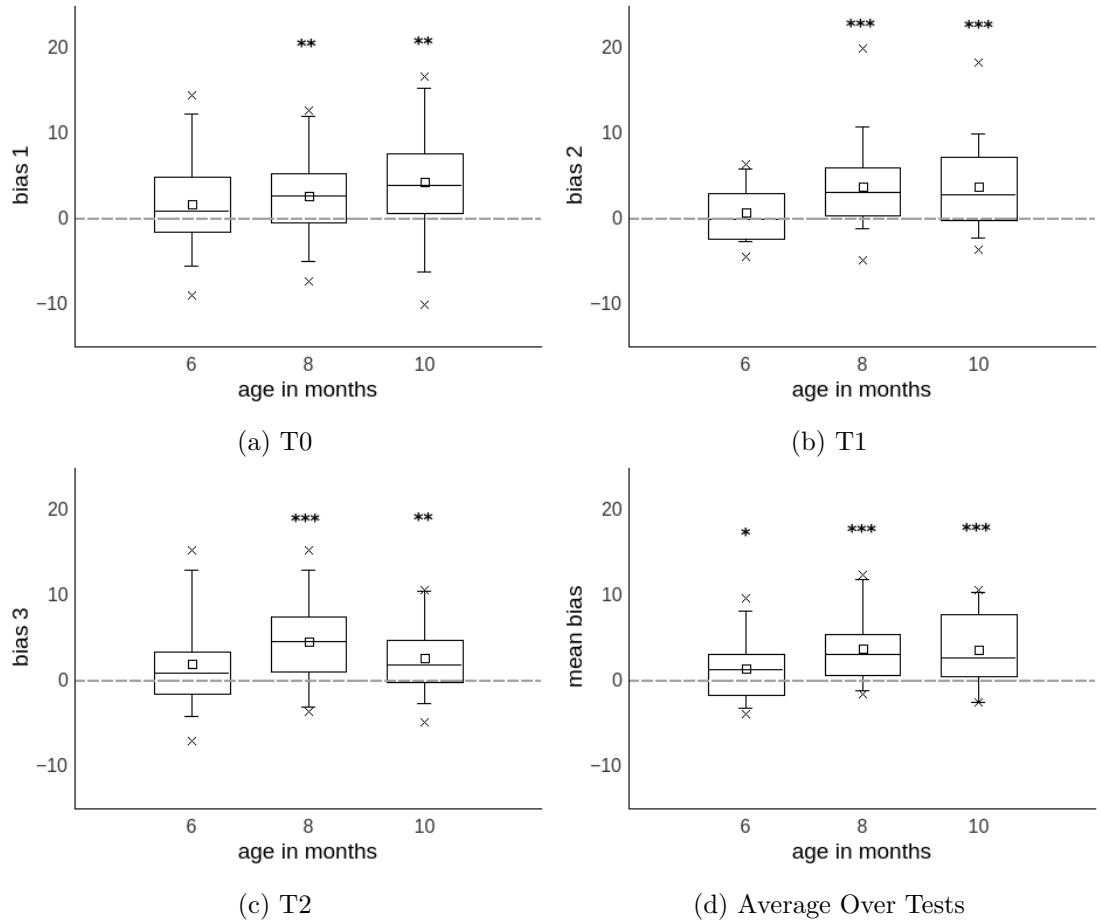


Figure 3.5: Functioning bias of different age groups during individual tests (a-c) and averaged over tests (d) as estimated by model 1. Dashed lines indicate level of unbiased behavior.

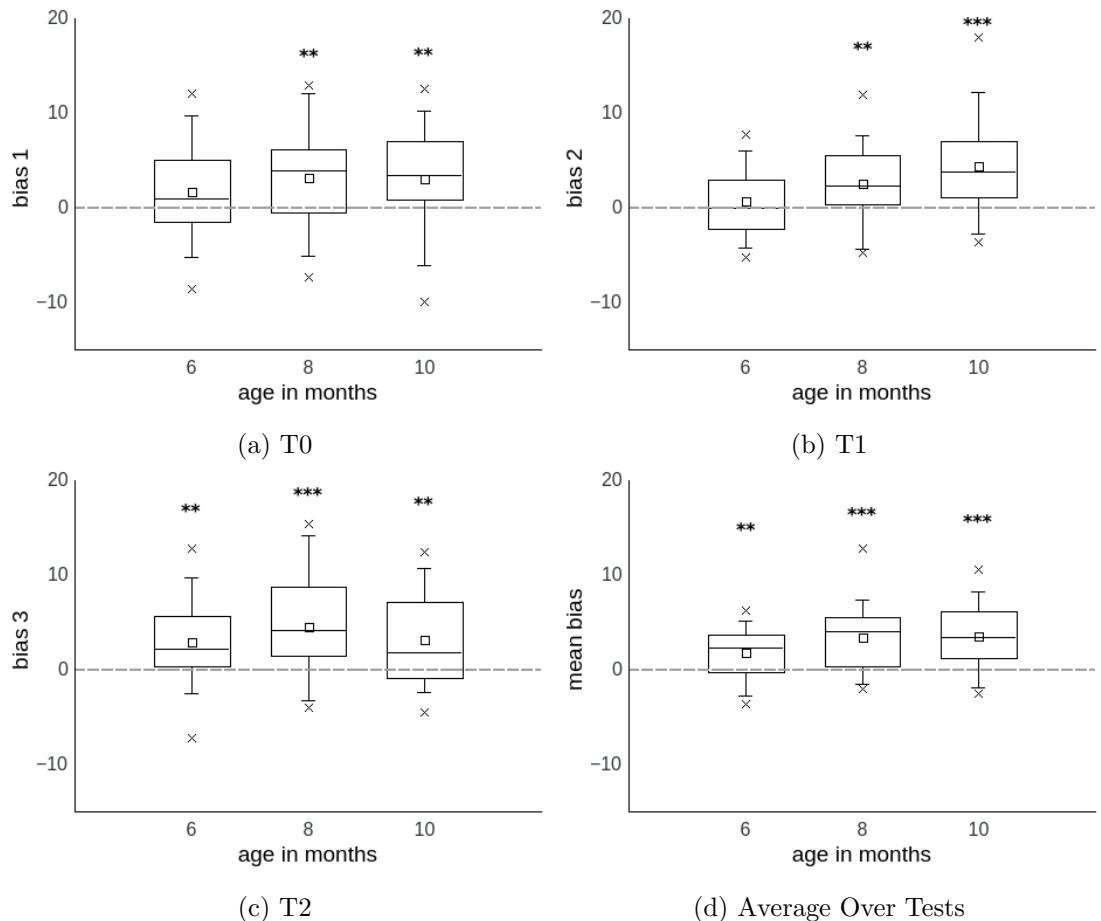


Figure 3.6: Functioning bias of different age groups during individual tests (a-c) and averaged over tests (d) as estimated by model 2. Dashed lines indicate level of unbiased behavior.

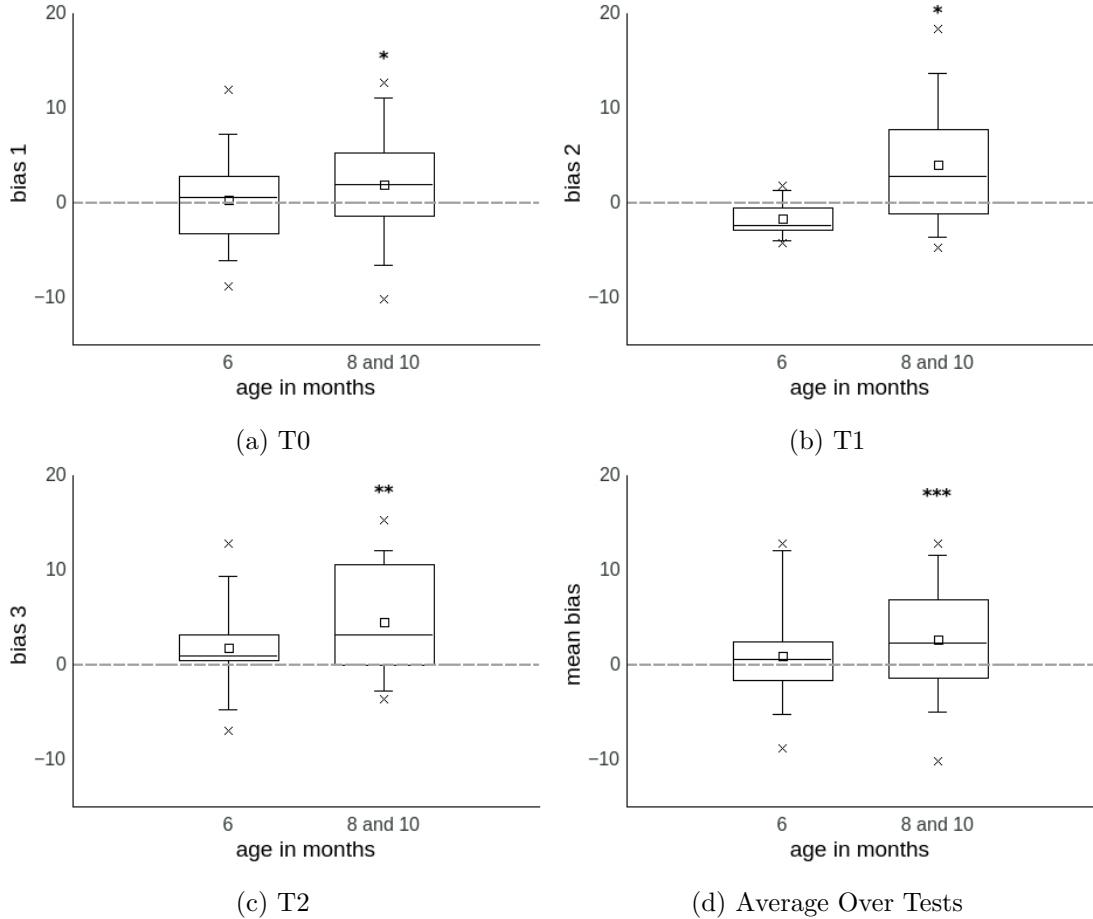


Figure 3.7: Empirical functioning bias of different age groups during individual tests (a-c) and averaged over tests (d). Dashed lines indicate level of unbiased behavior.

6-month-olds during T1. However, Model 1 estimates unbiased behavior of the 6-month-olds during T2 and a significantly lower one overall, whereas Model 2 estimates a similar bias level during T2 and overall. In summary, both models reproduce the consistent functioning bias exhibited by the older subjects. The models indicate that the difference is highest during T1 and predict that also the 6-month-olds show a functioning bias on average.

Differences in simulated behavior should be reflected in the underlying model parameters. Even though slight differences in the fitted parameters exist, none of them are significant (see Figs. 3.10 and 3.11).

Let us turn to studying the functioning bias on the level of individual subjects, which promises deeper insights.

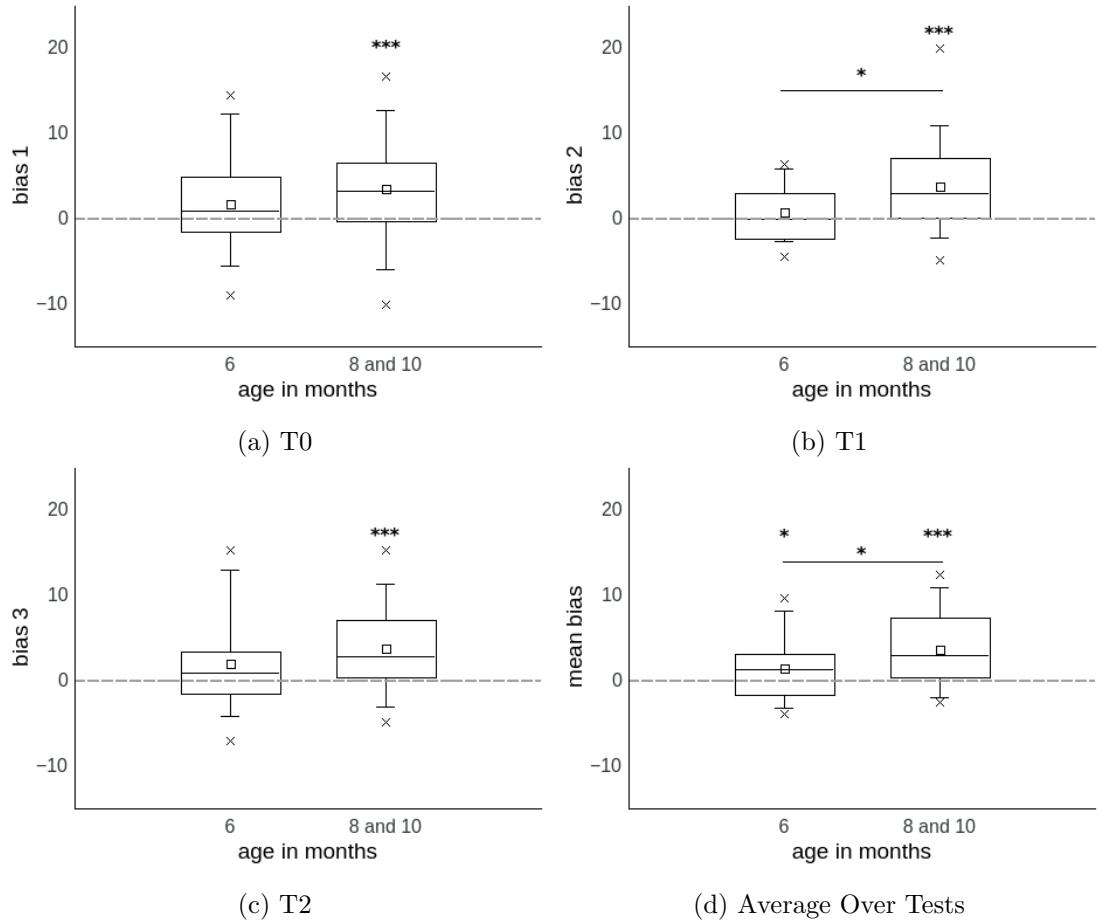


Figure 3.8: Functioning bias of different age groups during individual tests (a-c) and averaged over tests (d) as estimated by model 1. Dashed lines indicate level of unbiased behavior.

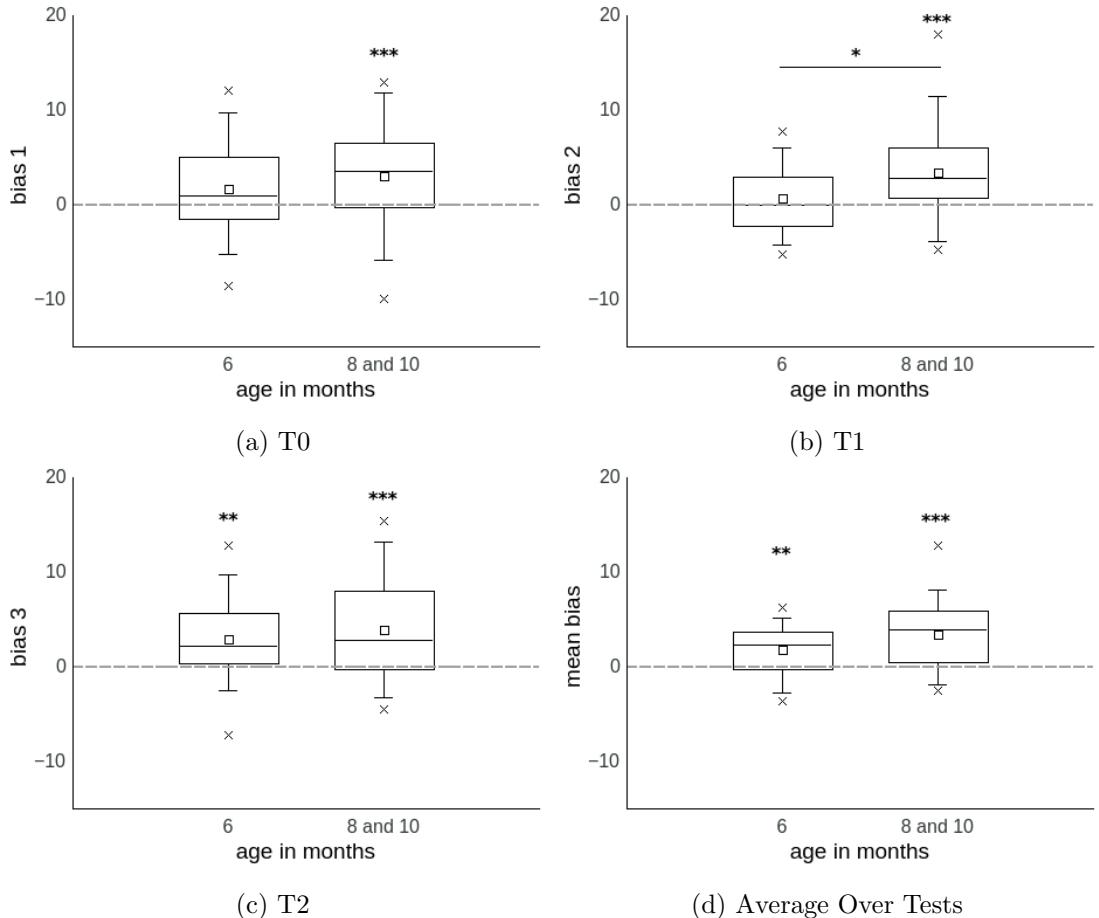


Figure 3.9: Functioning bias of different age groups during individual tests (a-c) and averaged over tests (d) as estimated by model 2. Dashed lines indicate level of unbiased behavior.

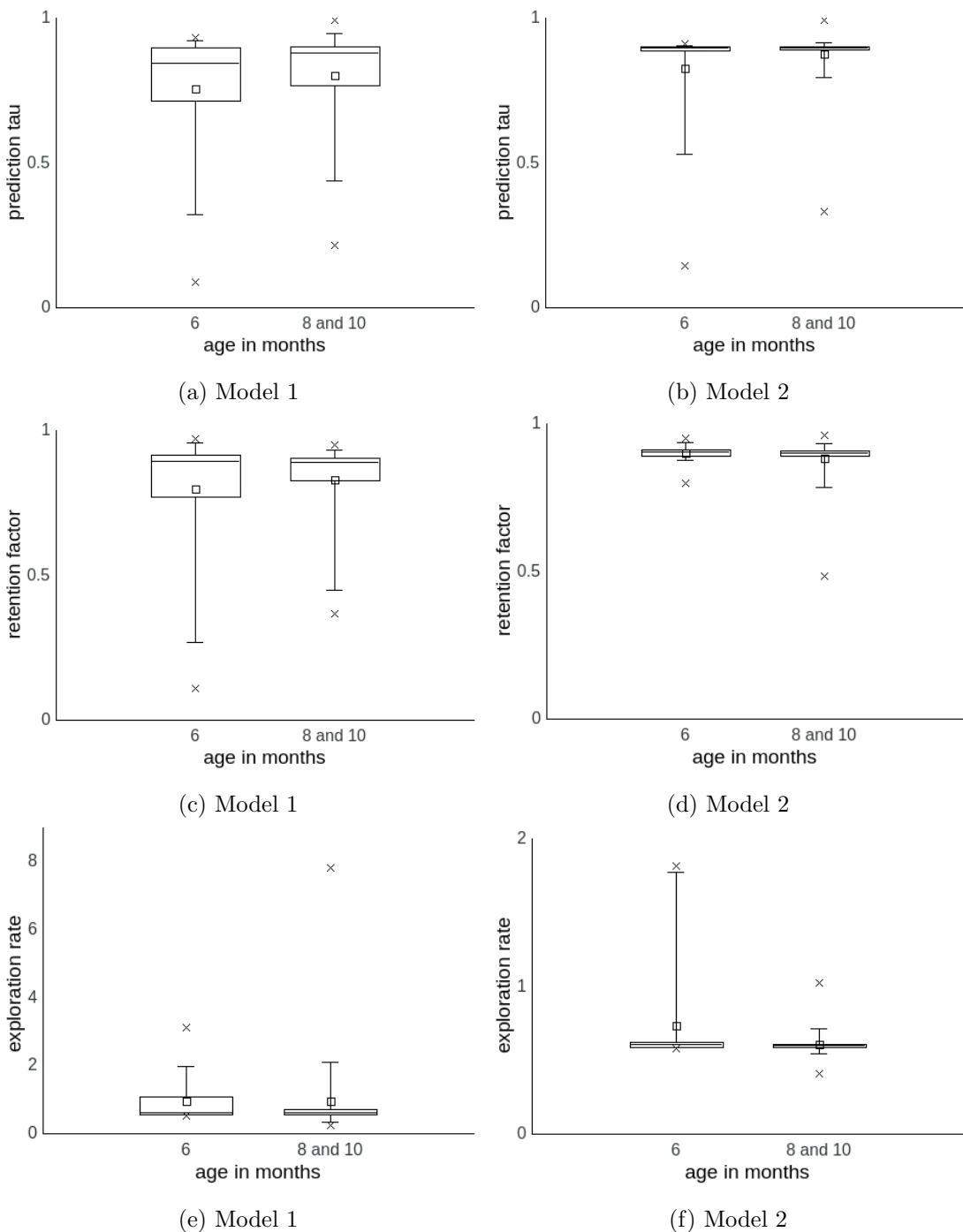


Figure 3.10: Distributions of the fitted parameters τ_p (a,b), k_{ret} (c,d), and k_{exp} (e,f) for 6-month-olds and older subjects.

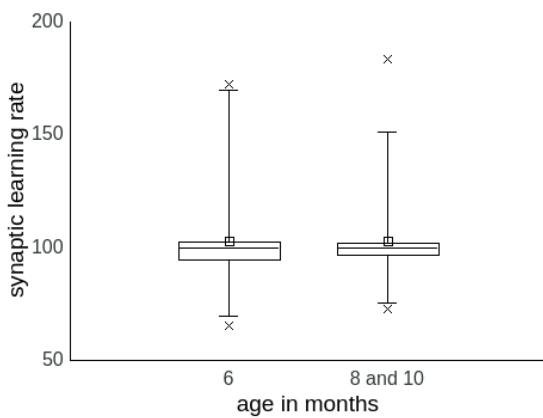


Figure 3.11: Distributions of the fitted synaptic learning rate η for 6-month-olds and older subjects. This parameter is not present in Model 2.

3.3 Individual Learning Progress

The main hypothesis of this study is that subjects acquire the gaze-contingency during the experiments, i.e., they learn to predict the appearance of an animal picture after they fixate the functioning disc. This process is modelled by adaptation of the prediction system. Specifically, the prediction variable y^* converges to 1 during learning (see Section 2.3.1). The more uncertain the agent is about the outcome of the action, i.e., the closer y^* is to 0.5, the higher its associated novelty salience and thus the higher the probability of performing that action. So the resulting behavioral bias is based on the transition of the internal prediction through the uncertain regime (y^* around 0.5) as indicated by the magnitude of the novelty salience.

Figure 3.12 shows the time evolution of the predictions and the corresponding novelty saliences for three simulated subjects representing three types of learning dynamics:

- The *fast learner* (see Fig. 3.12a) quickly acquires the contingency within the first session, as indicated by the rapid convergence of y^* to 1. As a result, the novelty salience quickly rises and falls in the beginning of the first session. The subject's internal model is fully adapted by the time session 2 starts, which is why no learning progress takes place and novelty salience is negligible during session 2. Depending on the memory retention factor, the subject may forget the contingency between T1 and T2, i.e., y^* drops after T1 and quickly rises again during T2. During this readaptation, the novelty salience may peak again and result in another period of biased behavior.
- The *slow learner's* prediction (see Fig. 3.12b) evolves similarly, but on a longer time scale. During T0, y^* is mostly in the uncertain regime, yielding a high level of novelty salience. The contingency is acquired by the end of the second session, which is still largely affected by the slowly decreasing novelty salience. Forgetting between T1 and T2 may again lead to a similar picture during T2 as during T0. In the end, high levels of novelty salience strongly bias the subject's behavior during all sessions, especially during T0 and T2.
- The *non-learner* (see Fig. 3.12c) steadily improves his prediction throughout the sessions but fails to acquire the contingency due to the slow rate of prediction adaptation. Because y^* lies in the uncertain regime for most of the experiments, a high level of novelty salience is quickly reached during T0 and retained for the rest of the sessions, which results in a strong, persistent bias from early on.

Confirming these arguments, both models reproduce and offer explanations for the different levels of bias observed during these subjects' sessions and estimate the bias during the missing sessions (see Fig. 3.13). Note that I chose these three subjects because they represent extreme cases of different predicted learning dynamics and thereby serve as illustrative examples of different types of model behavior. Considering that in fewer than 10% of all cases, subjects performed all three tests with failure rate less than 10% (see Table 2.2), it is not likely to find good examples only within subjects with all three

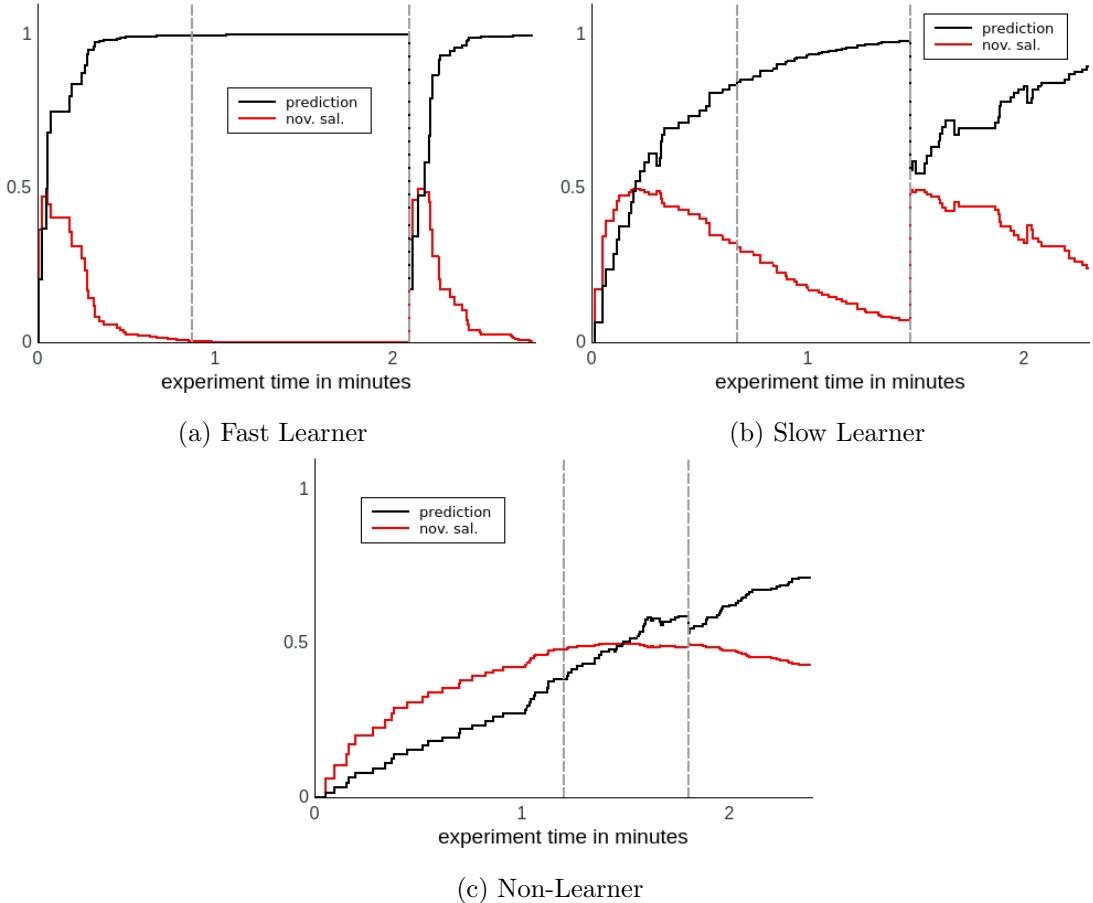


Figure 3.12: Time evolution of the prediction variable y^* (black) and the corresponding novelty salience (red) during simulated experiments for three individual subjects. Vertical dashed lines indicate session onsets. Note that a decrease of y^* is either due to forgetting between T1 and T2 or due to trigger failures (see Section 2.2), which contradict the contingency and lead to transient reversals of the prediction adaptation.

tests remaining in the analysis. It is not required since we are focussing on the model behavior at this point.

Both models estimate the same qualitative behavior for these subjects during the missing sessions, which indicates that the biological component of Model 1 does not have a large impact on the finally chosen action. Note the interesting behavior of the fast learner, who starts with an initial negative bias and then shows a positive bias at T1. As we have just observed, the fast learner is least influenced by novelty salience since acquisition is rapidly over. This implies that he is more influenced by intrinsic salience, which is subject to habituation, which in turn favors equal proportions of fixations on both discs. Since he starts with an excess of fixations on the non-functioning disc (as required by the experimental data), due to habituation its intrinsic salience is much lower at the start of T1 than the one of the functioning disc. This is why both models would agree on a functioning bias for the fast learner during T1. Ultimately, this is an effect due to habituation and this particular initial condition and not much related to learning itself.

To quantify the speed with which each subject transitions through the uncertain regime and acquires the contingency, I introduce two measures, which are extracted from the subject's simulated predictions:

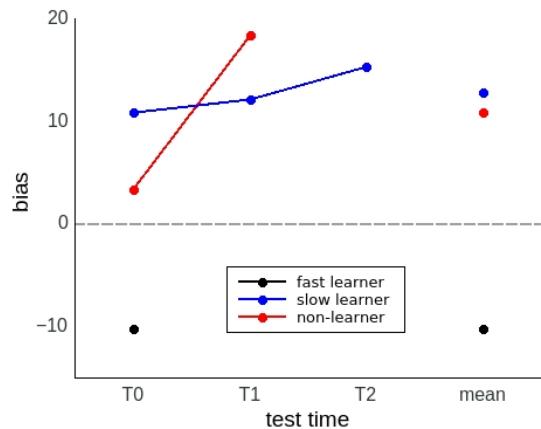
- $t5^3$ is the normalized experiment time at which y^* first exceeds 0.5 (or 50%). This is the moment of maximal uncertainty and thus maximal novelty salience.
- $t9$ is the normalized experiment time at which y^* first exceeds 0.9 (or 90%). $t9$ indicates when a subject has reached a high confidence of predicting the contingent response. Because y^* never reaches 1 (due to the exponential prediction adaptation, see Section 2.3.1), $t9$ may be interpreted as an approximation of the contingency acquisition time.

A time point's *normalized experiment time* is computed by dividing its absolute simulated time by the session's duration, which allows to compare different sessions' time points on a relative scale. The normalized experiment time lies within $[0, 3[$ with:

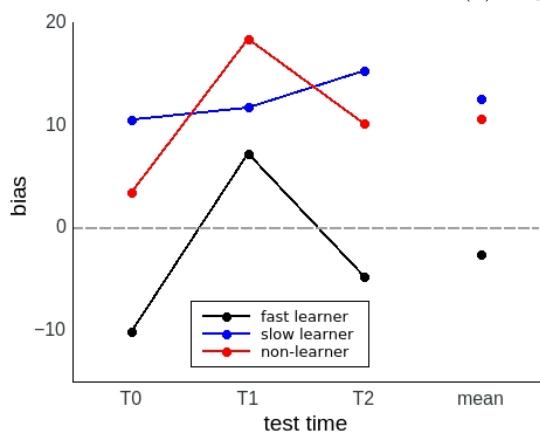
$$\text{normalized experiment time} \in \begin{cases} [0, 1[& \text{during T0,} \\ [1, 2[& \text{during T1,} \\ [2, 3[& \text{during T2.} \end{cases} \quad (3.2)$$

Figure 3.14 visualizes these measures for the fast learner, the slow learner, and the non-learner. In line with the previous discussion, the fast learner exhibits the smallest $t5$ whereas the non-learner exhibits the largest one. Additionally, the fast learner's $t9$ falls within T0 and the slow learner's $t9$ within T1, confirming the previously indicated times of contingency acquisition. Also, the non-learner's $t9$ is not defined since y^* does not reach 0.9 during the sessions.

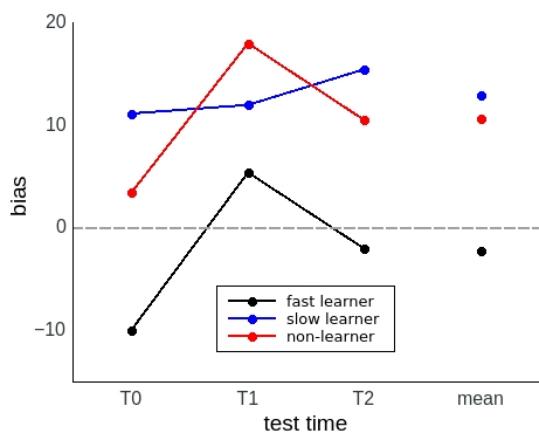
³Other fitting names for this quantity might be $t0.5$ or $t50$. However, I will stick with this notation for brevity's sake. The same holds for $t9$.



(a) Experiment



(b) Model 1



(c) Model 2

Figure 3.13: Functioning bias of the subjects of Fig. 3.12: experimental data (a) and simulated behavior (b and c). Dashed lines indicate level of unbiased behavior.

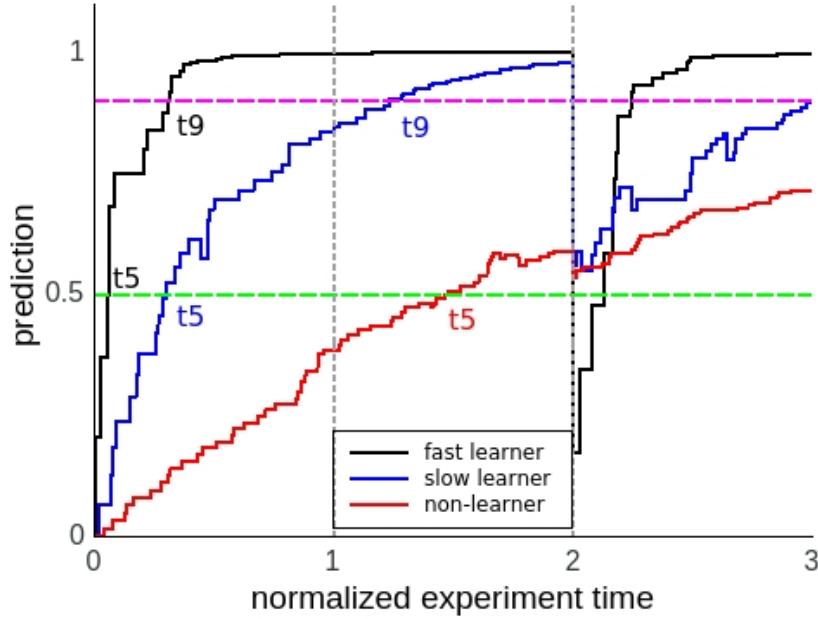


Figure 3.14: Prediction variables of Fig. 3.12 plotted over normalized experiment time. Vertical dashed lines indicate session onsets. The green dashed line marks $y^* = 0.5$; crossing this line for the first time defines a subject's t_5 . The purple dashed line marks $y^* = 0.9$; crossing this line for the first time defines a subject's t_9 .

The analysis of this chapter indicates that the adaptation rate of the internal prediction may be a key quantity determining the strength of the functioning bias. This idea is studied more closely in the next section.

3.4 Learning Speed

3.4.1 Median Split

To determine the effect of the prediction adaptation rate on the functioning bias within the whole sample, I first divided the subjects into two groups using a median split with respect to fitted prediction adaptation time constant τ_p : Half of the subjects had below-median τ_p and formed the *fast learners* group; the other half had above-median τ_p and formed the *slow learners* group. The resulting group bias is plotted in Figure 3.15:

- Forming the fast and slow learners groups with respect to the Model 1 fits of τ_p yields the Model 1 estimates of the group bias (see Fig. 3.15a). The model predicts a qualitatively different behavior of the two groups: While the slow learners exhibit a highly significant functioning bias throughout all sessions, the fast learners show unbiased behavior on average and during T0. While the bias approaches a similar level at T1, the slow learners' functioning bias is much higher during T0, T2, and on average.

- Considering the experimental data of the same subject groups exposes an even stronger discrepancy (see Fig. 3.15b). The slow learners again exhibit a significant functioning bias throughout all sessions, confirming the model prediction. However, the fast learners show a consistently unbiased behavior throughout all sessions. The bias difference is significant during each session.
- Conversely, forming the two groups with respect to the Model 2 fits of τ_p yields the Model 2 estimates of the group bias (see Fig. 3.15c). Model 2 predicts a smaller difference between groups. Only the fast learners at T0 show unbiased behavior. On average, the slow learners exhibit a more biased behavior than the fast learners, which also holds for T0 only.
- Again, the corresponding median split of the experimental data yields more pronounced differences than the model predicts (see Fig. 3.15d). During all sessions, the slow learners exhibit a significant functioning bias whereas the fast learners do not. The differences are significant during T0, T2, and on average.

To summarize, the median split analyses of both models indicate that slow adaptation of the internal prediction leads to a pronounced behavioral bias whereas fast adaptation does less so or not at all.

3.4.2 Bias Correlations

To further investigate the relation between the functioning bias and the speed of learning, I analysed the correlations between the bias and τ_p . Figure 3.16 shows the bias within the fully simulated data plotted over the fitted τ_p according to Model 1 for each session as well as averaged over sessions. These model results indicate clear correlations between the bias and τ_p for T0, T2, and on average. To study the robustness of these correlations, I performed the same analyses after removing the outliers with respect to the fitted τ_p parameter (see Fig. 3.17). These analyses confirm the correlations at T0, T2, and on average. Additionally, a correlation is found during T1, which may indicate atypical outlier behavior during this session.

The experimental data confirm this model estimation for T0 and on average (see Figs. 3.18 and 3.19). Tendencies are indicated for T1 and T2. The lack of significance might be due to the small sample size for these sessions.

The Model 2 data reveal a different picture. At first glance, they agree on the correlations at T0 and on average (see Fig. 3.20). However, the sharp-peaked distribution of fitted τ_p values implies a strong dependence of the correlations on the corresponding outliers. Hence, no correlations are found after removing the outliers (see Fig. 3.21). The same holds for the experimental data mapped onto the corresponding Model 2 τ_p fits (see Figs. 3.22 and 3.23).

In summary, the two models indicate different relationships between the functioning bias and the prediction adaptation time scale. While Model 1 confirms a causal interac-

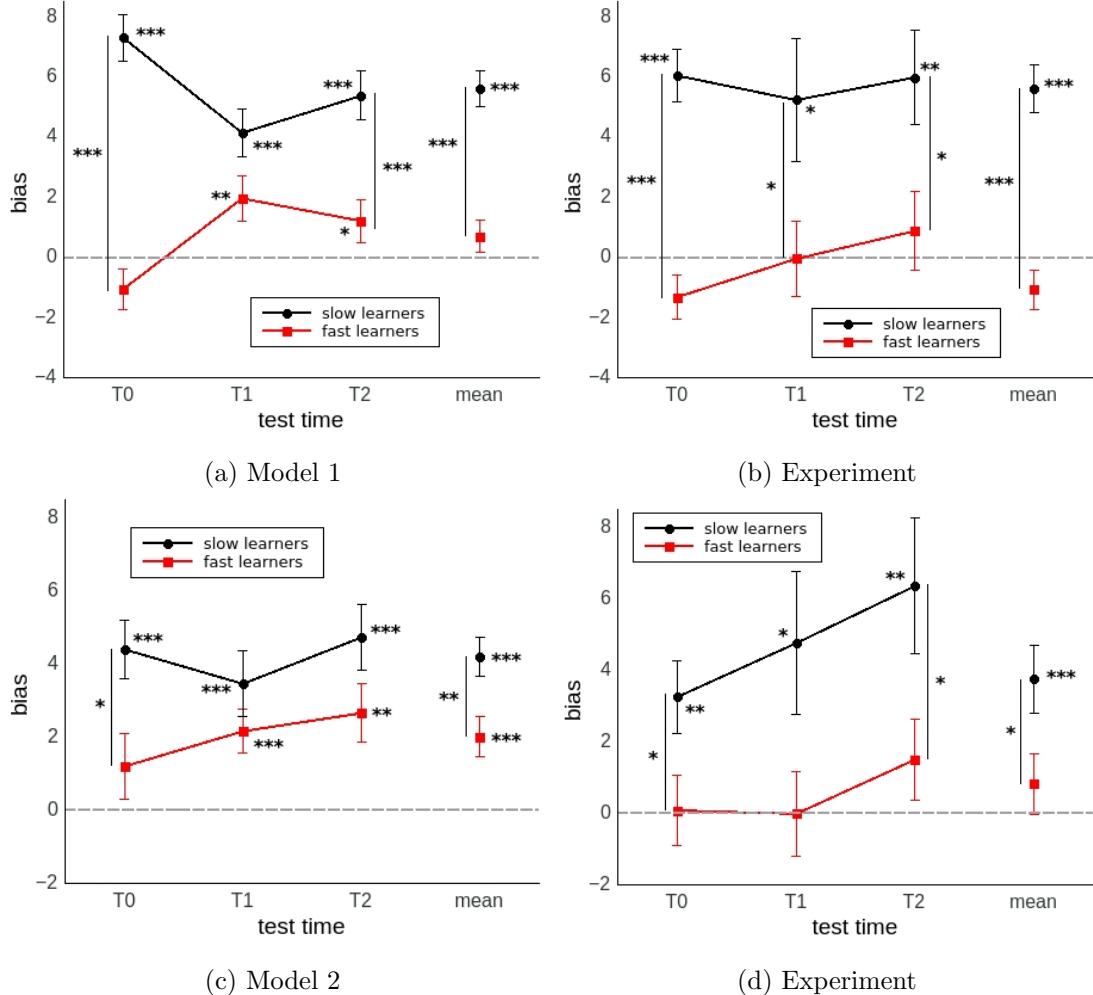


Figure 3.15: Functioning bias of subject groups defined by median split of fitted τ_p based on Model 1 fits (a,b) and Model 2 fits (c,d). The groups' bias is plotted using the full simulation data (a,c) and the corresponding experimental data (b,d). The dashed lines indicate the level of unbiased behavior.

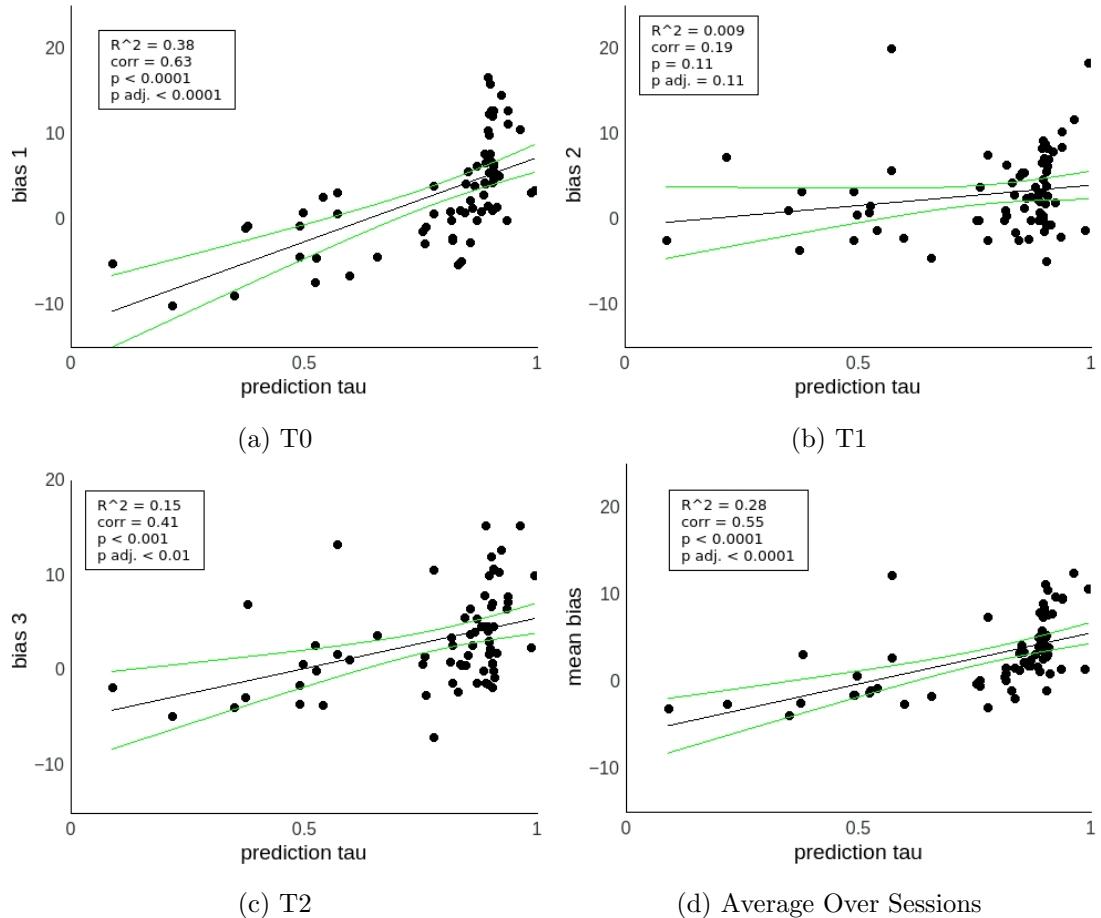


Figure 3.16: Functioning bias of full Model 1 simulations plotted over Model 1 τ_p fits for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.

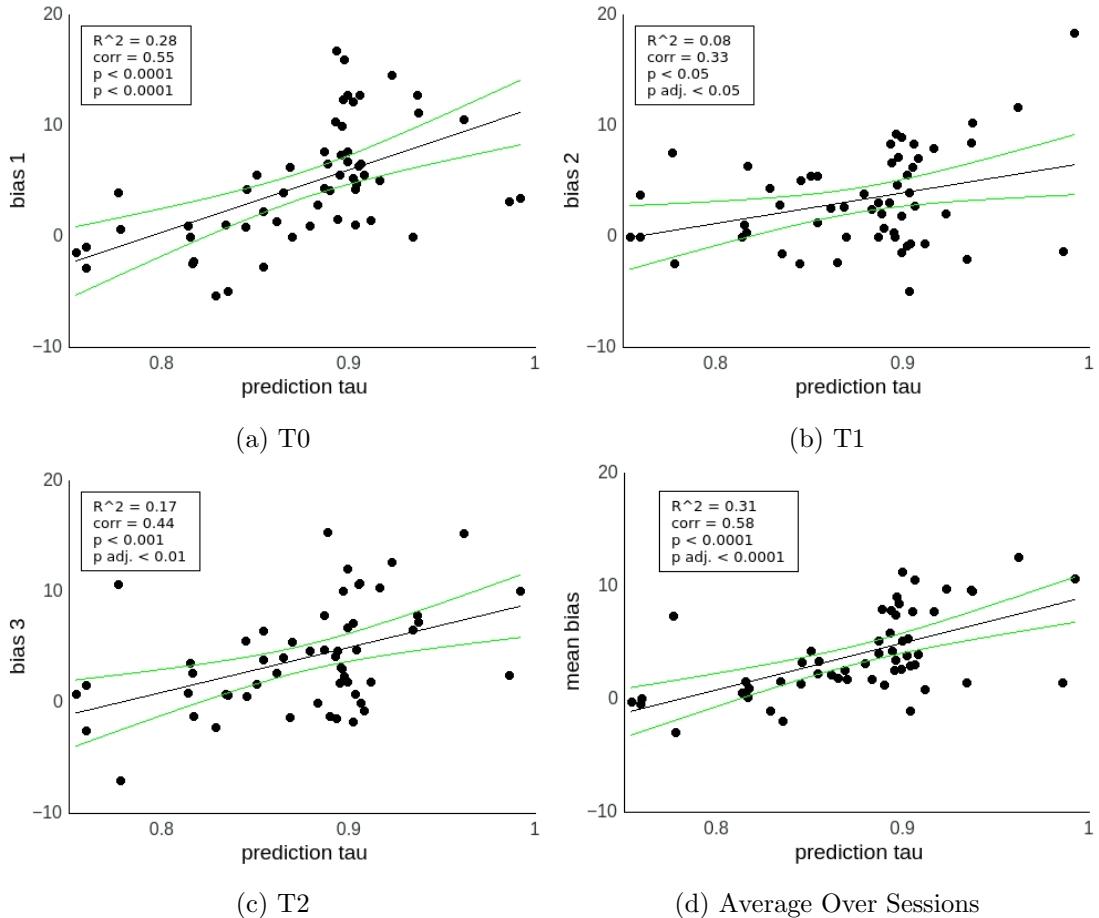


Figure 3.17: Functioning bias of full Model 1 simulations plotted over Model 1 τ_p fits without outliers for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.

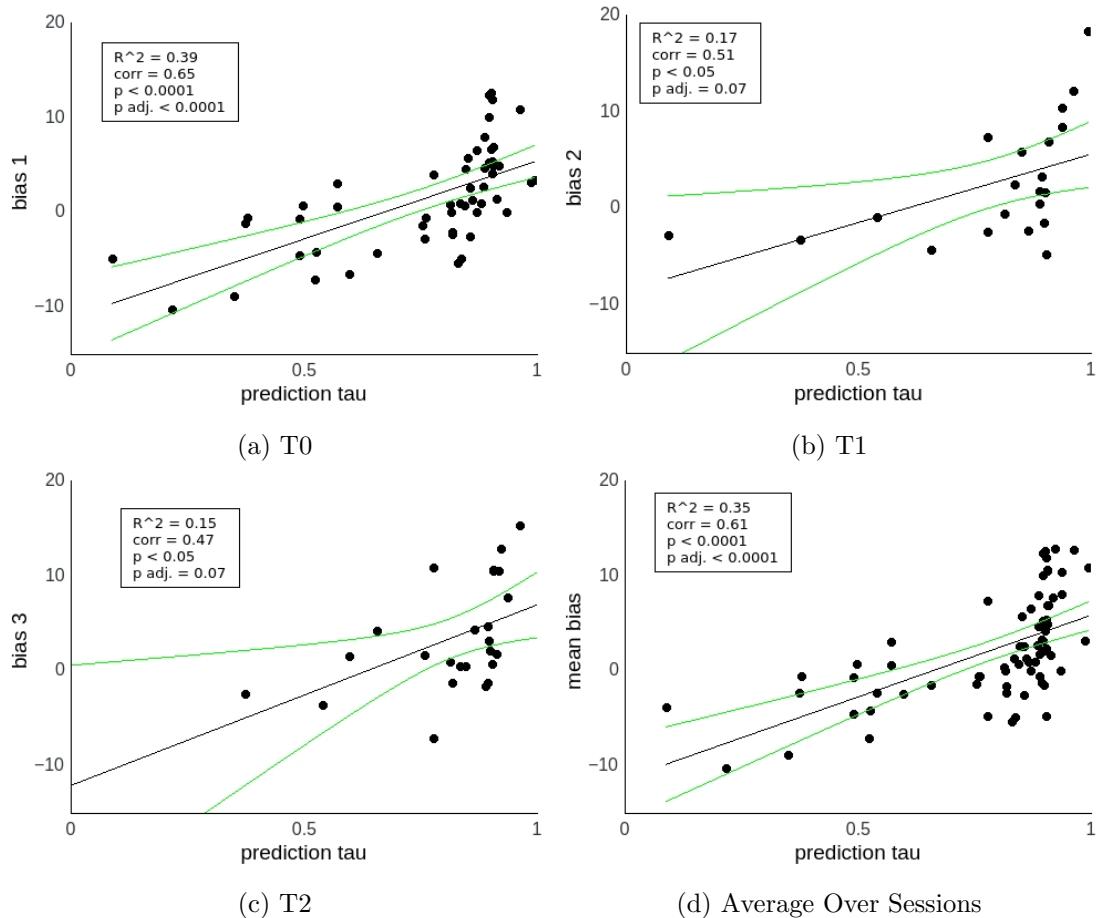


Figure 3.18: Functioning bias during experiments plotted over Model 1 τ_p fits for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.

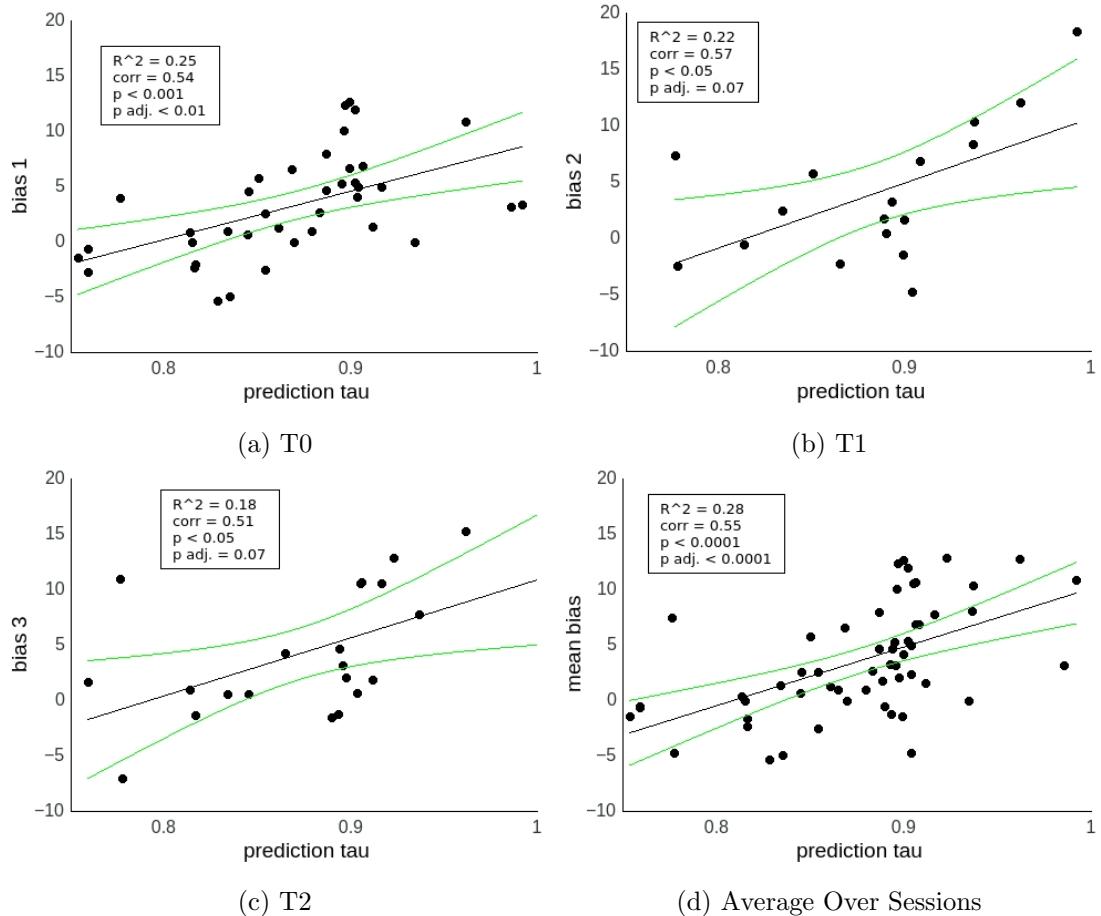


Figure 3.19: Functioning bias during experiments plotted over Model 1 τ_p fits without outliers for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.

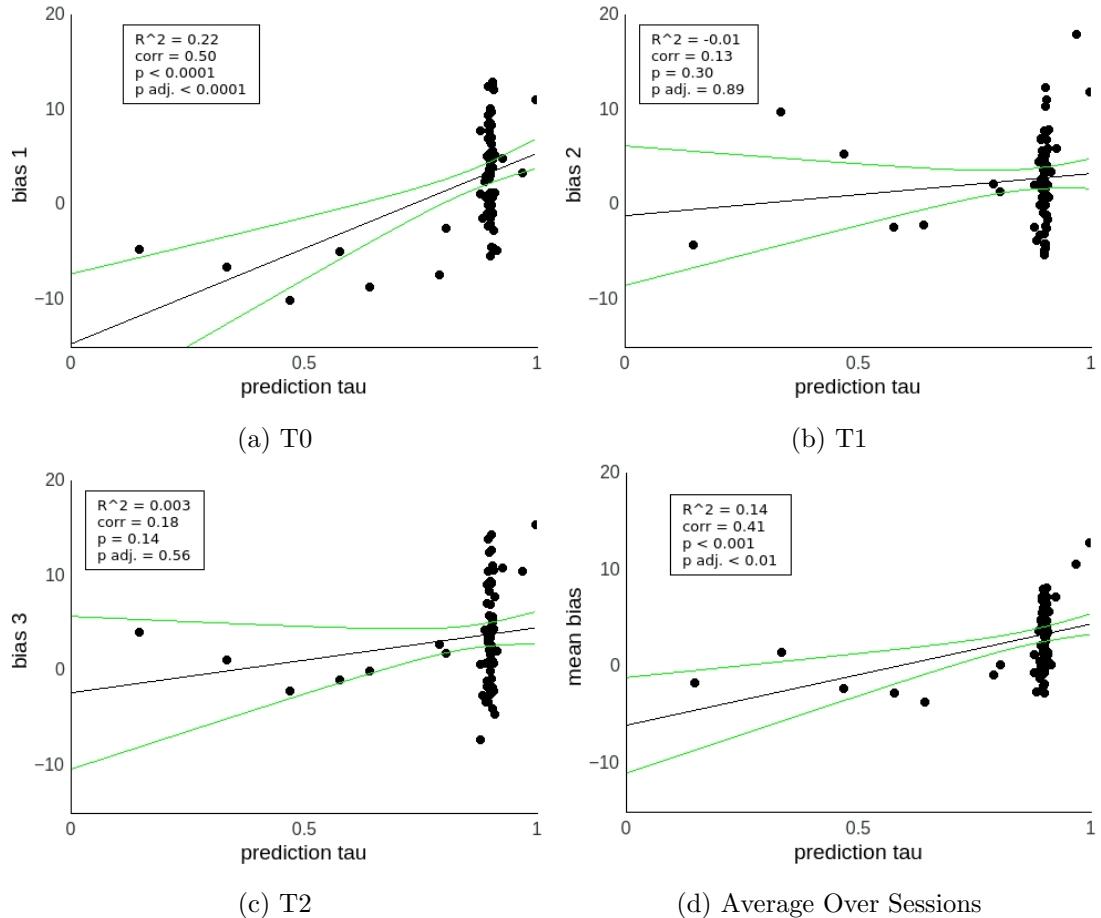


Figure 3.20: Functioning bias of full Model 2 simulations plotted over Model 2 τ_p fits for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.

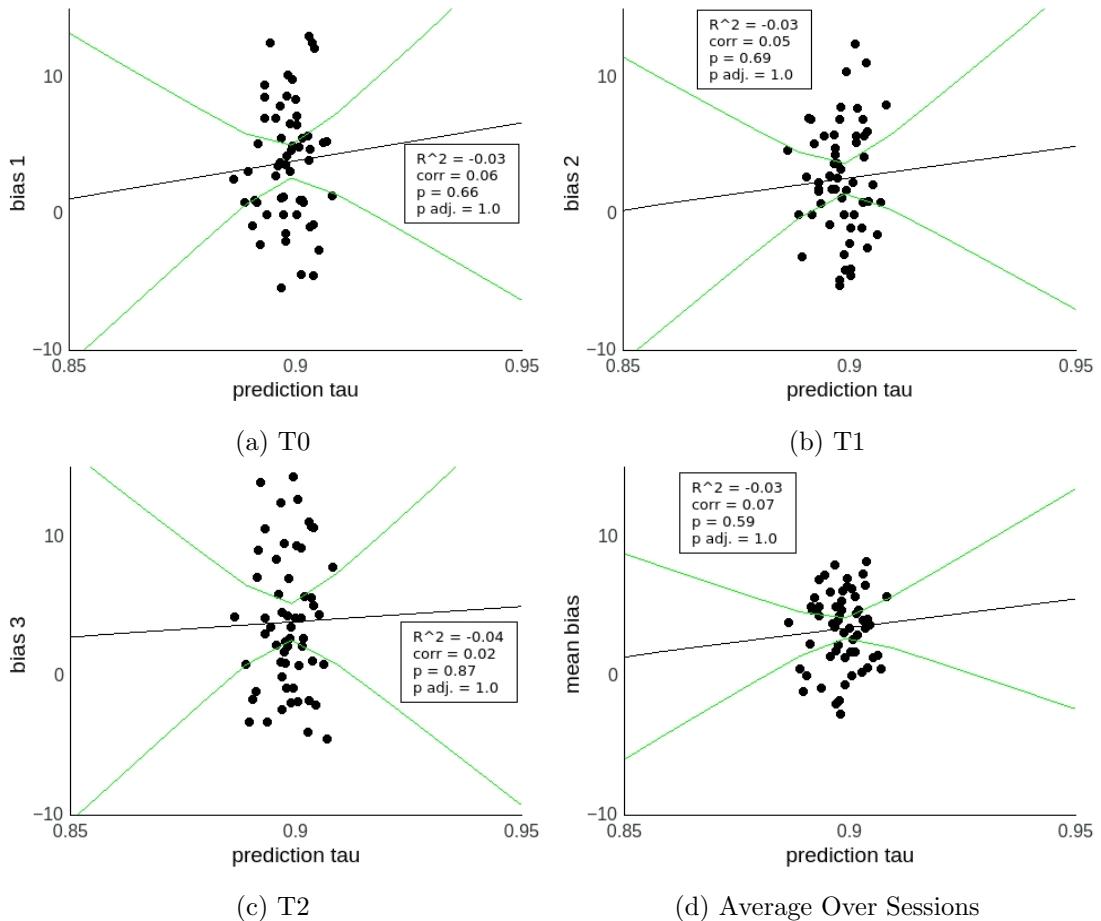


Figure 3.21: Functioning bias of full Model 2 simulations plotted over Model 2 τ_p fits without outliers for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.

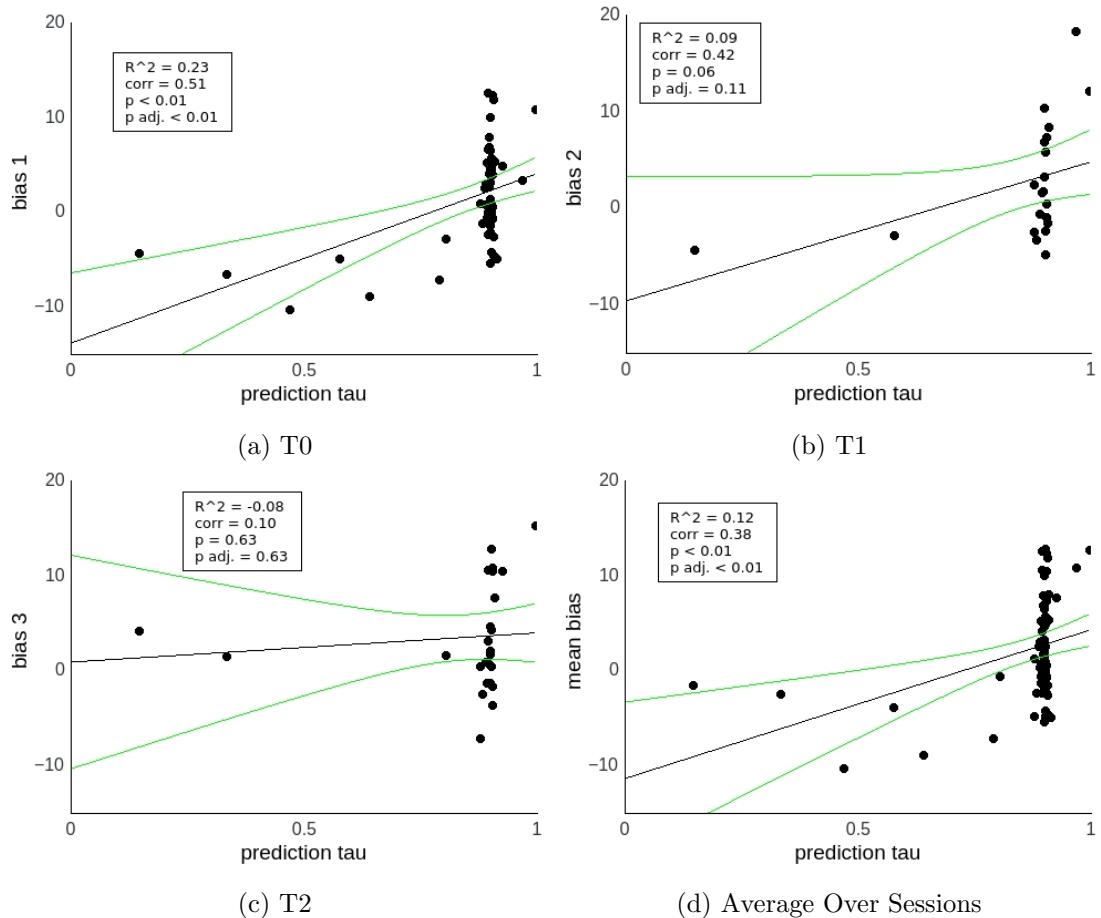


Figure 3.22: Functioning bias during experiments plotted over Model 2 τ_p fits for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.

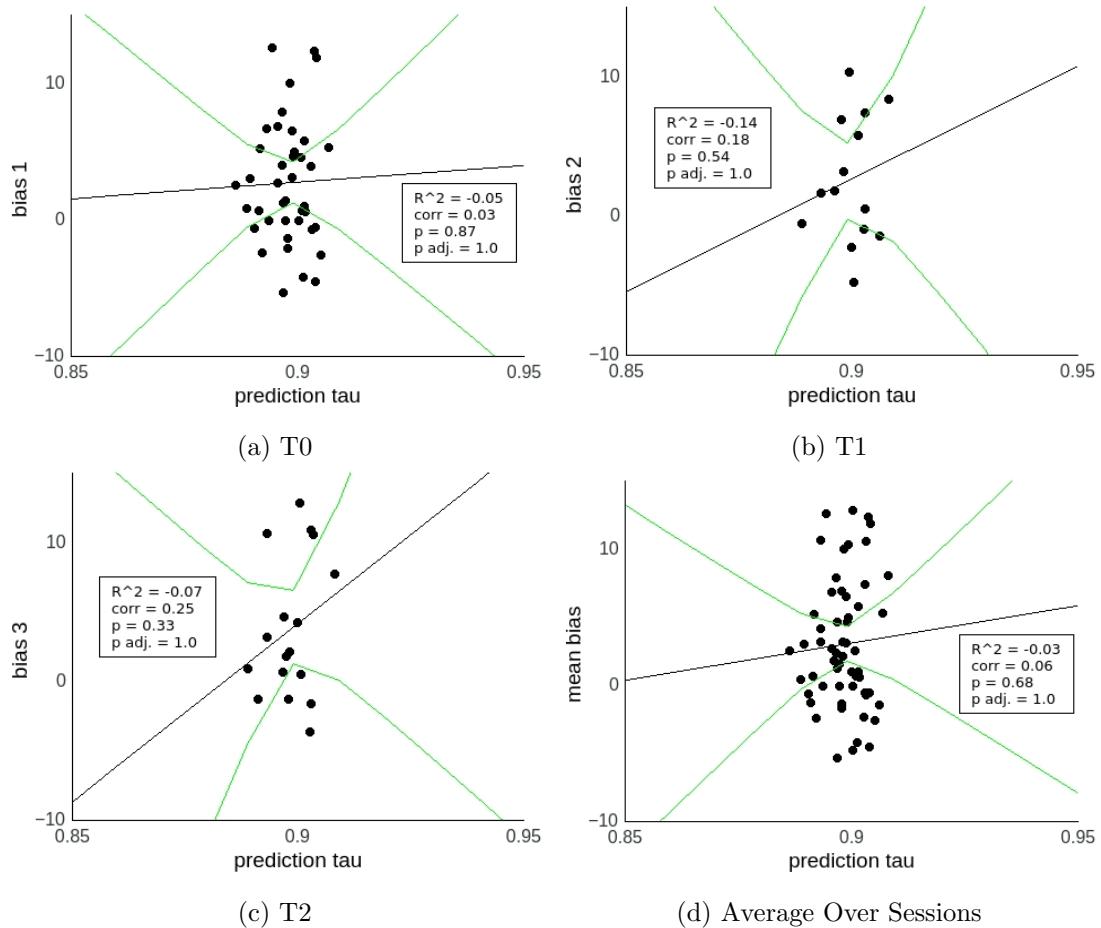


Figure 3.23: Functioning bias during experiments plotted over Model 2 τ_p fits without outliers for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.

tion⁴ between time scale and bias for T0 and on average, Model 2 finds this interaction only in the presence of outliers. Let us now analyse the behavior of these outliers.

3.4.3 Outlier Analysis

As we have seen, the distribution of fitted τ_p values can be described in terms of a cluster of samples around 0.9 and a number of outliers with considerably lower values (see Figs. 3.16 and 3.20). These outliers signify subjects with high prediction adaptation rates and contribute to a large degree to apparent bias correlations (see Figs. 3.20 and 3.22). Table 3.1 lists the numbers of outliers (on the low end of τ_p) and inlier samples (remaining samples) per session within the experimental data.

Group	Subjects	Tests		
		T0	T1	T2
Model 1 outliers	15	14	4	4
Model 1 inliers	56	39	17	21
Model 2 outliers	7	7	2	3
Model 2 inliers	64	46	19	22
all	71	53	21	25

Table 3.1: Number of analyzed τ_p outliers and inliers in experimental data per session.

Figure 3.24 shows the bias of these outliers and inliers:

- The full simulation data of Model 1 indicate that the inliers exhibit a strong functioning bias throughout all sessions (see Fig. 3.24a). The outliers, on the other hand, show a *nonfunctioning bias* during T0 and unbiased behavior at other times. Significant differences between those groups exist during T0, T2, and on average.
- The experimental data of the Model 1 outliers reveal an even larger discrepancy (see Fig. 3.24b). Whereas the inliers show the consistent functioning bias throughout the experiment, the outliers exhibit the nonfunctioning bias during T0 as well as during T1 and on average.
- The outliers with respect to the Model 2 fits behave in a very similar fashion as indicated by Model 1 (see Figs. 3.24c and d). The main difference is the missing nonfunctioning bias during T1 for the experimental data of the Model 2 outliers.

These results are reminiscent of the median split results (see Fig. 3.15). Where, previously, we observed unbiased behavior of the fast learners in some of the cases, the outliers never exhibit the functioning bias and even prefer the nonfunctioning over the functioning disc at T0.

⁴Correlation implies causality here since one quantity is an external model parameter while the other one is an observable extracted from the simulated behavior. Thus the former affects the latter, not the other way around.

Ultimately, splitting the subjects based on their outlier statistics within the τ_p fitting distribution reveals an even larger bias separation than after median splitting. The resulting qualitative difference between inliers (regular slow learners) and outliers (particularly fast learners) further underscores the notion that the functioning bias results from slow prediction adaptation. Interestingly, particularly fast prediction adaptation may lead to an initial opposite bias.

3.4.4 Uncertainty Timing

The time of maximal uncertainty regarding the subject's prediction is given by t_5 . As subjects with fast prediction adaptation should reach the uncertain regime earlier, one would expect a positive correlation between τ_p and t_5 . Figure 3.25 shows the models' resulting t_5 values plotted over fitted τ_p . A correlation is found for Model 2 but not for Model 1. However, analyzing the data after removing the outliers (both with respect to τ_p and to t_5) reveals that the Model 1 outliers overshadow a clear correlation whereas the Model 2 correlation is largely based on the corresponding outliers. Ultimately, the models do not seem to provide strong support for a correlation between τ_p and t_5 . However, the following analyses indicate that this may be due to the outliers with respect to t_5 and not to τ_p , which supports the correlation in both models (see Fig. 3.25c, where the interfering t_5 outliers are removed, and Fig. 3.25b, where the supporting τ_p outliers and the single t_5 outlier are retained).

Figure 3.26 shows the distribution of t_5 values for the outlier and inlier groups as defined in Section 3.4.3 as estimated by both models. While Model 1 does not quite estimate a significant difference between both groups' t_5 values, the inliers of the Model 2 estimates do exhibit significantly larger t_5 values. So the notion of slower prediction adaptation leading to later uncertain regimes is confirmed at least by Model 2 on the group level.

To further study the relationship between the functioning bias and prediction adaptation, I analyzed the interaction between the bias and t_5 for each session:

- The fully simulated Model 1 data indicate clear correlations for T1, T2, and on average (see Fig. 3.27). After removing the t_5 outliers, strong correlations are found for all sessions including T0, which indicates atypical behavior of the t_5 outliers (see Fig. 3.28).
- Analyzing the experimental bias data combined with Model 1's t_5 estimates confirms these simulation results, i.e., clear correlations only in the absence of t_5 outliers (see Figs. 3.29 and 3.30).
- Model 2 draws a very different picture. The fully simulated Model 2 data indicate correlations for all sessions, which essentially depend on the outliers—both the extremely large one and the small ones close to 0 (see Figs. 3.31 and 3.33).
- The same is found for the experimental data combined with Model 2's t_5 estimates (see Figs. 3.33 and 3.34).

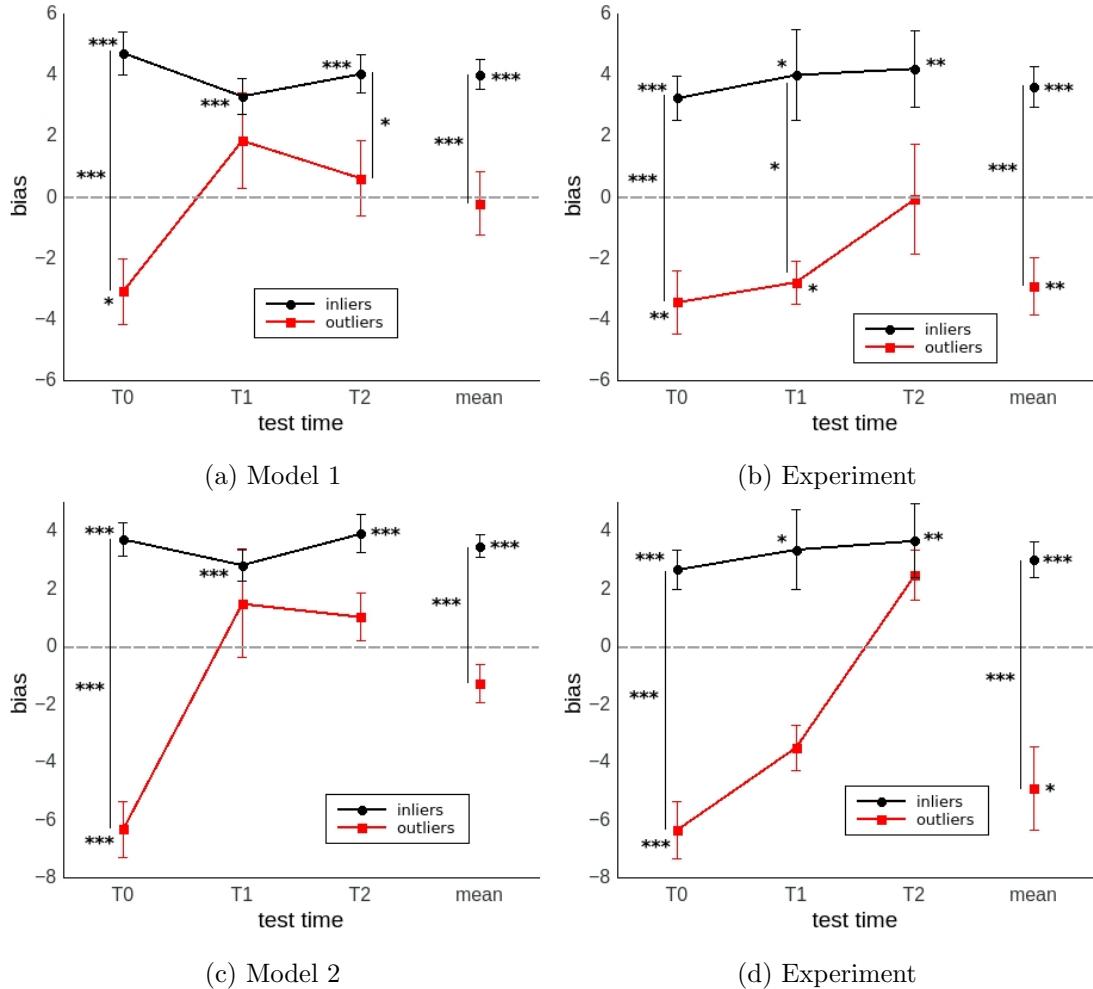


Figure 3.24: Functioning bias of lower τ_p outliers (smaller than first quartile minus 1.5 times interquartile range) with respect to τ_p values fitted using Model 1 (a,b) and Model 2 (c,d) as well as inliers (remaining samples). The groups' functioning bias is plotted using the full simulation data (a,c) and the corresponding experimental data (b,d). The dashed lines indicate the level of unbiased behavior.

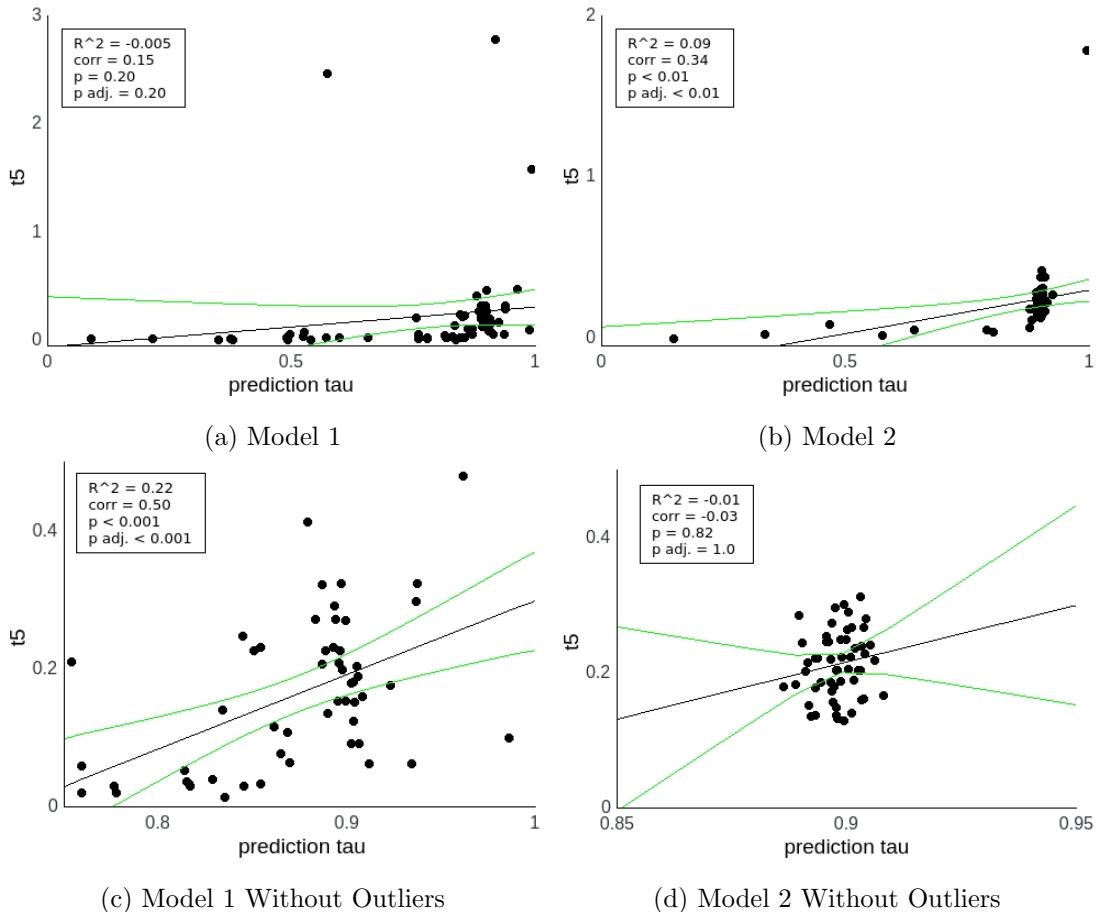


Figure 3.25: t_5 values extracted from simulated experiments based on Model 1 (a) and Model 2 (b) plotted over the corresponding fitted τ_p values. The same data are plotted after removing both the t_5 and τ_p outliers (c,d). Lines indicate correlation fit and 95% confidence bands.

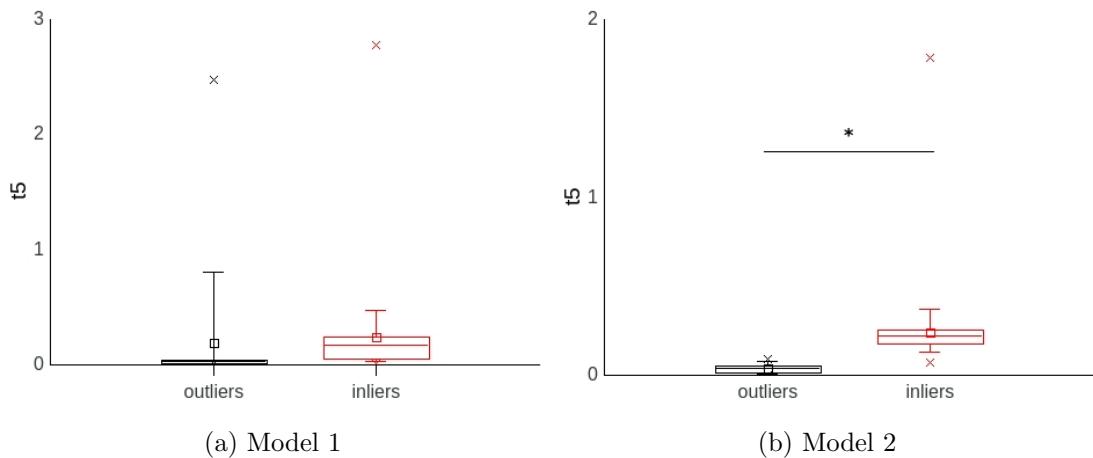


Figure 3.26: The distributions of t_5 values of simulated Model 1 outliers and inliers (a) and Model 2 outliers and inliers (b).

To summarize, both models indicate correlations between t5 and the functioning bias, i.e., the later the uncertain regime is reached, the higher the level of observed bias. In case of Model 1, outliers with very high t5 values interfere with the correlation, whereas the correlation with Model 2 depends on outliers both on the low and the high end of the t5 spectrum, which is likely due to the strong clustering of most t5 values around a single value.

The difficulty of finding conclusive relationships between t_5 on one hand and prediction adaptation rate and the functioning bias on the other hand may be due to the fact that t_5 occurs relatively early, i.e., halfway between the beginning and ending of successful acquisition. t_5 values thus cover a relatively small range of values, which hampers correlation analyses as those values are more prone to perturbations by exploratory noise. A more reliable measure in this respect might be the acquisition time t_9 , which will be analysed in the next section.

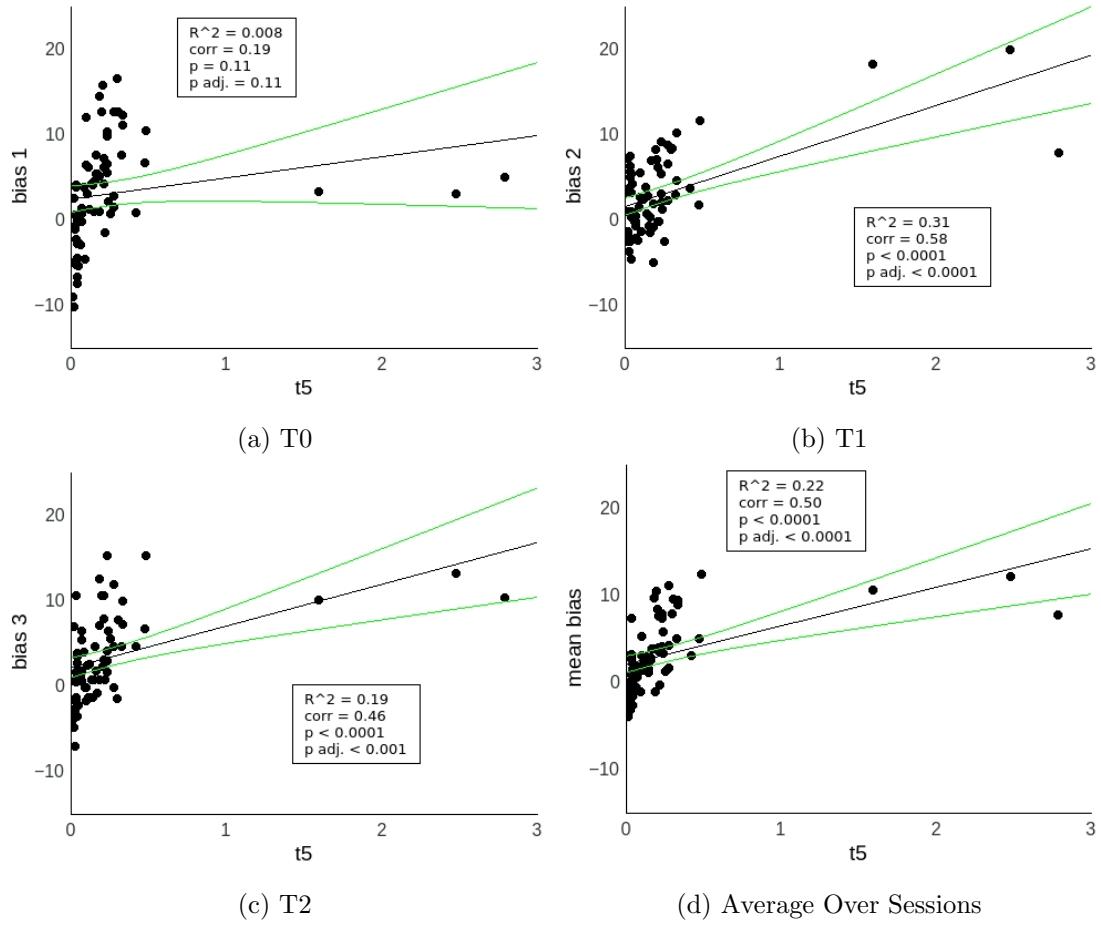


Figure 3.27: Functioning bias of full Model 1 simulations plotted over corresponding t_5 values for each session. Lines indicate correlation fit and 95% confidence bands.

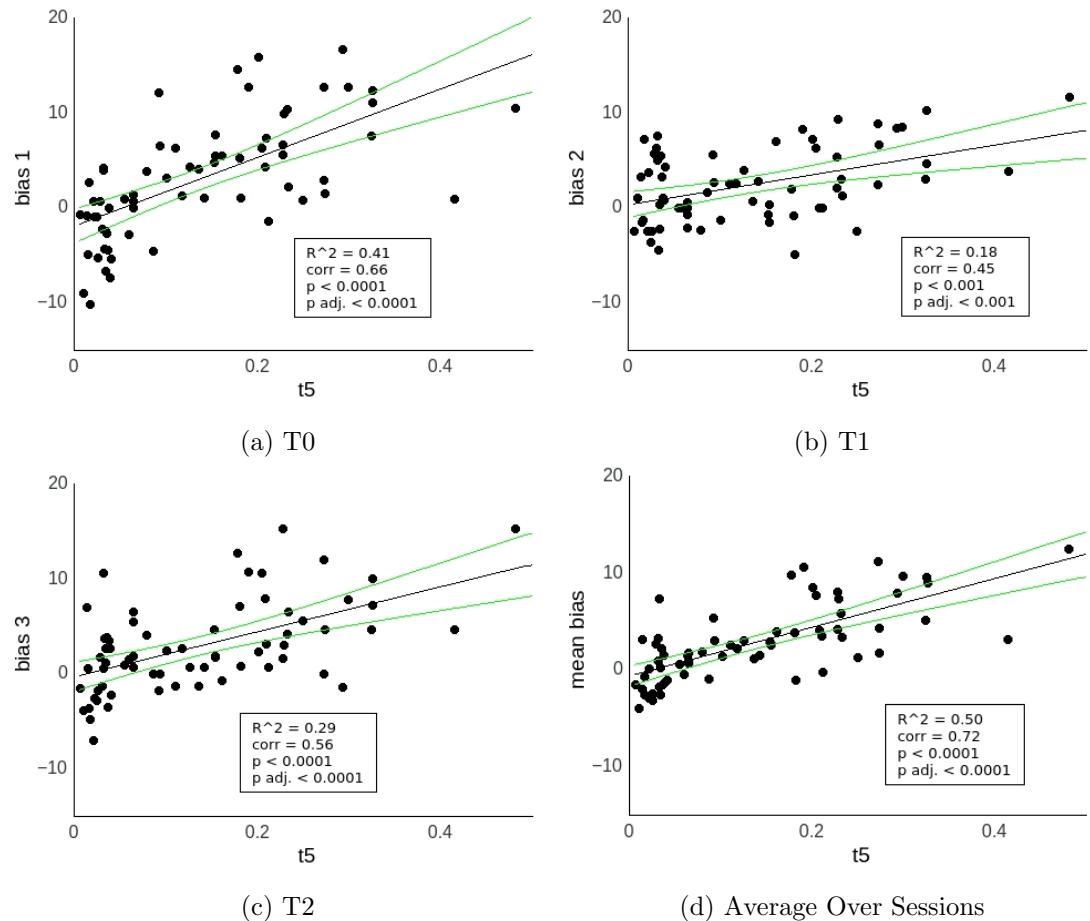


Figure 3.28: Functioning bias of full Model 1 simulations plotted over corresponding $t5$ values for each session. $t5$ outliers are removed. Lines indicate correlation fit and 95% confidence bands.

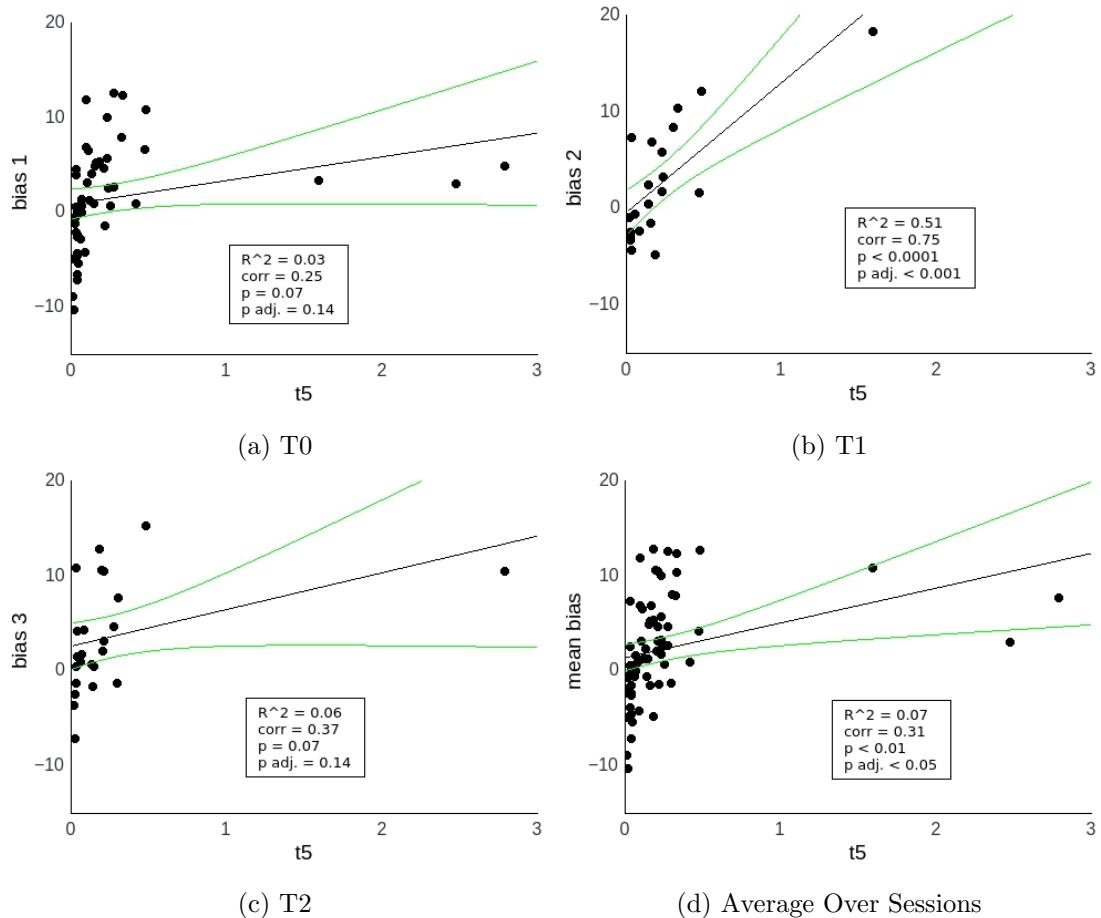


Figure 3.29: Functioning bias of experimental data plotted over corresponding t_5 values estimated by Model 1 for each session. Lines indicate correlation fit and 95% confidence bands.

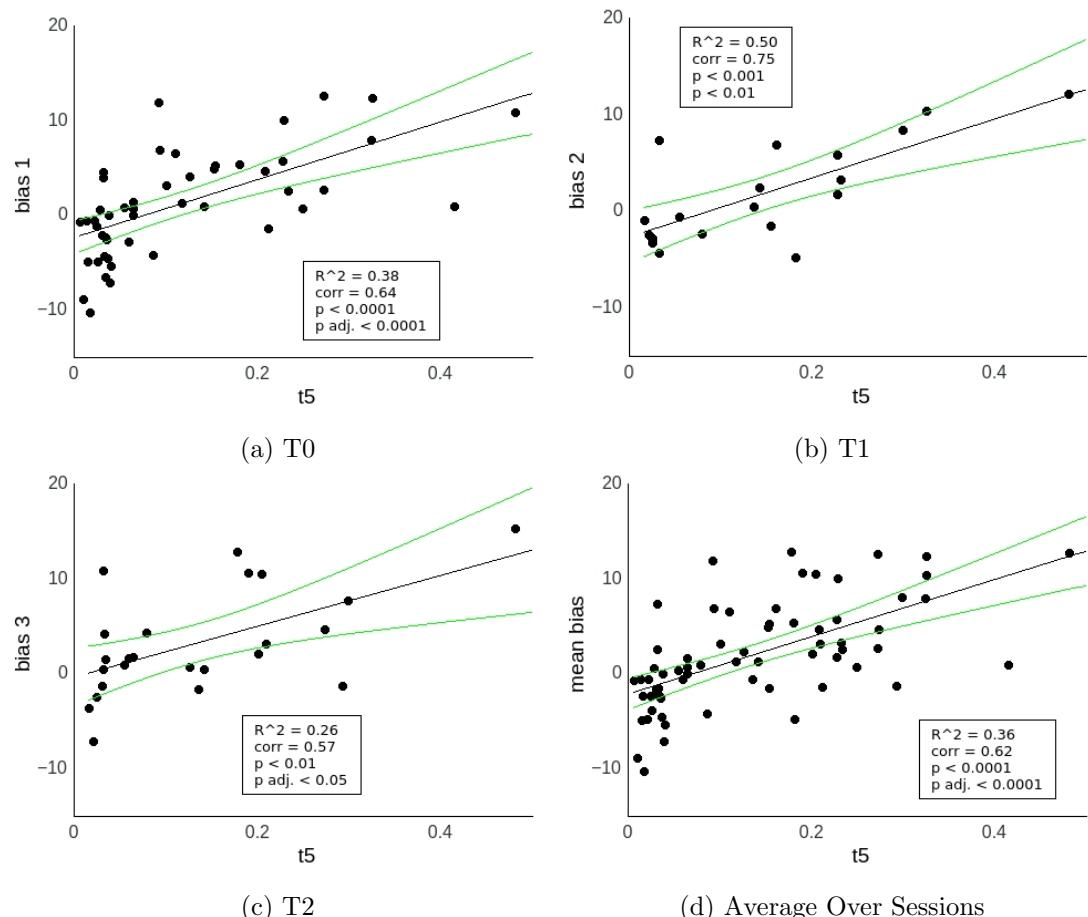


Figure 3.30: Functioning bias of experimental data plotted over corresponding t_5 values estimated by Model 1 for each session. t_5 outliers are removed. Lines indicate correlation fit and 95% confidence bands.

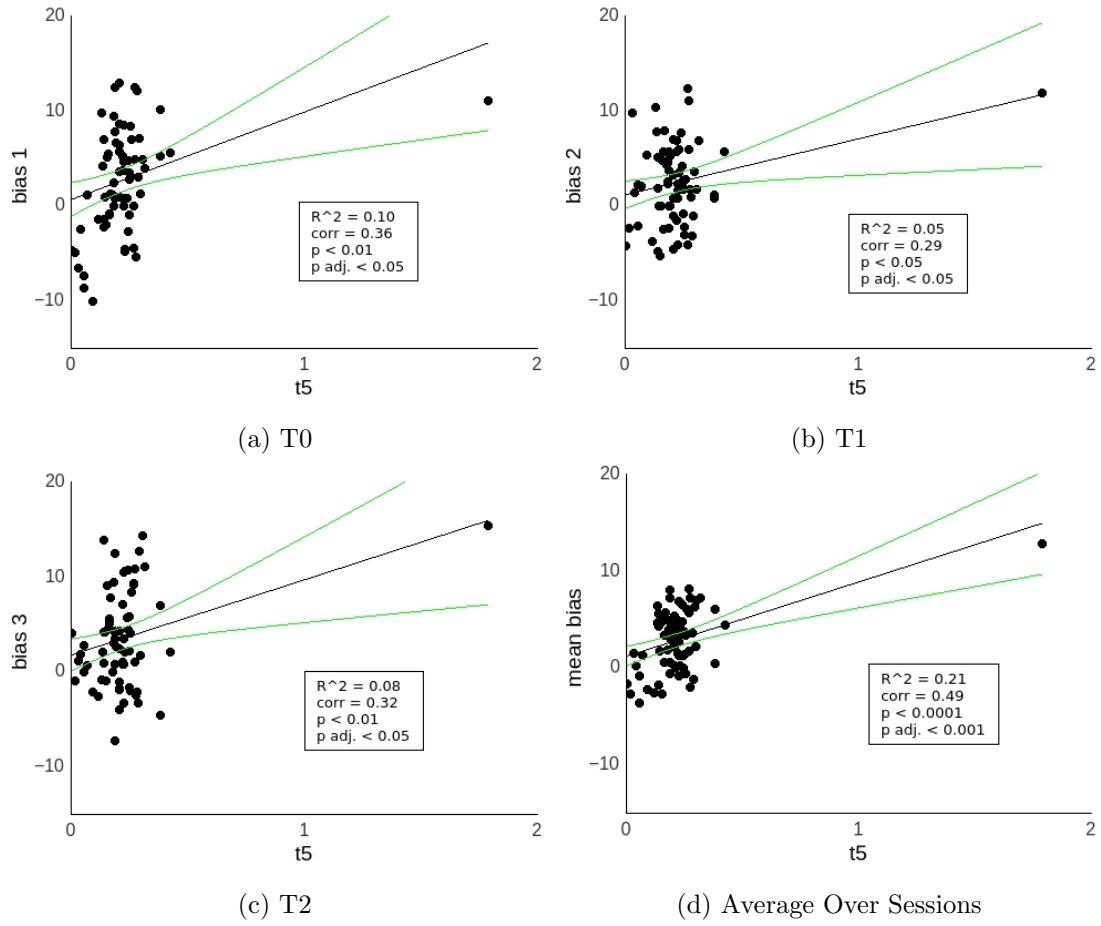


Figure 3.31: Functioning bias of full Model 2 simulations plotted over corresponding t_5 values for each session. Lines indicate correlation fit and 95% confidence bands.

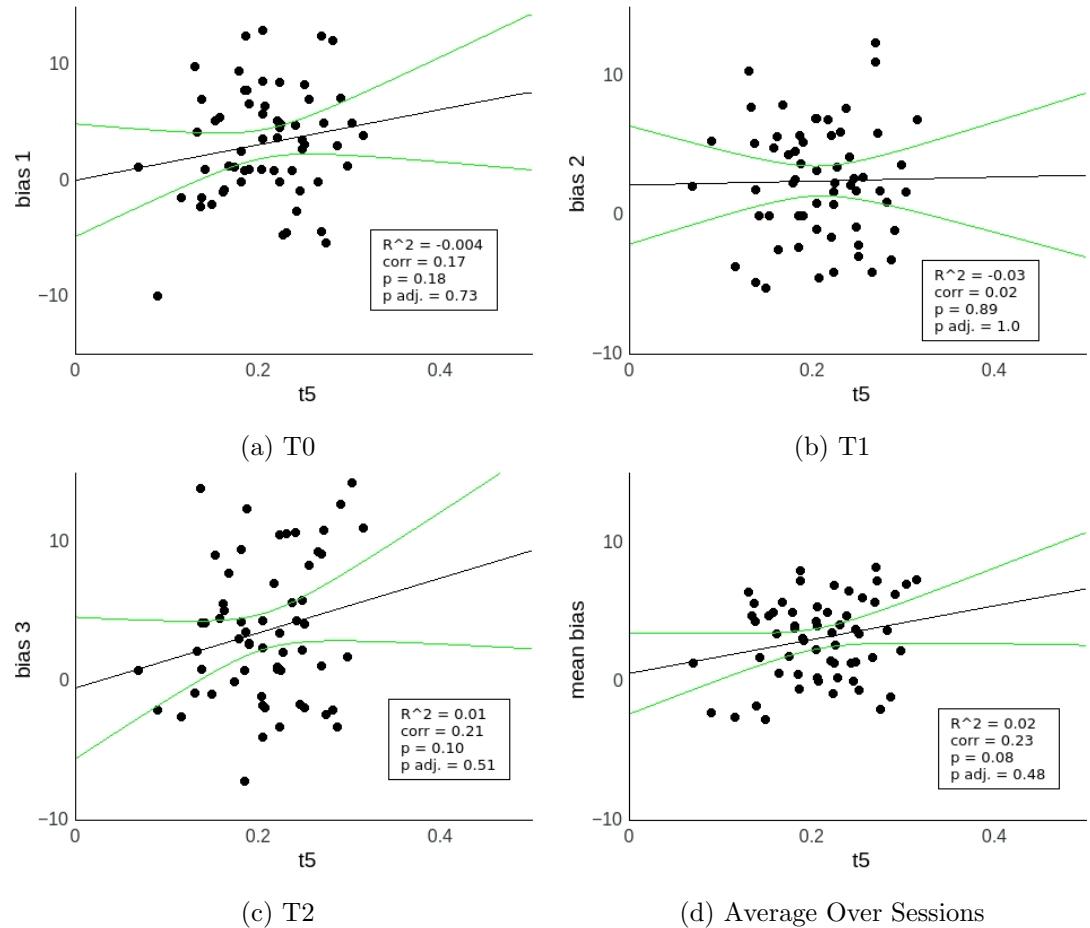


Figure 3.32: Functioning bias of full Model 2 simulations plotted over corresponding $t5$ values for each session. $t5$ outliers are removed. Lines indicate correlation fit and 95% confidence bands.

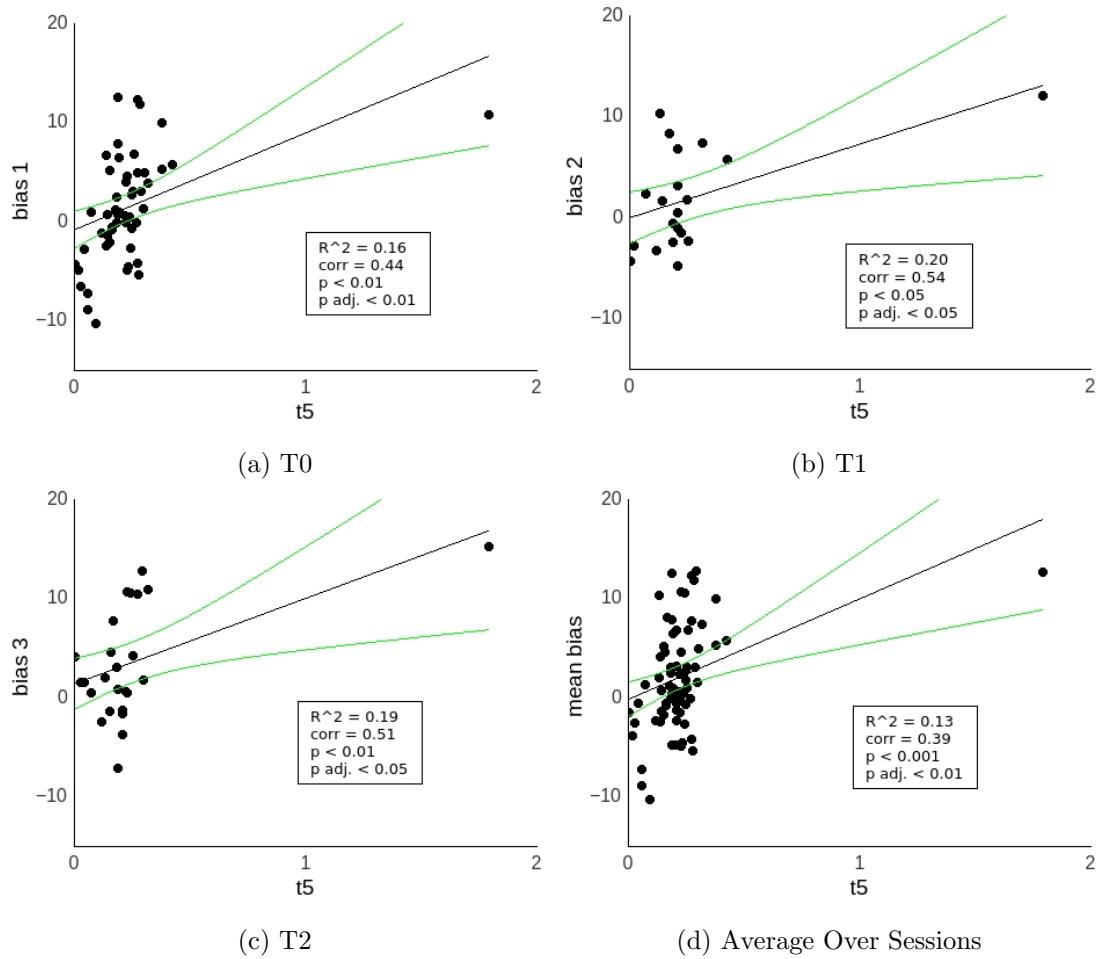


Figure 3.33: Functioning bias of experimental data plotted over corresponding t_5 values estimated by Model 2 for each session. Lines indicate correlation fit and 95% confidence bands.

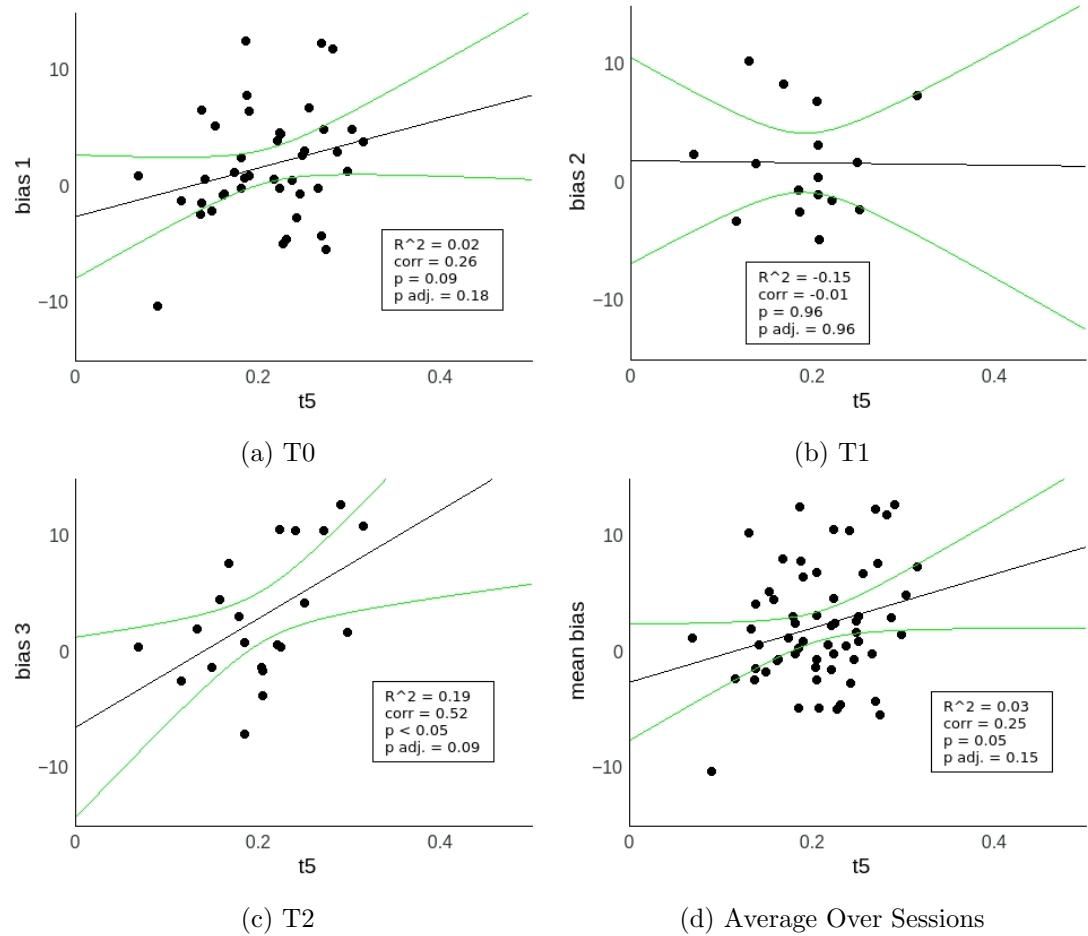


Figure 3.34: Functioning bias of experimental data plotted over corresponding $t5$ values estimated by Model 2 for each session. $t5$ outliers are removed. Lines indicate correlation fit and 95% confidence bands.

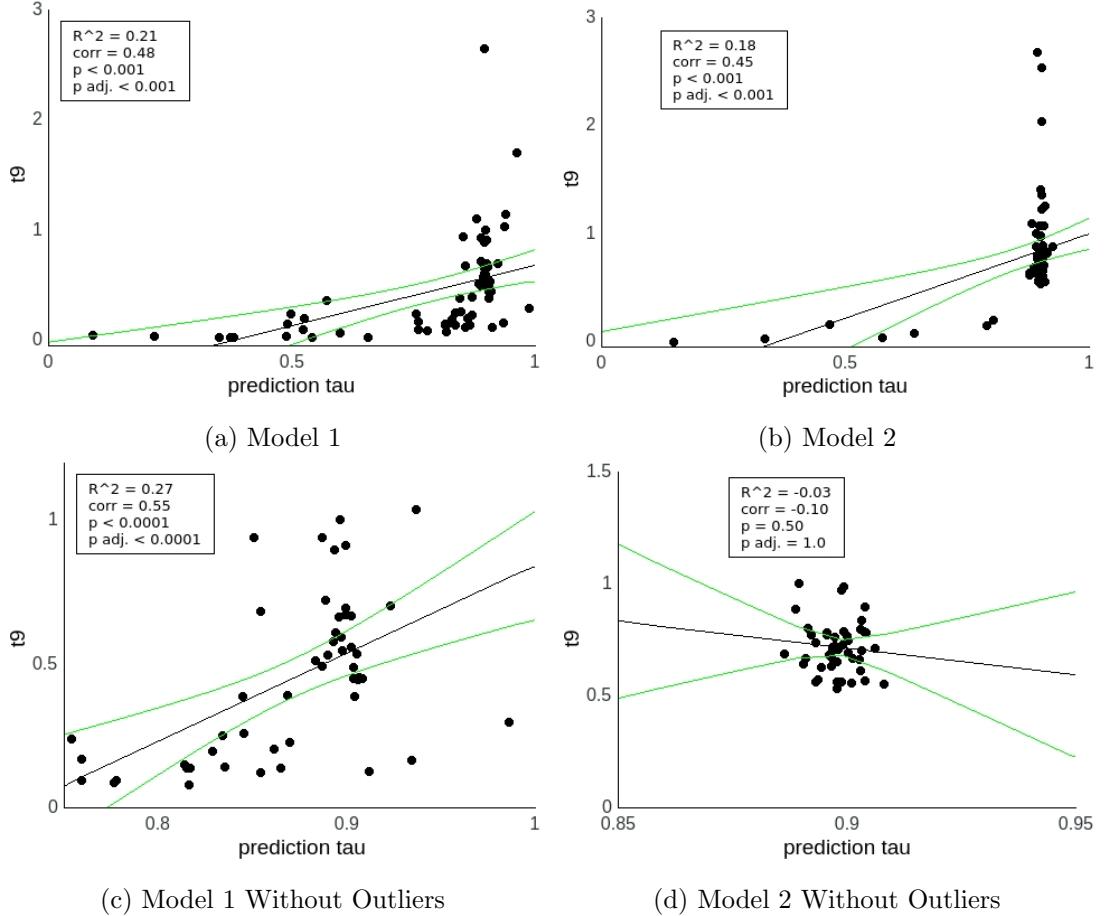


Figure 3.35: t_9 values extracted from simulated experiments based on Model 1 (a) and Model 2 (b) plotted over the corresponding fitted τ_p values. The same data are plotted after removing both the t_9 and τ_p outliers (c,d). Lines indicate correlation fit and 95% confidence bands.

3.4.5 Acquisition Time

The time of contingency acquisition is approximated by t_9 . Thus slow learning is indicated by high t_9 values, which are then expected to be correlated with the prediction adaptation time constant τ_p . Figure 3.35 shows the t_9 values of both models' simulated data plotted over corresponding fitted τ_p values. Both models confirm the correlation between t_9 and τ_p (see Fig. 3.35a,b). After removing both the t_9 and the τ_p outliers, the correlation within the Model 1 data remains, indicating a robust finding (see Fig. 3.35c). However, the Model 2 correlation is not present after removing the outliers, indicating a dependence on both the outliers and the cluster of regular t_9 - τ_p values (see Fig. 3.35d).

As before, comparing the rapidly adapting τ_p outliers (as defined in Section 3.4.3) with the regular, slow adapters should reveal systematically lower t_9 values for the outliers if the acquisition time depends on prediction adaptation. Indeed, both models indicate

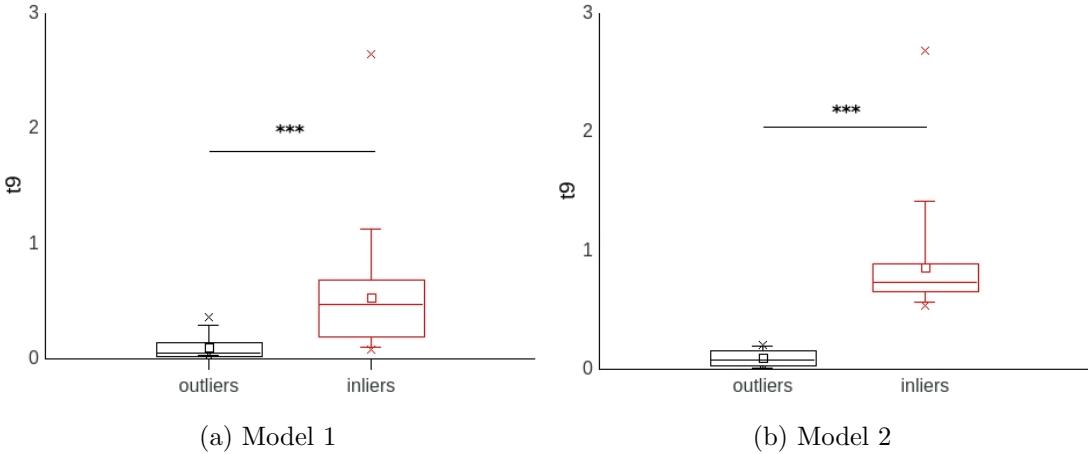


Figure 3.36: The distributions of t_9 values of simulated Model 1 outliers and inliers (a) and Model 2 outliers and inliers (b).

a corresponding, highly significant difference in t_9 values between these groups (see Fig. 3.36).

If the functioning bias is based on slow prediction adaptation, we might expect to see high bias levels in subjects that acquire the contingency late. In other words, we expect a correlation between t_9 and the functioning bias. The analyses reveal:

- The simulated Model 1 data indicate correlations between t_9 and the functioning bias for all sessions (see Fig. 3.37). These correlations are robust with respect to outlier removal (see Fig. 3.38).
- Based on these estimated t_9 values, the experimental data confirm these correlations for all sessions (see Fig. 3.39). However, correlations during T1 and T2 vanish if outliers are removed, which is probably due to the resulting insufficient sample size in these cases (see Fig. 3.40).
- Model 2 does not predict strong correlations. The simulation data indicate weak, not quite significant correlations during T0 and on average (see Fig. 3.41). Consistent with previous observations, removing the outliers of the Model 2 data removes any trace of correlations, as expected given the sharp-peaked t_9 distribution (see Fig. 3.42).
- Based on Model 2's t_9 estimates, the experimental data indicate a correlation during T0 (see Fig. 3.43). As before, no correlations are found after removing the outliers (see Fig. 3.44).

To summarize, Model 1 predicts correlations between t_9 and the functioning bias throughout all sessions and thereby confirms the notion that the functioning bias is based on late contingency acquisition, i.e., slow prediction adaptation. However, the Model 2 data are not conclusive in this respect. Based on the Model 2 estimates of the acquisition time, the experimental data confirm the correlation for T0 only.

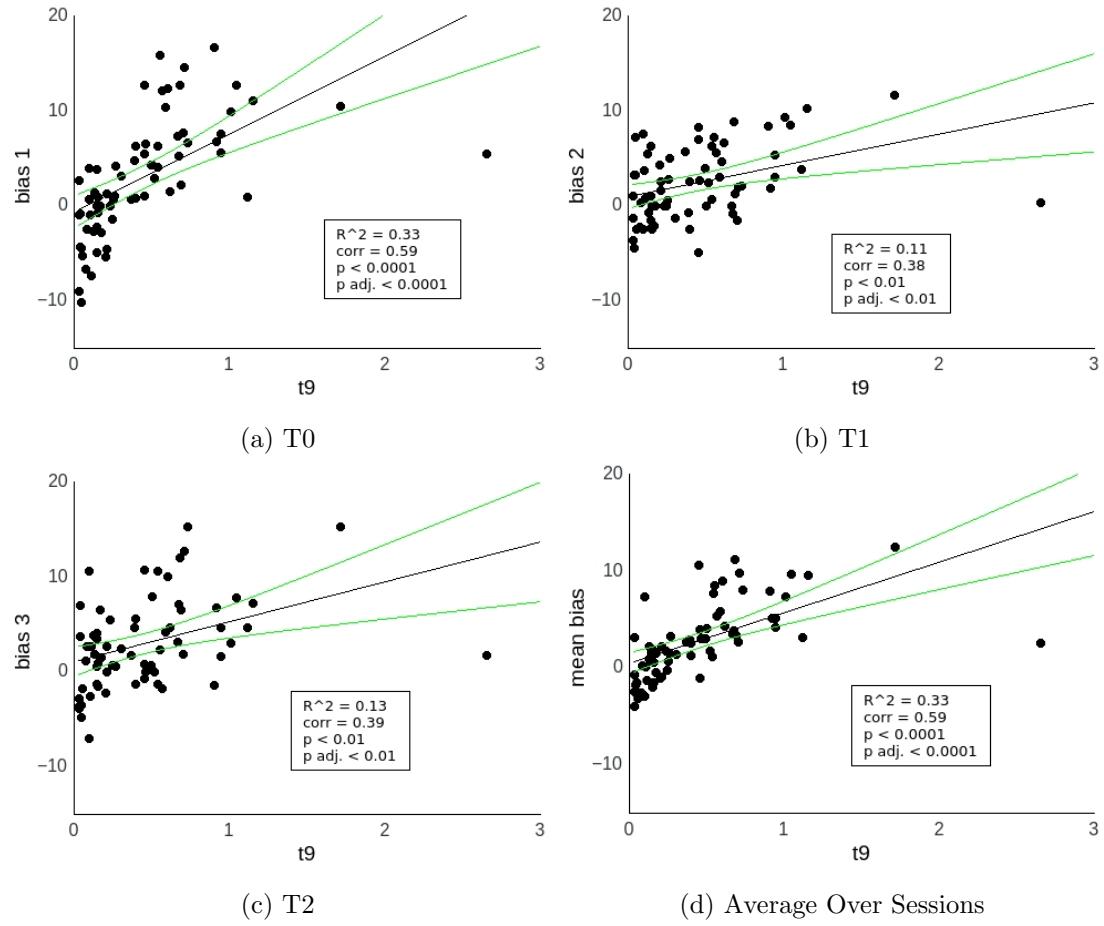


Figure 3.37: Functioning bias of full Model 1 simulations plotted over corresponding t_9 values for each session. Lines indicate correlation fit and 95% confidence bands.

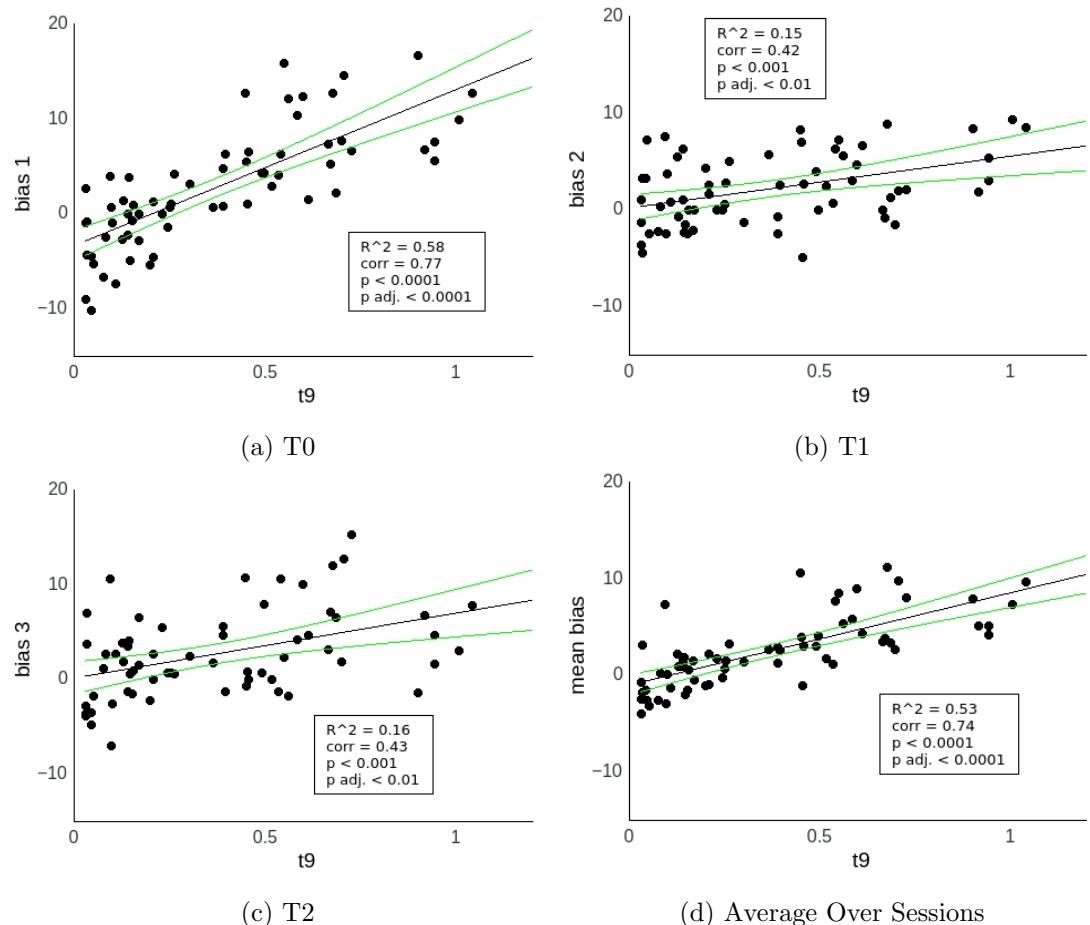


Figure 3.38: Functioning bias of full Model 1 simulations plotted over corresponding t_9 values for each session. t_9 outliers are removed. Lines indicate correlation fit and 95% confidence bands.

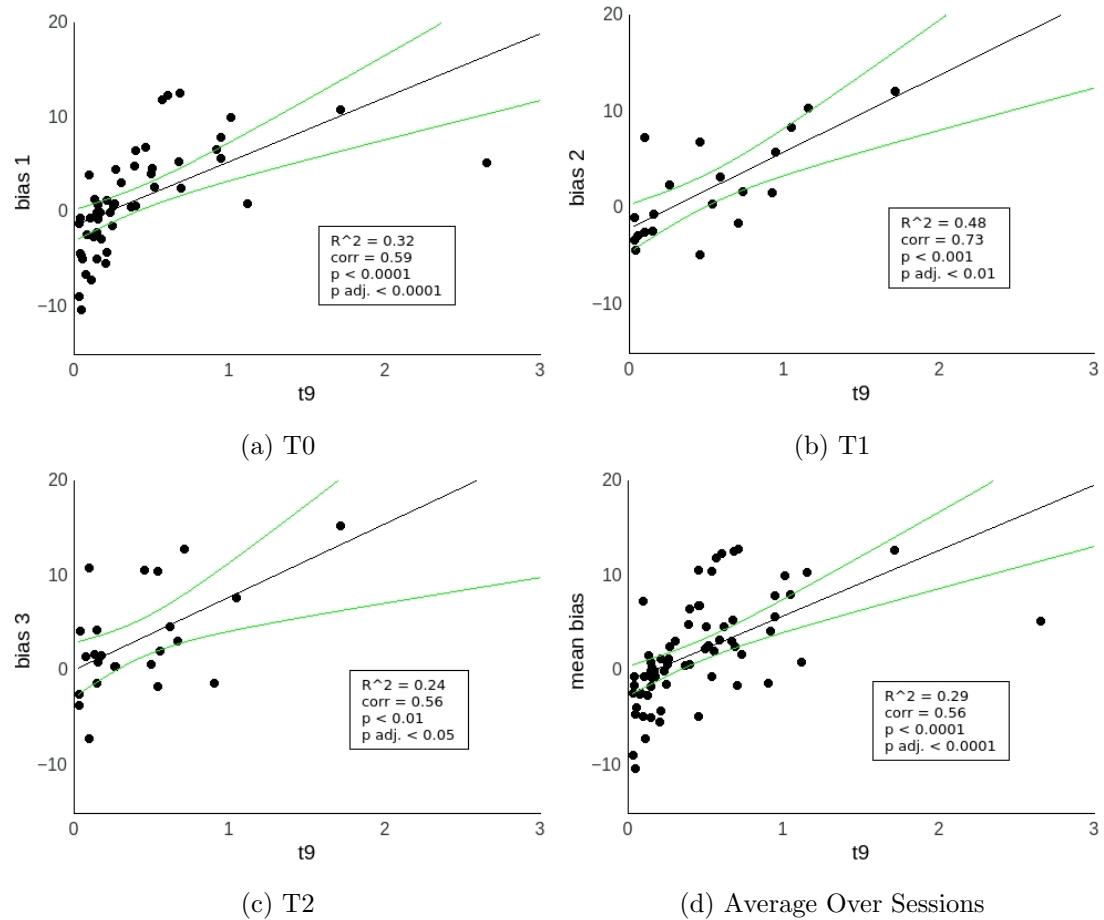


Figure 3.39: Functioning bias of experimental data plotted over corresponding t_9 values estimated by Model 1 for each session. Lines indicate correlation fit and 95% confidence bands.

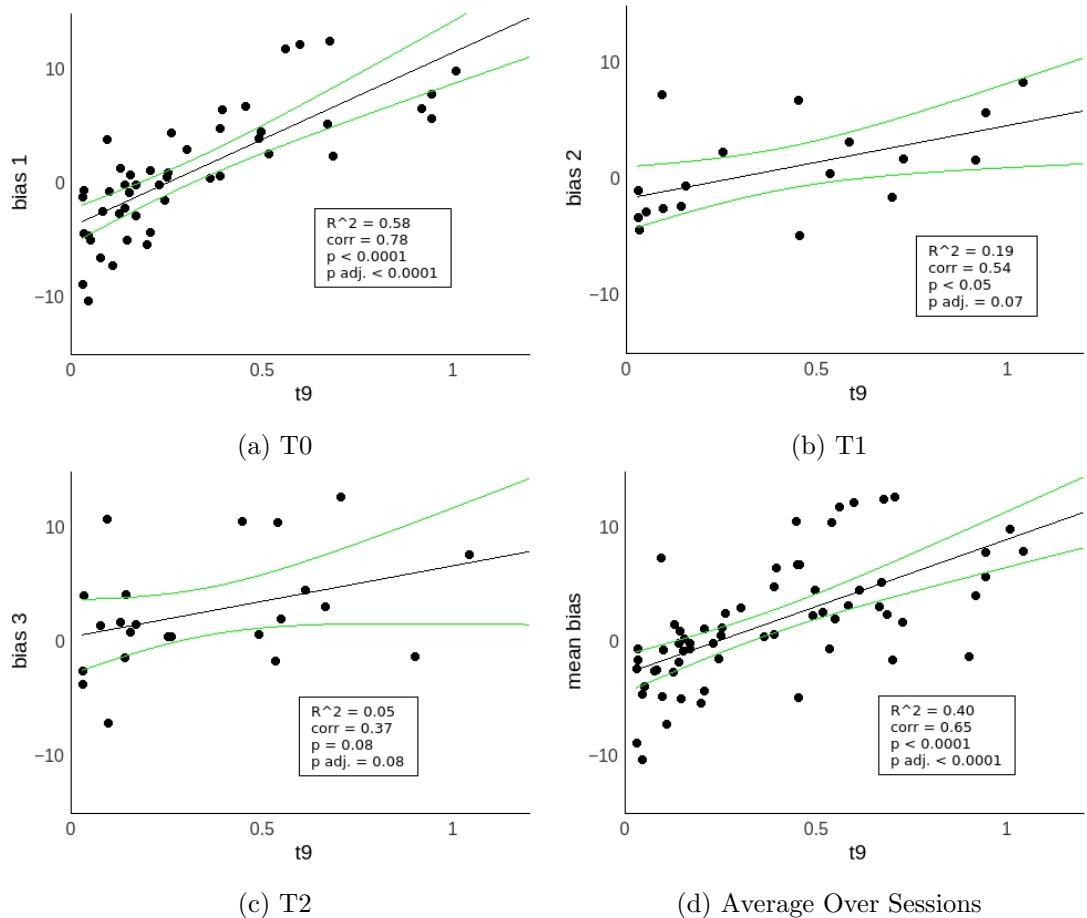


Figure 3.40: Functioning bias of experimental data plotted over corresponding t_9 values estimated by Model 1 for each session. t_9 outliers are removed. Lines indicate correlation fit and 95% confidence bands.

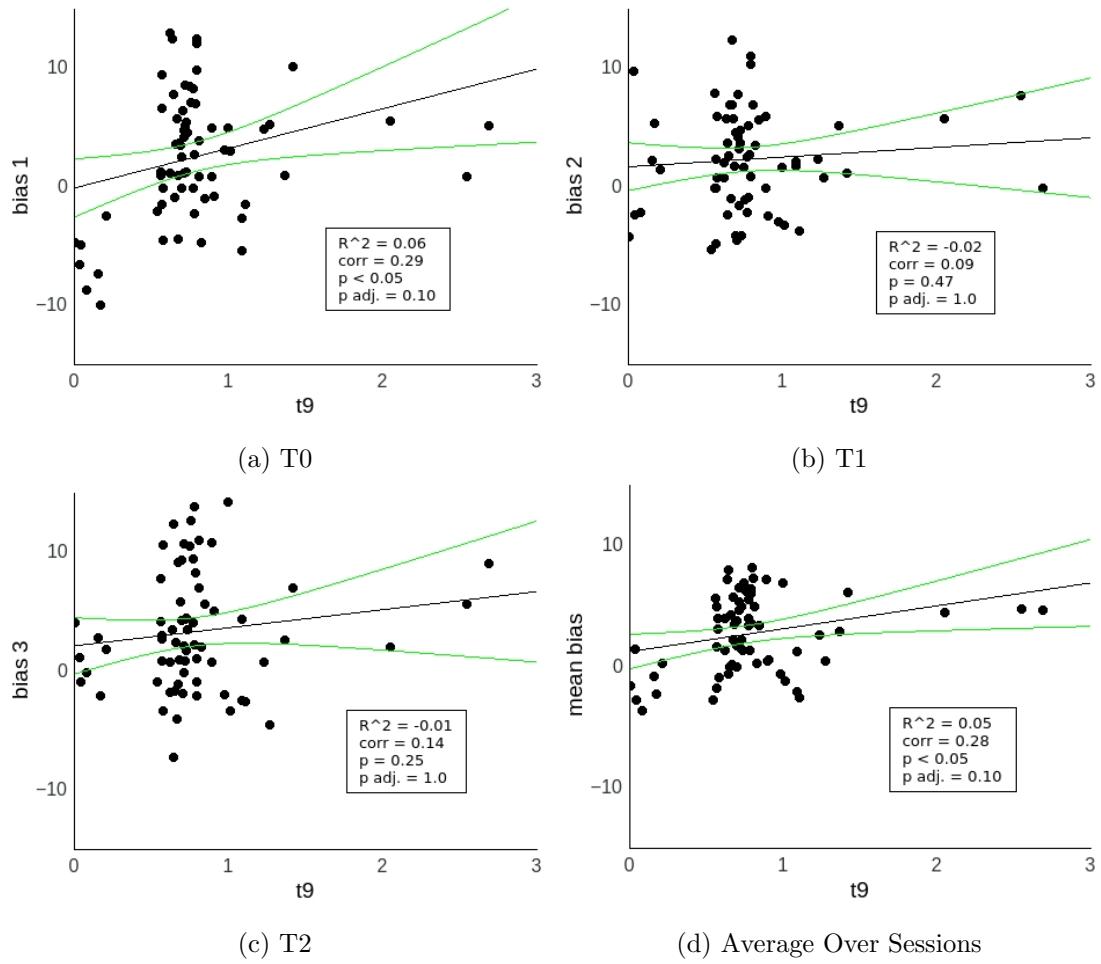


Figure 3.41: Functioning bias of full Model 2 simulations plotted over corresponding t_9 values for each session. Lines indicate correlation fit and 95% confidence bands.

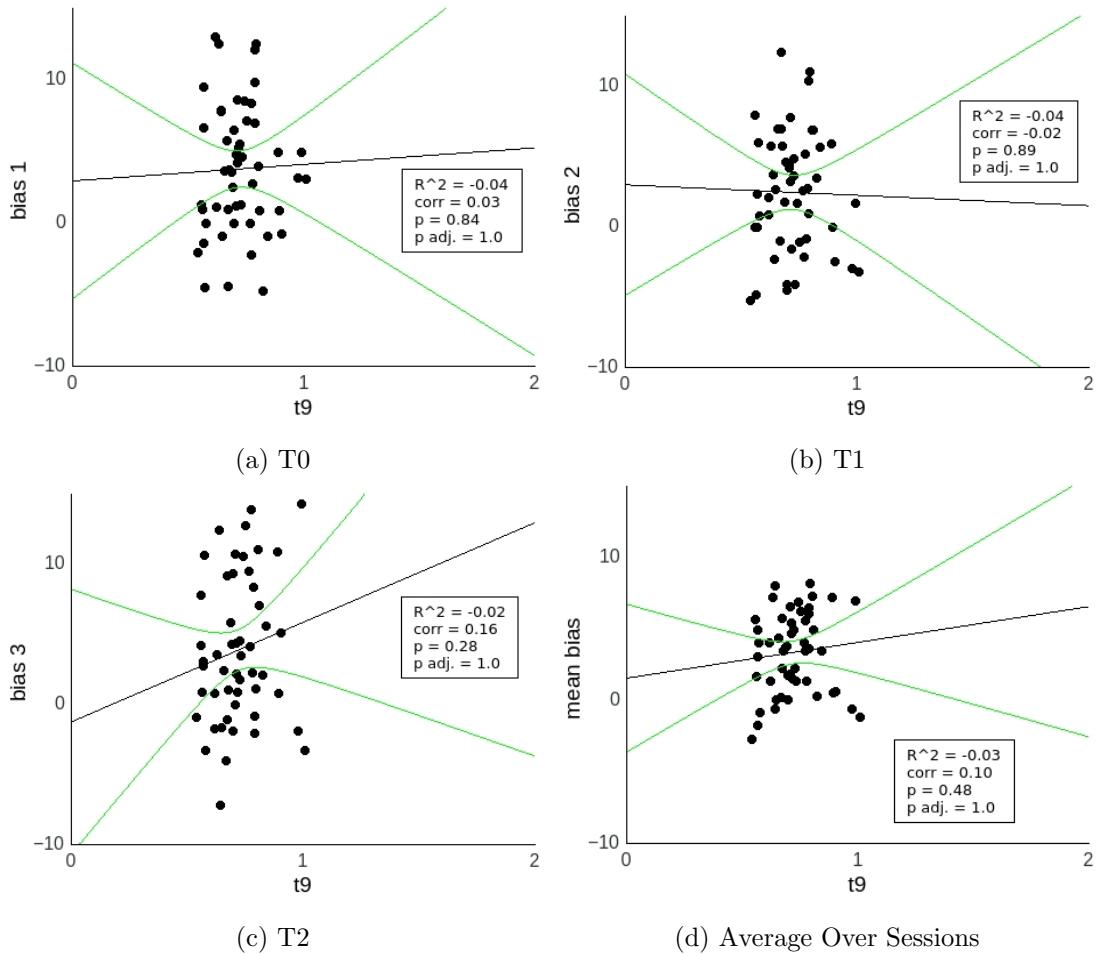


Figure 3.42: Functioning bias of full Model 2 simulations plotted over corresponding t_9 values for each session. t_9 outliers are removed. Lines indicate correlation fit and 95% confidence bands.

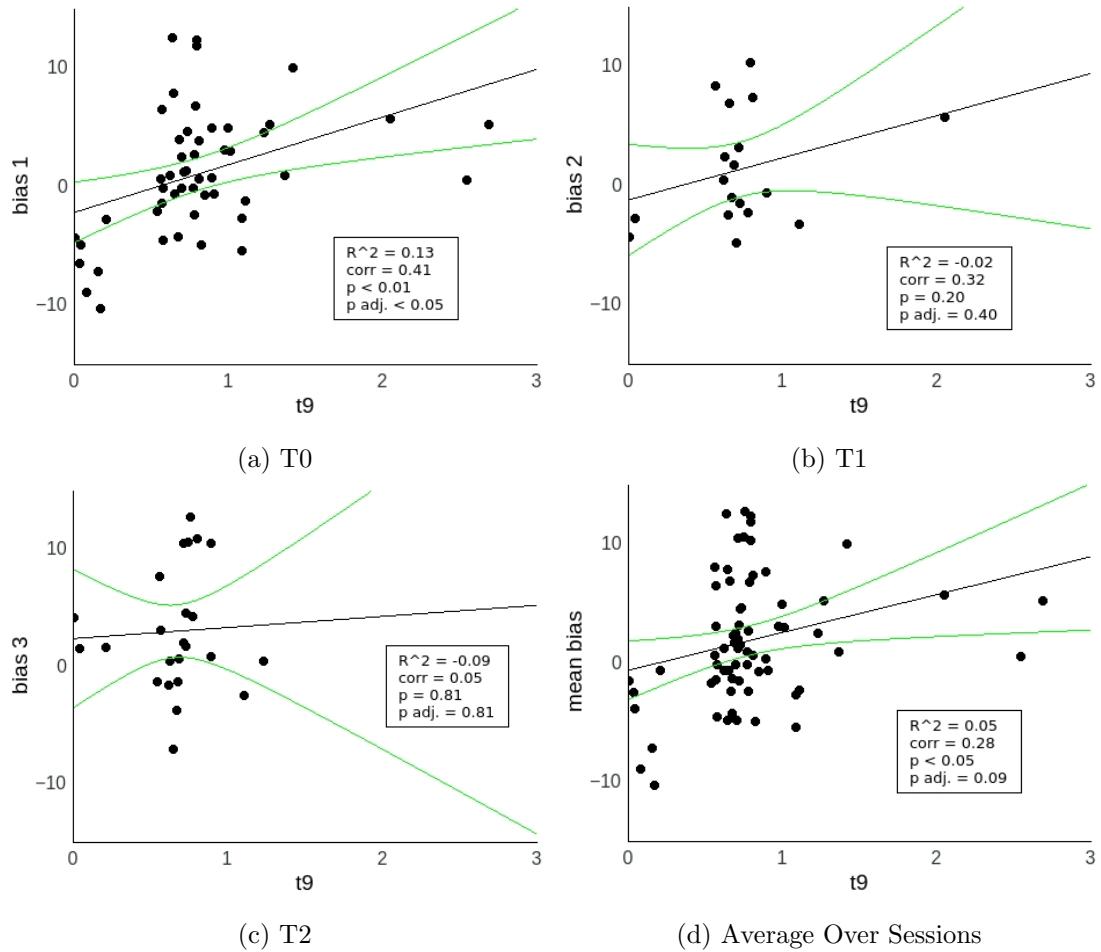


Figure 3.43: Functioning bias of experimental data plotted over corresponding t_9 values estimated by Model 2 for each session. Lines indicate correlation fit and 95% confidence bands.

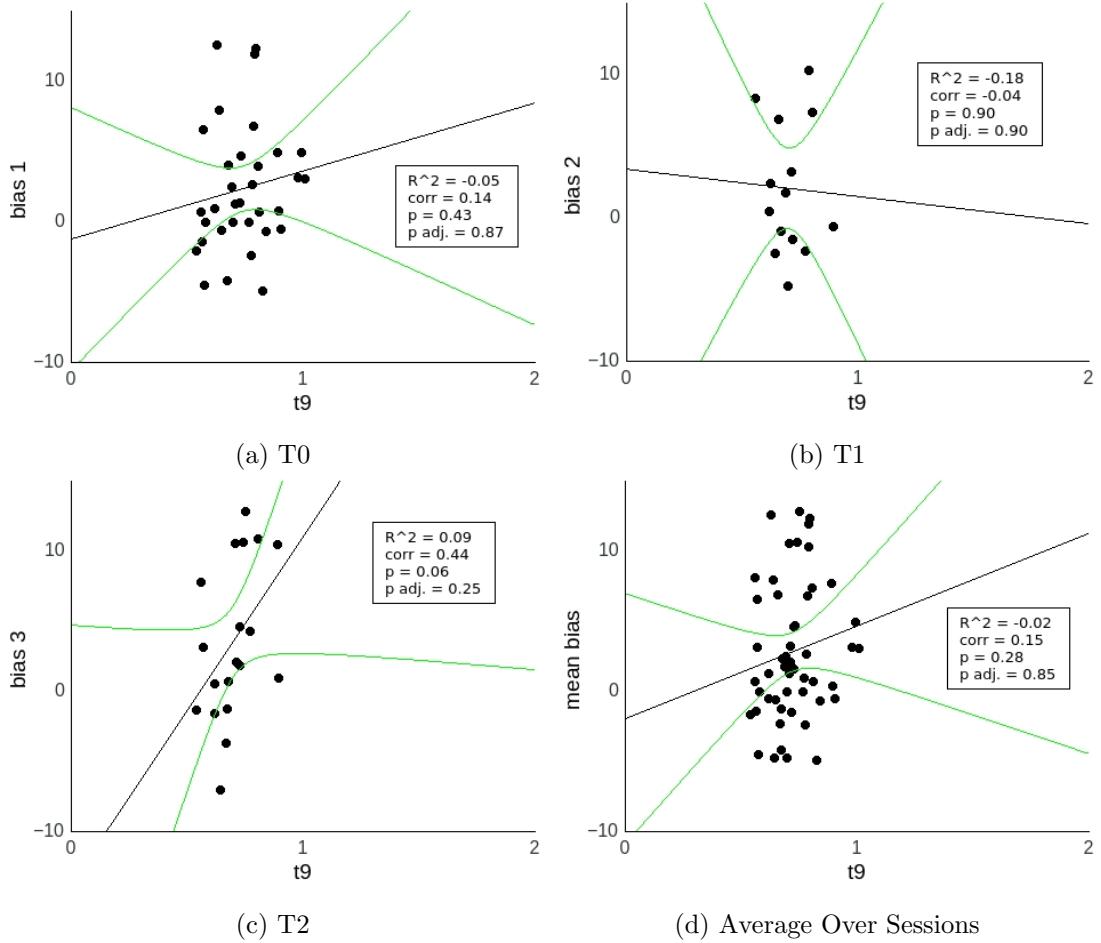


Figure 3.44: Functioning bias of experimental data plotted over corresponding t_9 values estimated by Model 2 for each session. t_9 outliers are removed. Lines indicate correlation fit and 95% confidence bands.

3.4.6 Conclusion

In this section, I analyzed how the speed of prediction adaptation affects the emergence of the functioning bias. To summarize the key results:

- Subjects with slow prediction adaptation consistently exhibit a significant functioning bias throughout all sessions, both in terms of the median split analysis (see Section 3.4.1) and the outlier analysis (see Section 3.4.3). These subjects show a significantly higher functioning bias level than those with fast prediction adaptation during T0 and when averaged over sessions.
- Both models indicate a correlation between the time scale of prediction adaptation and the functioning bias during T0 and when averaged over sessions (see Section 3.4.2). This implies a causal link: the slower the prediction adaptation, the higher the resulting functioning bias during T0 and on average.
- In general, the later the subjects' predictions reach the point of maximal uncertainty, the stronger their functioning bias throughout the experiment (see Section 3.4.4). Atypical cases are predicted by Model 1.
- According to Model 1, the later the contingency is finally acquired, the stronger their functioning bias throughout the experiment (see Section 3.4.5). The Model 2 data are inconclusive in this respect.

In summary, both models offer the same mechanistic explanation of the functioning bias: The functioning bias arises as a consequence of ongoing prediction adaptation. To generalize this statement and refer to the models' key assumption: *Progressing information maximization leads to the emergence of behavioral preferences.*

3.5 Exploration Rate

After investigating the role of prediction adaptation in the emergence of the functioning bias, I turn to the role of exploratory behavior: How does exploration contribute to the emergence of the behavioral bias?

3.5.1 Median Split

As before, the first analysis is based on splitting the sample of subjects along the median of the fitted exploration rate k_{exp} . The group of subjects with below-median k_{exp} is then termed *non-explorers*, those with above-median k_{exp} are the *explorers*. Figure 3.45 shows the functioning bias of these groups for each session based on both model fits:

- The simulation data of Model 1 indicate similar bias levels of both groups except during the first session, during which the explorers show no functioning bias whereas the non-explorers exhibit a significantly higher bias level (see Fig. 3.45a).
- The same subject grouping applied on the experimental data indicates a significant difference when averaged over sessions (see Fig. 3.45b). The non-explorers show biased behavior whereas the explorers don't.
- Based on the Model 2 simulations, no differences between the two groups are expected (see Fig. 3.45c).
- This is confirmed by the experimental data using the Model 2 median split (see Fig. 3.45d).

To summarize, Model 1 reveals a qualitative difference within the experimental data regarding the mean functioning bias: The non-explorers exhibit a functioning bias on average whereas the explorers don't. On the other hand, Model 2 does not indicate any difference.

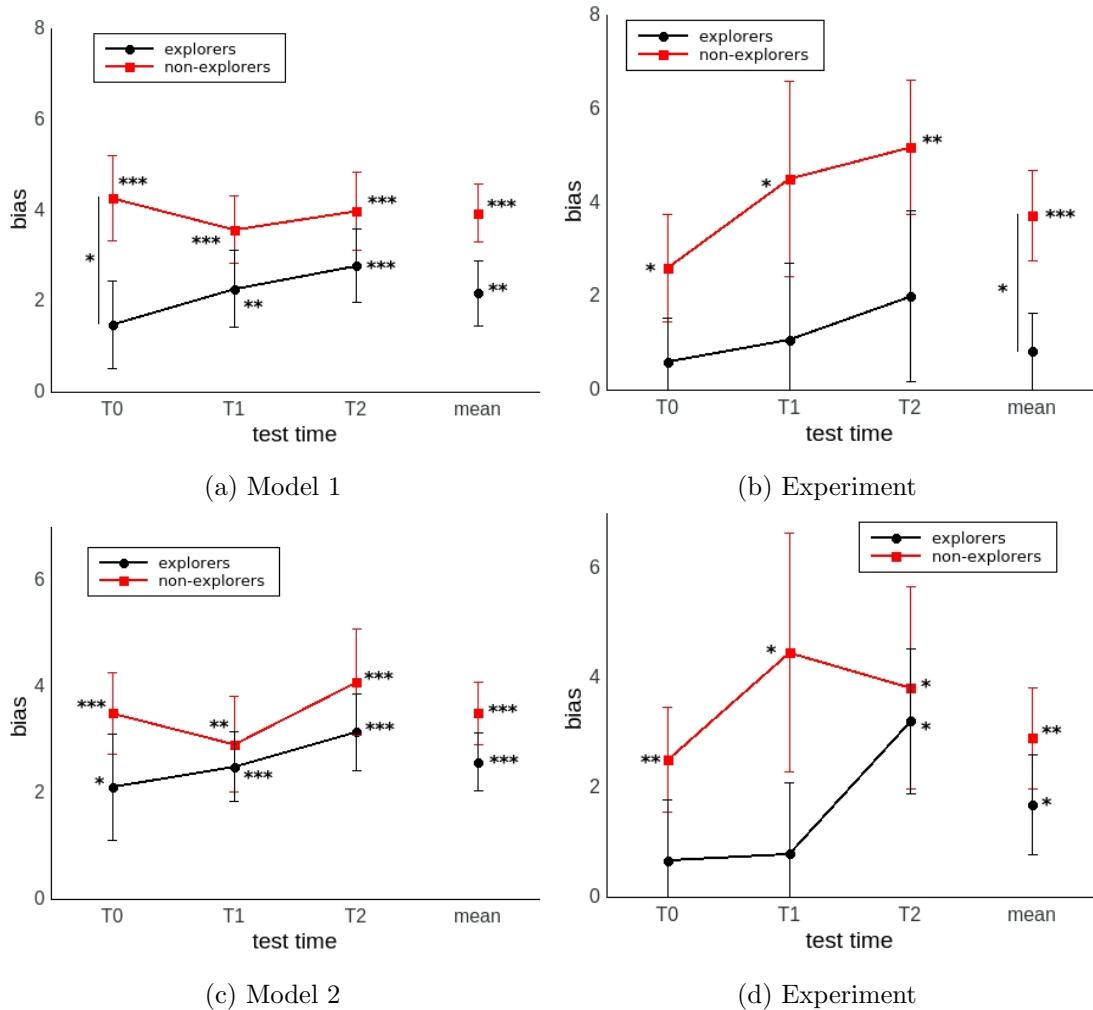


Figure 3.45: Functioning bias of subject groups defined by median split of fitted k_{exp} based on Model 1 fits (a,b) and Model 2 fits (c,d). The groups' bias is plotted using the full simulation data (a,c) and the corresponding experimental data (b,d).

3.5.2 Bias Correlations

To study the interaction between exploration and the functioning bias on a deeper level, I tested whether the bias is correlated with k_{exp} . The results indicate:

- Based on the simulation data, Model 1 predicts negative correlations for T0 and on average (see Fig. 3.46). However, these correlations largely depend on outliers, as indicated by testing after outlier removal (see Fig. 3.47).
- Based on the Model 1 fits, the experimental data cannot quite confirm these correlations (see Fig. 3.48). Clear tendencies arise during T0 and on average; the correlations are not significant after α -correction, though. These tendencies also vanish after removing outliers (see Fig. 3.49).
- Investigating Model 2 yields similar results. Based on the simulation data, Model 2 also predicts negative correlations for T0 and on average (see Fig. 3.50). Again, these correlations depend on outliers (see Fig. 3.51).
- In case of the Model 2 fits, the experimental data confirm these correlations (see Fig. 3.52). Once again, outlier removal eliminates these correlations (see Fig. 3.53).

To summarize, both models predict negative correlations between the functioning bias and k_{exp} during T0 and on average, i.e., the more the subjects exhibit exploratory behavior, the less they seem to prefer the functioning disc in these cases. The experimental data confirm these correlations based on the Model 2 fits but not on the Model 1 fits, though tendencies are observable in the latter case. In all cases, the correlations strongly depend on the presence of k_{exp} outliers, i.e., a small group of persistently exploring individuals, who exhibit a lower functioning bias than the rest of the subjects.

These results are not surprising, considering that exploration acts like noise and thereby tends to abolish clear biases.

One remaining question is: How do prediction adaptation and exploration interact within our sample of subjects?

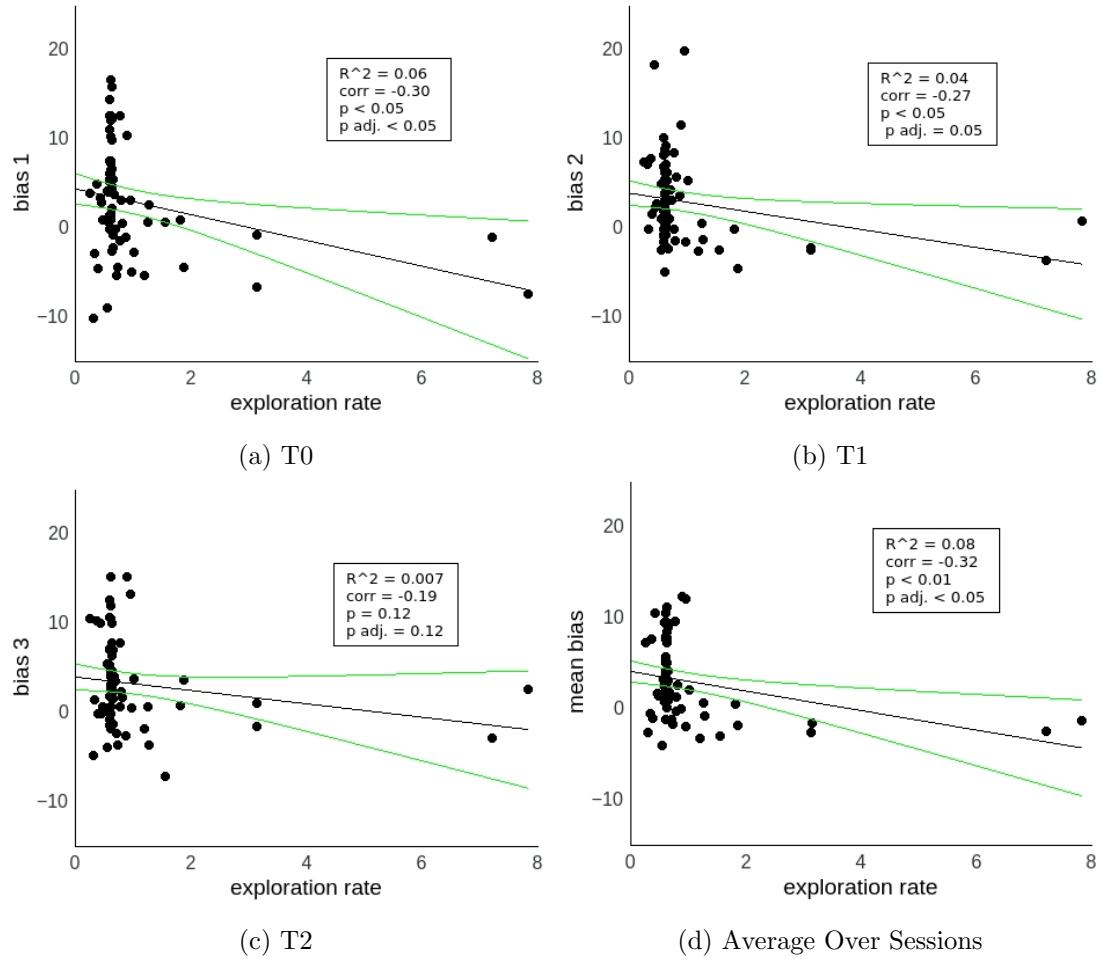


Figure 3.46: Functioning bias of full Model 1 simulations plotted over Model 1 k_{exp} fits for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.

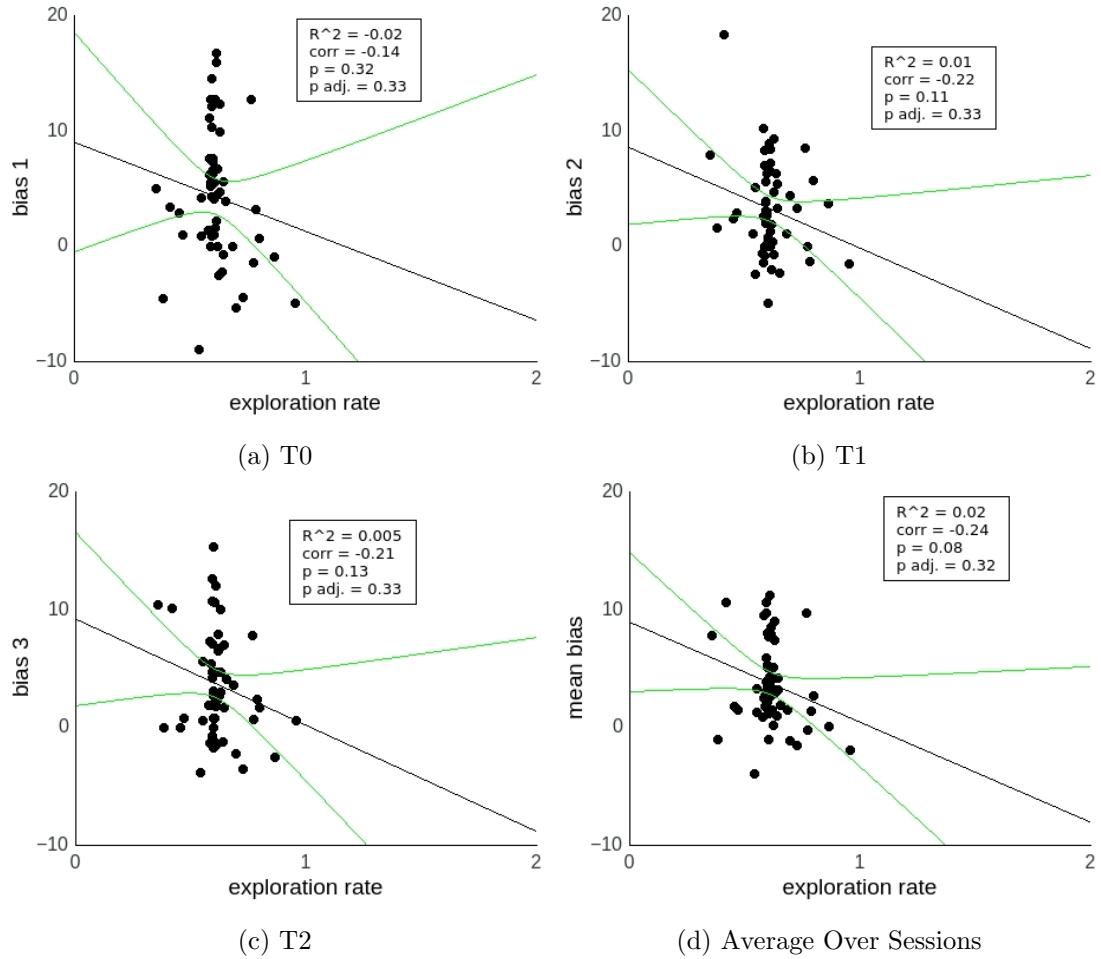


Figure 3.47: Functioning bias of full Model 1 simulations plotted over Model 1 k_{exp} fits without outliers for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.

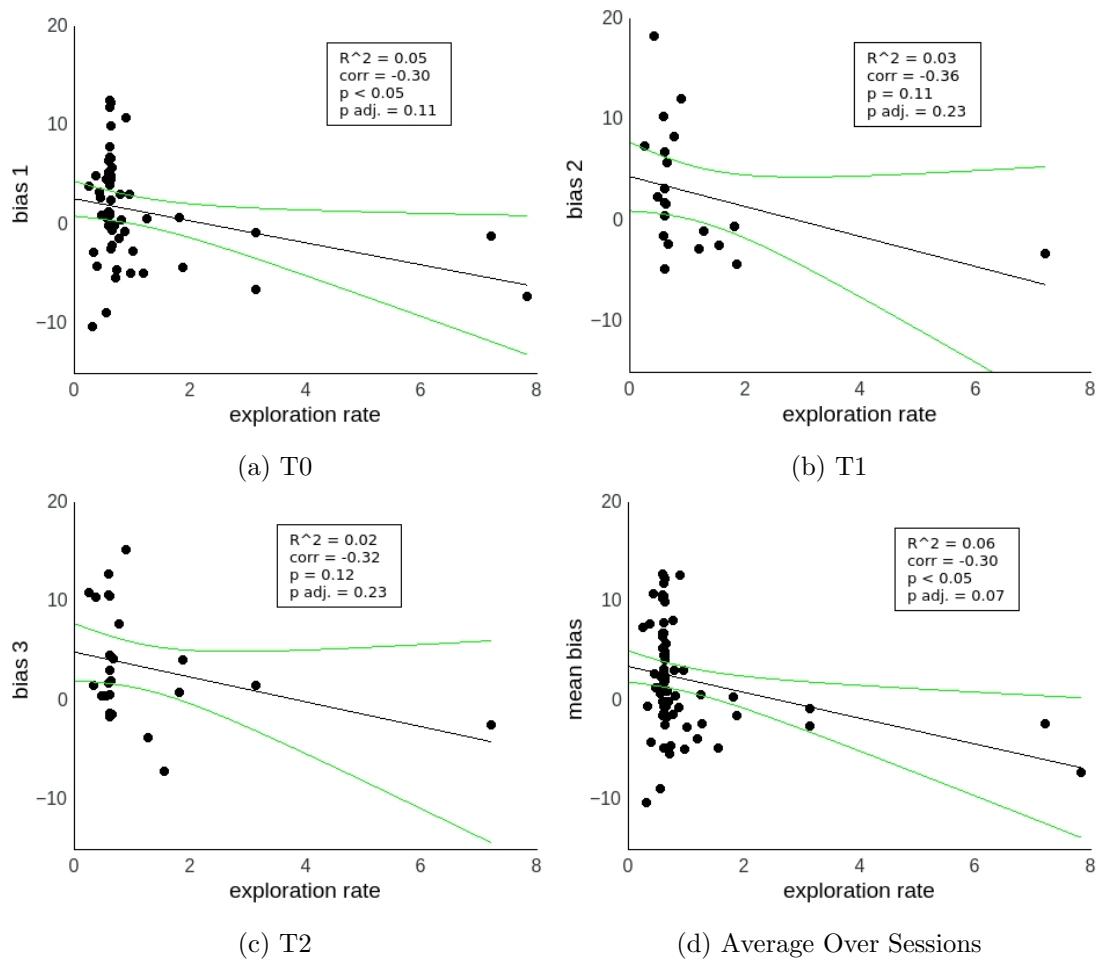


Figure 3.48: Functioning bias during experiments plotted over Model 1 k_{exp} fits for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.

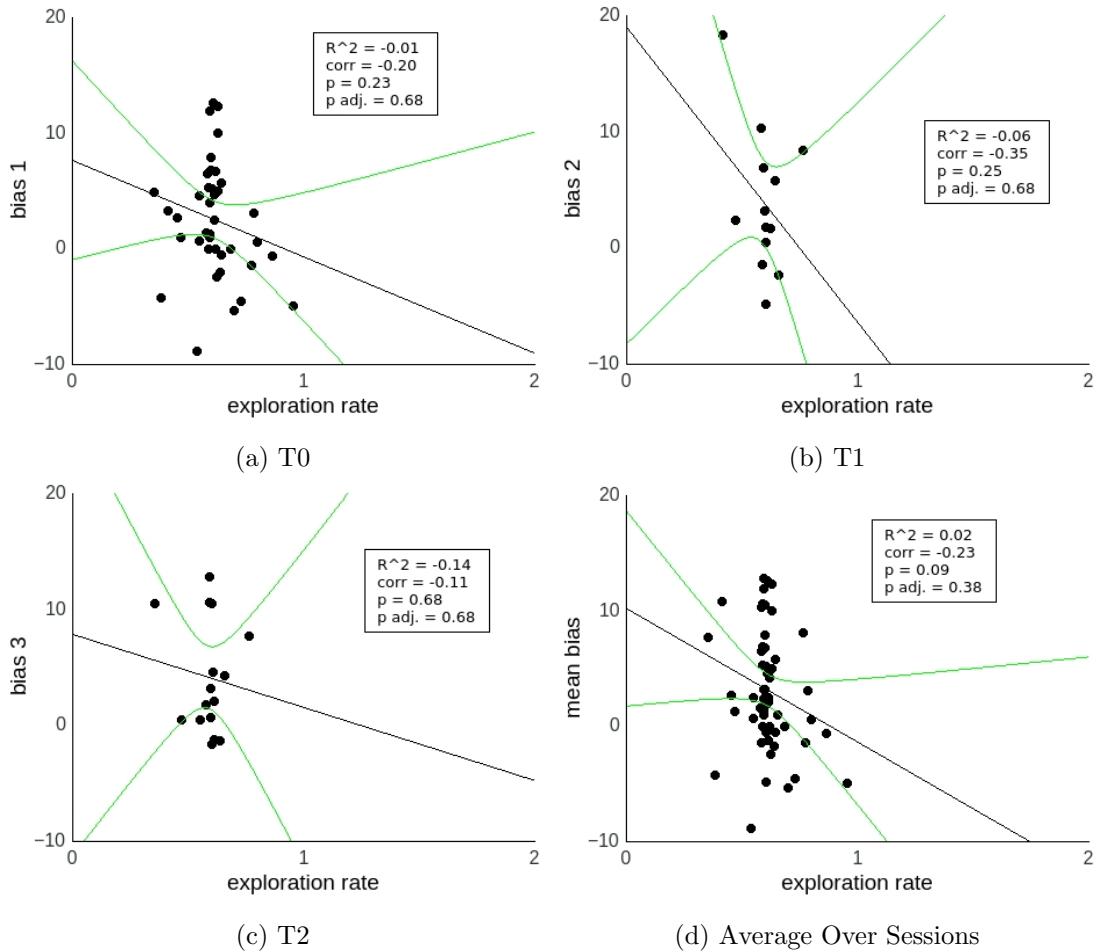


Figure 3.49: Functioning bias during experiments plotted over Model 1 k_{exp} fits without outliers for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.

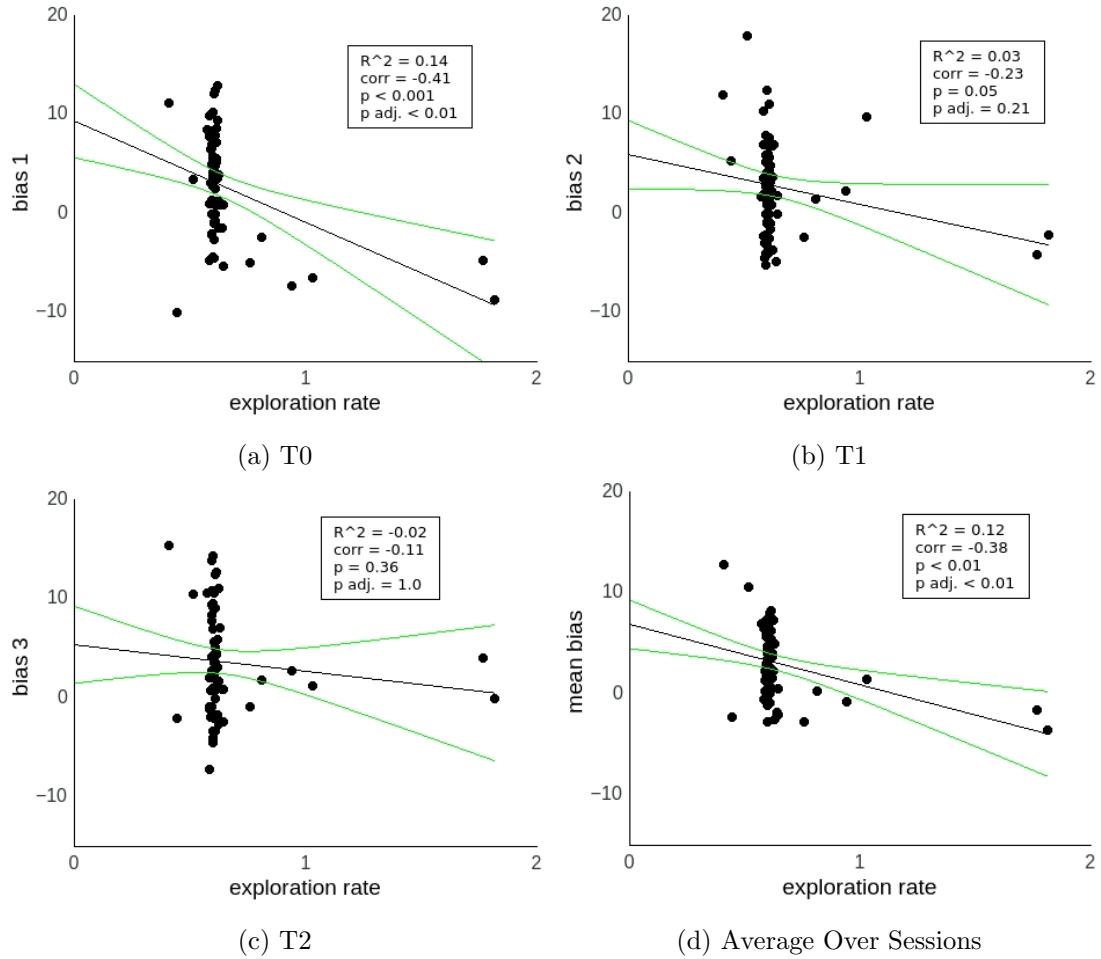


Figure 3.50: Functioning bias of full Model 2 simulations plotted over Model 2 k_{exp} fits for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.

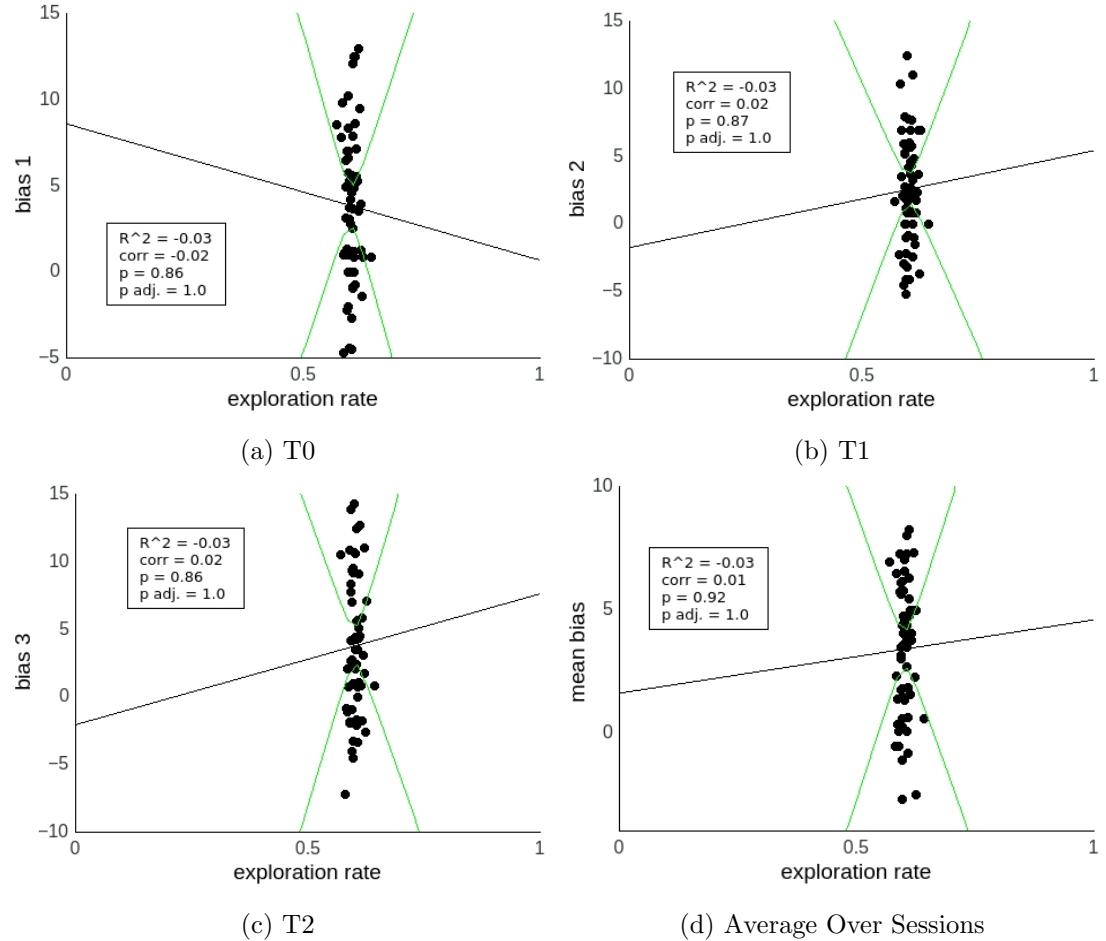


Figure 3.51: Functioning bias of full Model 2 simulations plotted over Model 2 k_{exp} fits without outliers for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.

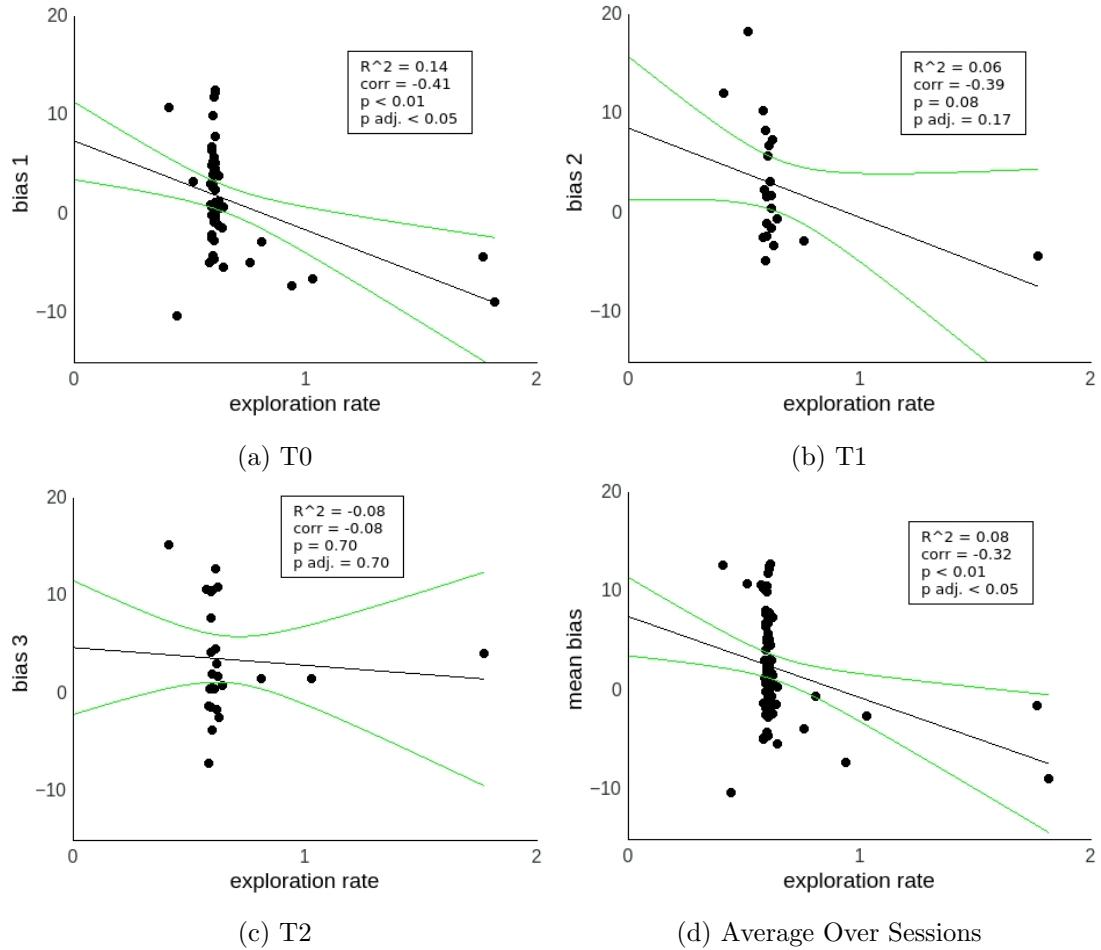


Figure 3.52: Functioning bias during experiments plotted over Model 2 k_{exp} fits for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.

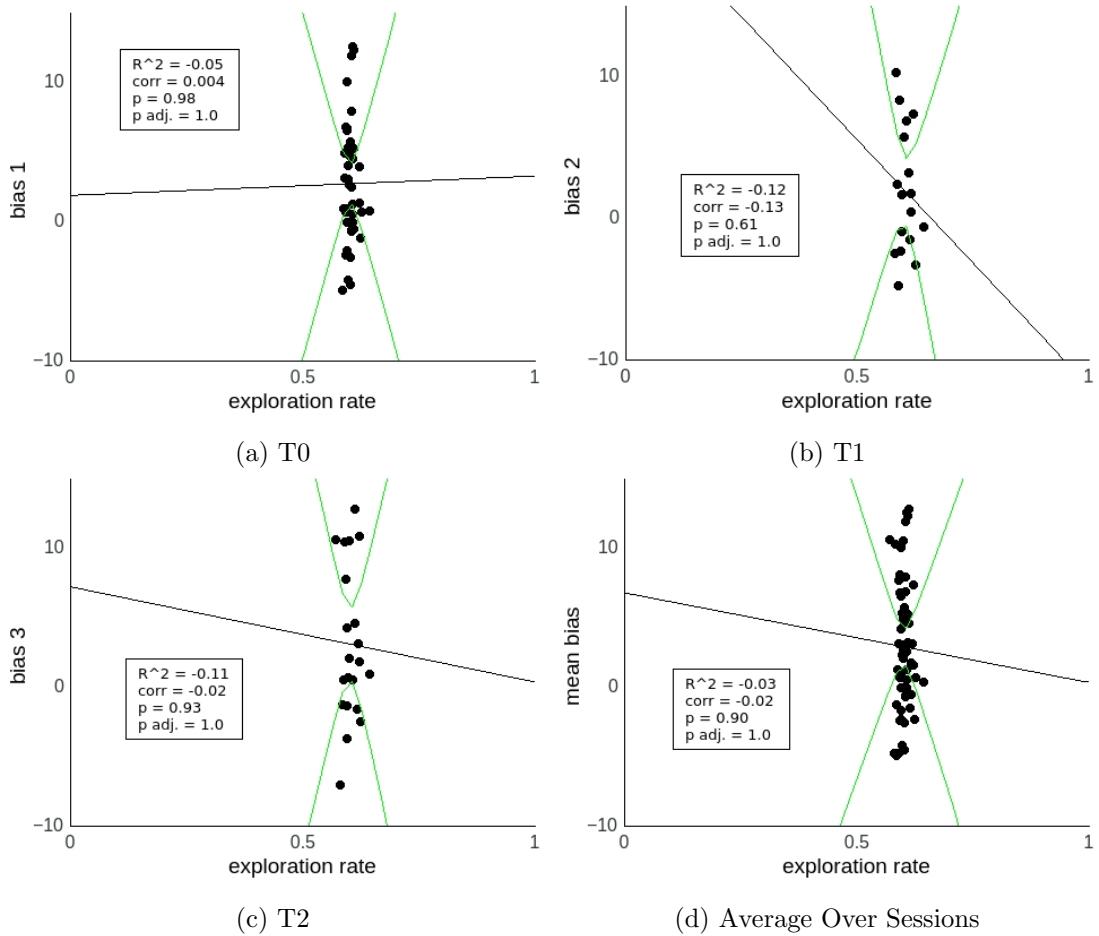


Figure 3.53: Functioning bias during experiments plotted over Model 2 k_{exp} fits without outliers for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.

3.5.3 Exploration and Prediction Adaptation

To study the relationship between prediction adaptation and exploration, I performed correlation analyses on the fitted τ_p and k_{exp} parameters. Figure 3.54 shows the results: Both models clearly indicate a negative correlation (see Figs. 3.54a,b). The Model 1 result is robust with respect to outlier removal (see Fig. 3.54c), whereas the correlation predicted by Model 2 depends on the outliers (see Fig. 3.54d).

This result makes sense on a pragmatic level: If you need more time to acquire a contingency, you should spend your time on improving the predictions and have less resources for exploring to discover novel effects. On the other hand, if you're a fast learner, you can afford more time exploring the environment. Notably, according to both models there are no explorers who are also slow learners, i.e., all subjects with slow prediction adaptation ($\tau_p \geq 0.8$) have relatively low exploration rates ($k_{exp} \leq 1$), which is in line with this interpretation.

Furthermore, this result offers the following explanation for the negative correlation between the functioning bias and the exploration rate discussed in the previous section: Subjects with high exploration rates show fast prediction adaptation, which leads to a reduced bias level (see Section 3.4). In other words, the negative correlation between the functioning bias and k_{exp} is the combined result of the positive correlation between the functioning bias and τ_p on the one hand and the negative correlation between τ_p and k_{exp} on the other hand. So the interaction between exploration rate and functioning bias might not indicate a direct, causal relationship, but rather a side effect of the direct link between prediction adaptation rate and functioning bias as well as the distribution of prediction adaptation rates and exploration rates within the subjects.

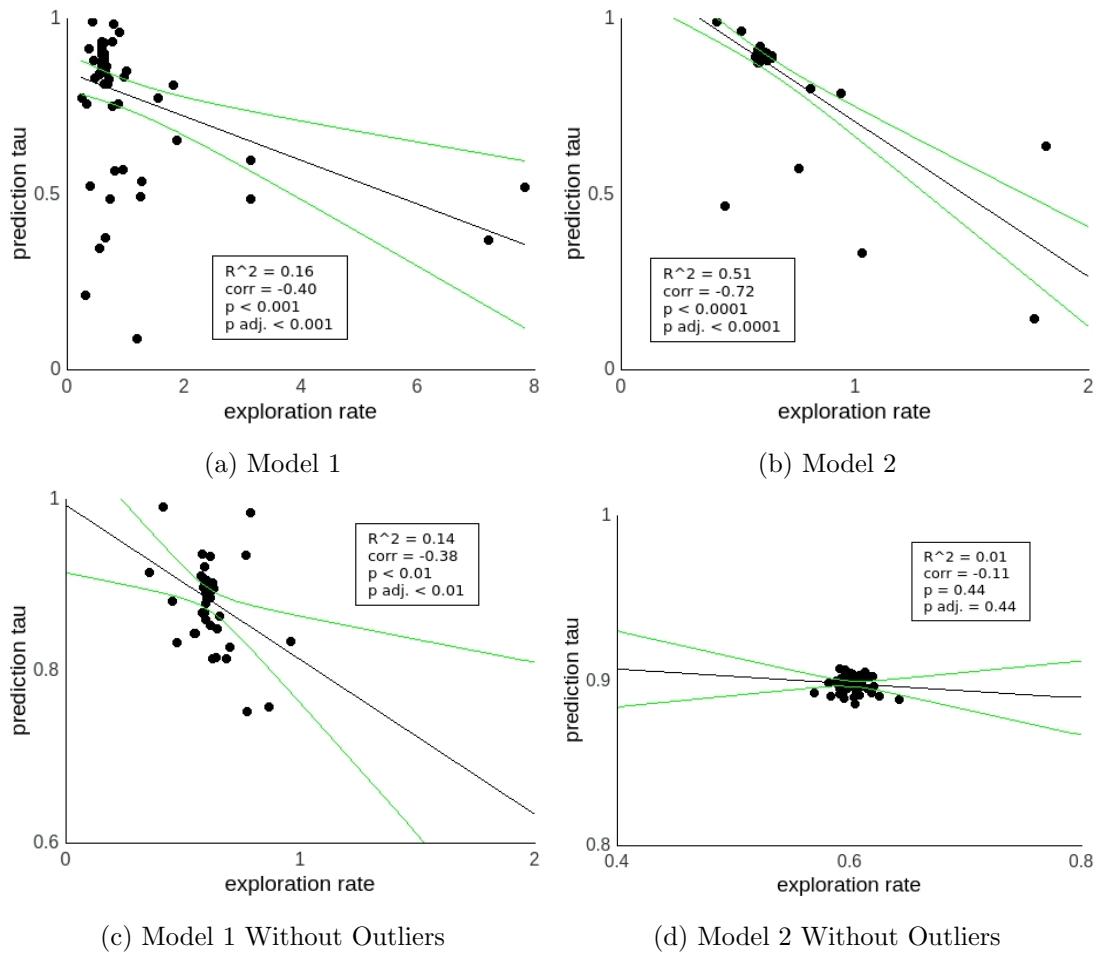


Figure 3.54: τ_p plotted over k_{exp} for each subject fitted by Model 1 (a) and Model 2 (b). Data are again plotted after removing outliers with respect to τ_p as well as outliers with respect to k_{exp} (c,d). Lines indicate correlation fit and 95% confidence bands.

4 Discussion

Why do infants develop a behavioral bias during gaze-contingent learning? How do inter-individual differences within their behavior arise? What are the roles of prediction and exploration?

To tackle these questions, I applied two computational models that allowed to simulate and reproduce the behavior of individual infants during gaze-contingent learning. Both models are based on the same key assumption: *Human beings are intrinsically motivated to optimize their predictions concerning interactions with the environment.* Specifically, *they act to maximize the amount of information they receive, based on their beliefs and expectations.*

Model 1 is an adaptation of previous work, which captured intrinsically motivated contingency learning in rats [59]. Apart from prediction adaptation, it simulates biological processes on the level of neuronal populations and includes neuronal adaptation based on dopamine-modulated plasticity. I contributed to this model by reformulating its previously heuristic intrinsic motivation aspect in terms of information theoretical measures. On the other side, Model 2 is an abstracted version of Model 1. It shares Model 1's prediction system but abstracts away certain biological details. It is more focussed on the functional aspects of prediction adaptation, less complex, and therefore suitable from a theoretical point of view.

Both models were able to reproduce each individual subject's behavior on the level of observed biases (see Section 3.1). The models also reproduced the qualitatively different behavior between the 6-month-olds and the 8- and 10-month-old subjects, i.e., a consistent functioning bias within the older subjects versus no or a very small bias within the 6-month-olds (see Section 3.2). As to the basis of this difference, the models remain inconclusive.

On the level of individual subject's behavior, the models indicated differences in learning dynamics that lead to qualitatively different behavioral patterns (see Section 3.3). This provided the first clue to the significance of the speed of prediction adaptation for the emergence of the behavioral bias. Multiple analyses revealed that the slower the prediction adaptation, the later the contingencies are acquired, and the stronger the behavior is biased during the experiment (see Section 3.4). This established the key result: *The functioning bias arises as a consequence of ongoing prediction adaptation.* In more general terms: *Progressing information maximization leads to the emergence of behavioral preferences.*

This result is in line with the repetition bias hypothesis put forward by Redgrave and Gurney [28]. They propose that action selection must be biased in favor of actions whose sensory outcomes are unpredicted for as long as the predictions can be improved. In this view, this transient repetition bias reflects ongoing prediction adaptation and is

abolished once the environmental responses match the internal predictions. Redgrave and Gurney locate this behavioral adaptation in the interface between the cortex and the basal ganglia, which are thought to be involved in action selection [100, 101, 102, 103]. This adaptation is realized by synaptic plasticity, which is modulated by dopamine, a neurotransmitter known to signal reward prediction errors [80]. However, they concede that this is only one way behavioral adaptation may take place. The other way is by changes of the cortical salience signals themselves. However, they don't make any explicit statements how those changes may come about; their focus of interest lies on the cortico-striatal plasticity. In order to generalize their framework to situations without external rewards, Redgrave and Gurney hypothesize a more general role for dopamine, namely as an indicator of a sensory prediction error [104]. Ultimately, Model 1 was devised to test these hypotheses in the context of an intrinsically motivated learning task in rats [59], to which it could be successfully be applied. The fact that both Model 1 and Model 2 are able to reproduce subjects' behavioral bias during gaze-contingent learning indicates that dopaminergic signalling might not be a crucial factor; instead, action selection based on prediction optimization appears to be sufficient for the formation of the repetition bias. This might seem strange, considering dopamine's role in signalling the prediction error. However, at least in the dynamical regime the model is operating in in this study, the modulated plasticity of the cortico-striatal projections seems to play a minor role compared to the novelty salience signal produced by the prediction system. In any case, prediction impacts behavior, either through the dopaminergic prediction error or through the expected information gain directly evaluated by the prediction system.

Finally, investigating the subjects' exploratory behavior indicated that slow learners explored less than fast learners tended to do (see Section 3.5). This is in line with the exploration-exploitation dilemma of reinforcement learning [61]: An agent who tries to maximize his reward may either exploit his knowledge and execute actions yielding a known reward, or he may explore by choosing actions with lower estimated rewards to maybe eventually find actions yielding an even higher reward. This is a fundamental consideration in problems with large action and state spaces, but it is also significant for small action and state spaces in case of non-stationary problems, i.e., if the rewards are changing over time. In this case, optimal action choices may change over time, and exploration is important to constantly evaluate competing actions. The equivalent of the reward in the presented models is the salience, which is dominated by the expected information gain (novelty salience), a non-stationary quantity. Because the initial expected information gain for each action is 0, exploration is important to find actions with unpredicted outcomes in the first place. As the information gain increases during contingency learning, exploitation dominates behavior to drive prediction improvement. Once the predictions are sufficiently optimized, the expected information gain is low, and exploratory behavior becomes important again to detect novel contingencies. As a consequence, slow learners, who spend most of their time exploiting (improving their predictions), are less dependent on exploration than fast learners, who rapidly acquire the contingency and may be motivated to search for novel contingencies afterwards.

In general, the observed effects were most pronounced during T0 and all times on average. On the one hand, this is explained by the smaller sample size during T1

(N=21) and T2 (N=25) compared to T0 (N=53) and the whole sample (N=71, see Table 2.1). On the other hand, this also reflects the fact that the analyzed subjects might have slightly different experiences during subsequent sessions. During T0, no analyzed subject had any prior experience with the experiment. The sample analyzed for T1 only contained subjects from the active-active group, so it was assumed they all had experienced the fully gaze-contingent experiment during the previous session. However, for 11 out of 21 of these subjects the first session had an estimated failure rate larger than 10% (see Table 2.2). So even though they faced the same experimental condition before, the contingency was less reliable, possibly resulting in different experiences and learning processes. For T2, the situation is even more complex: Only 6 out of the 25 analyzed subjects encountered highly reliable contingencies during the previous sessions, whereas 7 subjects only experienced less reliable ones before. This sample heterogeneity during T1 and T2 needs to be taken into account when interpreting the results for these sessions. In the end, the same effects as for T0 might be present during T1 and T2, but they might be harder to extract within these smaller and more heterogeneous samples. On the other hand, the homogeneity of the T0 data lends confidence to the analysis of this sample.

Another recurring observation throughout the analyses was the sharp-peaked distribution of fitted parameter values for Model 2, resulting in a strong dependence of observed correlations on outliers. Model 2 is far simpler and less constrained than Model 1, which is why one would expect the Model 2 parameters to vary more than for Model 1. However, we observe the opposite. With the exception of the synaptic learning rate in Model 1, both models share the same set of fitting parameters (see Tables 2.3 and 2.7). In conclusion, the biological processing including the synaptic plasticity seems to broaden the dynamic range in which Model 1 operates and produces behavior that fits the experimental data.

This study is based on models of *intrinsically motivated* behavior. Even though they take into account the extrinsic visual salience of the screen areas, they don't include inference processes for optimizing extrinsic rewards. Since the experiment was designed to provide visually attractive incentives (the animal pictures) for the subjects, another plausible, complementary approach would be to interpret gaze-contingent learning not as prediction optimization (*intrinsic motivation*), but rather as maximization of visual rewards (*extrinsic motivation*). In the latter case, inference of the values of actions and sensory states would be used to learn optimal behaviors for receiving external rewards. This way, the agent would also acquire the contingency between the functioning disc and the animal picture, but on a more implicit level, i.e., as part of state and action values (in terms of conventional reinforcement learning algorithms). This is then expected to generate behavior that also exhibits a functioning bias. However, a major difference between these approaches would be the resulting behavior after learning: Whereas the intrinsically motivated agent stops preferring the functioning disc after contingency acquisition and searches novel, unexpected contingencies to acquire, the extrinsically motivated agent would continue preferring the functioning disc for as long as fixating it triggers novel animal pictures. This results from the fact that the extrinsically motivated learning problem is well approximated by a stationary model, whereas the intrinsically

motivated learning problem is non-stationary (see above). Experimentally disentangling the significance of the external incentive and the intrinsic motivation is non-trivial since the intrinsic motivation depends on the occurrence of unpredicted sensory events, which are to some degree always inherently salient. Ultimately, a more realistic and comprehensive model might contain both an intrinsic aspect of motivation and an inference, or planning, process focussing on optimizing extrinsic rewards.

One limitation of the presented models is their susceptibility to noisy processes. The outcome of a perfect noise source is maximally uncertain and thereby always carries the most information. The model should therefore develop a strong bias for producing such noisy outcomes for all times, even though the internal predictions are not improved after an initial learning phase. This problem of unlearnability has been extensively discussed in the literature [92, 8]. However, in this study this limitation is not a major problem since the modelled data were filtered by contingency reliability such that only sessions were considered during which fixations of the functioning disc triggered animal pictures at least 90% of the time. This way, only a highly reliable contingency needed to be considered. To make the model more generally applicable, one could formulate the intrinsic motivation in terms of *learning progress*, i.e., a measure of how much the internal model improves over time [19, 18]. In the case of the perfect noise source, the agent would then act to trigger the noisy outcomes until the internal prediction matches the real outcome uncertainty. Even though the expected information gain would remain high, prediction adaptation and thus the learning progress would then vanish and the behavioral bias steadily decrease, allowing the agent to detach from the noise source and start exploring the environment again.

To formulate intrinsic motivations based on information gain is not a new idea. In his seminal work leading back to the 1990s, Schmidhuber proposed an implementation of curiosity based on prediction improvement [19, 92]. Specifically, his agent aims to maximize the compression of his internal representations and thereby acts to maximize his information intake. His work is cited as one of the most influential in the field as he has developed a coherent theory on intrinsic motivations from the perspective of machine learning [10], and his ideas about prediction improvement left clear traces within my own work.

Another model that was used to explain gaze-contingent learning was proposed by Marrappa et al. [105]. They modelled the one-disc experiment version of Wang et al.'s gaze-contingency paradigm [1] and could successfully reproduce anticipatory gaze shifts observed during the experiments. Because different experiment versions were modelled, the modelling results are difficult to compare: Marrappa et al.'s focus lay on the temporal dynamics and emerging anticipatory gaze behavior, which could not be evaluated in the two-disc version due to the slow fading out of the animal picture. On the other hand, my work focusses on the development of a behavioral bias, which could not be measured in the one-disc version due to the lack of the nonfunctioning control disc. On the conceptual level of modelling, a common theme exists in the integration of bottom-up and top-down attentional control. A major difference lies in the assumed motivation of the modelled agents: Whereas my work emphasizes prediction optimization as the intrinsic motivation driving gaze-contingent learning, their approach interprets gaze-contingent learning as

visual saliency-focussed, extrinsically motivated behavioral adaptation. Ultimately, even though gaze-contingent learning is addressed in both cases, the two modelling approaches differ widely in their assumptions as well as the questions they answer.

An interesting connection can be drawn between this work and the work by Bruce and Tsotsos, who have proposed the Attention by Information Maximization (AIM) model [106]. This model was also created to explain gaze behavior in humans, and it also draws on the idea of behavior guided by information maximization. However, in my approach the models evaluate information content with respect to an internal model of environmental processes, whereas the AIM model's notion of information is based on the statistics of the visual stimulus. These two approaches exemplify the classic distinction between top-down and bottom-up processes: My approach interprets gaze behavior as prediction- or knowledge-driven (top-down), whereas the AIM model's behavior is driven purely by properties of the current stimulus (bottom-up)¹. However, an interesting extension of the models I applied would be to implement the AIM model for extracting the intrinsic saliences of the visual input. In the current version, the visual saliences are estimated in a very simplistic and crude fashion. Using a detailed bottom-up model such as AIM would allow applications on a much more realistic level and possibly even on a robotic platform, and it would realize a two-fold implementation of a single principle, namely behavior based on information maximization.

¹In my approach, bottom-up processing is present as well, as implemented by the intrinsic salience signal (see Section 2.3.2)

List of Figures

2.1	(a) Schematic of the stimulus set-up during the gaze-contingent learning task. Once the subject fixates the functioning disc, an animal picture appears within 80 ms and fades out within 17 seconds. Adapted from [1]. (b) Estimating trigger failures. In case of inaccurate eye tracking, there is a discrepancy between where the subject perceives the red disc (striped circle) and where the eye tracker defines it (red disc). Intended disc fixations fall within the striped circle and are missed by the tracker if they fall within the gray area (false negatives). If more than 10% of all intended functioning disc fixations are false negatives, the test data are discarded.	9
2.2	Spatial distribution of fixations during a gaze-contingent test. The black dots indicate recorded fixations and the color map the estimated fixation density. The red circles and the central rectangle mark the actual positions of the red discs and the central image area, respectively. The open circle on the right marks the fitted subjective functioning disc fitted around the local maximum in the right portion of the screen. 4 fixations out of 33 within the subjective disc are marked as trigger failures, yielding a failure rate of 12%. Adapted from [88].	11
2.3	Schematic of Model 1. The prediction system stores and processes the internal beliefs about the environment's behavior. It assigns to each possible action a novelty salience, which is combined with intrinsic salience and fed into the neural circuit model. Depending on its output, the current action is chosen and executed. Based on the environment's response, the prediction system is updated and the next iteration begins. Thin arrows denote propagation of scalar values, thick arrows indicate information flow within each parallel action channel.	13
2.4	Processing within the prediction system. (A) Cartoon of a prediction trajectory. Every time the considered action is executed (circles), the prediction variable is updated based on the sensory outcome (red bars). (B) Mapping from prediction onto novelty salience. The binary entropy function indicates the expected information gain based on the prediction variable, so uncertain outcomes yield higher saliences. The function is scaled for model consistency. (C) The novelty salience corresponding to the prediction variable in (A) . Based on the mapping function in (B) , this salience reflects the evolution of the expected information gain for this action. (D) The sensory prediction error corresponding to the prediction variable in (A) . Adapted from [59].	14

2.5	(A) Schematic of the overall neural circuit model indicating the sensory and motor cortex, the thalamic system comprising the thalamic reticular nucleus (TRN) and the ventrolateral thalamus (VL), and the basal ganglia and brainstem. (B) Schematic of the basal ganglia network model indicating subthalamic nucleus (STN), globus pallidus internal and external segment (GPi and GPe), substantia nigra pars reticulata and pars compacta (SNr and SNc). Compartments represent action channels with population activity level (gray bars). Note that four action channels are present in my model implementation. Adapted from [59].	17
2.6	Schematic of Model 2, which is essentially Model 1 without the neural circuit model (see Fig. 2.3, same conventions).	23
3.1	Aggregate fixation distribution during all gaze-contingent tests (a) and yoked control tests at T0 (b). The central rectangle indicates the location of the animal pictures, the circles the locations of the red discs. In (a), fixation locations during tests with the functioning disc on the left were mirrored along the central vertical axis such that the functioning disc is visualized on the right. The density estimation is described in Section 2.2. Adapted from [88].	26
3.2	(a) Functioning bias of all subjects tested in the gaze-contingent condition. (b) Experimental data and their model fits. (c) Experimental and full simulation data.	27
3.3	Histograms of subjects' fitness errors for model 1 (a) and model 2 (b). . .	28
3.4	Empirical functioning bias of different age groups during individual tests (a-c) and averaged over tests (d). Dashed lines indicate level of unbiased behavior.	29
3.5	Functioning bias of different age groups during individual tests (a-c) and averaged over tests (d) as estimated by model 1. Dashed lines indicate level of unbiased behavior.	30
3.6	Functioning bias of different age groups during individual tests (a-c) and averaged over tests (d) as estimated by model 2. Dashed lines indicate level of unbiased behavior.	31
3.7	Empirical functioning bias of different age groups during individual tests (a-c) and averaged over tests (d). Dashed lines indicate level of unbiased behavior.	32
3.8	Functioning bias of different age groups during individual tests (a-c) and averaged over tests (d) as estimated by model 1. Dashed lines indicate level of unbiased behavior.	33
3.9	Functioning bias of different age groups during individual tests (a-c) and averaged over tests (d) as estimated by model 2. Dashed lines indicate level of unbiased behavior.	34
3.10	Distributions of the fitted parameters τ_p (a,b), k_{ret} (c,d), and k_{exp} (e,f) for 6-month-olds and older subjects.	35

3.11	Distributions of the fitted synaptic learning rate η for 6-month-olds and older subjects. This parameter is not present in Model 2.	36
3.12	Time evolution of the prediction variable y^* (black) and the corresponding novelty salience (red) during simulated experiments for three individual subjects. Vertical dashed lines indicate session onsets. Note that a decrease of y^* is either due to forgetting between T1 and T2 or due to trigger failures (see Section 2.2), which contradict the contingency and lead to transient reversals of the prediction adaptation.	38
3.13	Functioning bias of the subjects of Fig. 3.12: experimental data (a) and simulated behavior (b and c). Dashed lines indicate level of unbiased behavior.	40
3.14	Prediction variables of Fig. 3.12 plotted over normalized experiment time. Vertical dashed lines indicate session onsets. The green dashed line marks $y^* = 0.5$; crossing this line for the first time defines a subject's t5. The purple dashed line marks $y^* = 0.9$; crossing this line for the first time defines a subject's t9.	41
3.15	Functioning bias of subject groups defined by median split of fitted τ_p based on Model 1 fits (a,b) and Model 2 fits (c,d). The groups' bias is plotted using the full simulation data (a,c) and the corresponding experimental data (b,d). The dashed lines indicate the level of unbiased behavior.	43
3.16	Functioning bias of full Model 1 simulations plotted over Model 1 τ_p fits for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.	44
3.17	Functioning bias of full Model 1 simulations plotted over Model 1 τ_p fits without outliers for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.	45
3.18	Functioning bias during experiments plotted over Model 1 τ_p fits for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.	46
3.19	Functioning bias during experiments plotted over Model 1 τ_p fits without outliers for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.	47
3.20	Functioning bias of full Model 2 simulations plotted over Model 2 τ_p fits for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.	48
3.21	Functioning bias of full Model 2 simulations plotted over Model 2 τ_p fits without outliers for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.	49
3.22	Functioning bias during experiments plotted over Model 2 τ_p fits for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.	50
3.23	Functioning bias during experiments plotted over Model 2 τ_p fits without outliers for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.	51

3.24 Functioning bias of lower τ_p outliers (smaller than first quartile minus 1.5 times interquartile range) with respect to τ_p values fitted using Model 1 (a,b) and Model 2 (c,d) as well as inliers (remaining samples). The groups' functioning bias is plotted using the full simulation data (a,c) and the corresponding experimental data (b,d). The dashed lines indicate the level of unbiased behavior.	54
3.25 t5 values extracted from simulated experiments based on Model 1 (a) and Model 2 (b) plotted over the corresponding fitted τ_p values. The same data are plotted after removing both the t5 and τ_p outliers (c,d). Lines indicate correlation fit and 95% confidence bands.	55
3.26 The distributions of t5 values of simulated Model 1 outliers and inliers (a) and Model 2 outliers and inliers (b).	56
3.27 Functioning bias of full Model 1 simulations plotted over corresponding t5 values for each session. Lines indicate correlation fit and 95% confidence bands.	57
3.28 Functioning bias of full Model 1 simulations plotted over corresponding t5 values for each session. t5 outliers are removed. Lines indicate correlation fit and 95% confidence bands.	58
3.29 Functioning bias of experimental data plotted over corresponding t5 values estimated by Model 1 for each session. Lines indicate correlation fit and 95% confidence bands.	59
3.30 Functioning bias of experimental data plotted over corresponding t5 values estimated by Model 1 for each session. t5 outliers are removed. Lines indicate correlation fit and 95% confidence bands.	60
3.31 Functioning bias of full Model 2 simulations plotted over corresponding t5 values for each session. Lines indicate correlation fit and 95% confidence bands.	61
3.32 Functioning bias of full Model 2 simulations plotted over corresponding t5 values for each session. t5 outliers are removed. Lines indicate correlation fit and 95% confidence bands.	62
3.33 Functioning bias of experimental data plotted over corresponding t5 values estimated by Model 2 for each session. Lines indicate correlation fit and 95% confidence bands.	63
3.34 Functioning bias of experimental data plotted over corresponding t5 values estimated by Model 2 for each session. t5 outliers are removed. Lines indicate correlation fit and 95% confidence bands.	64
3.35 t9 values extracted from simulated experiments based on Model 1 (a) and Model 2 (b) plotted over the corresponding fitted τ_p values. The same data are plotted after removing both the t9 and τ_p outliers (c,d). Lines indicate correlation fit and 95% confidence bands.	65
3.36 The distributions of t9 values of simulated Model 1 outliers and inliers (a) and Model 2 outliers and inliers (b).	66

3.37	Functioning bias of full Model 1 simulations plotted over corresponding t9 values for each session. Lines indicate correlation fit and 95% confidence bands.	67
3.38	Functioning bias of full Model 1 simulations plotted over corresponding t9 values for each session. t9 outliers are removed. Lines indicate correlation fit and 95% confidence bands.	68
3.39	Functioning bias of experimental data plotted over corresponding t9 values estimated by Model 1 for each session. Lines indicate correlation fit and 95% confidence bands.	69
3.40	Functioning bias of experimental data plotted over corresponding t9 values estimated by Model 1 for each session. t9 outliers are removed. Lines indicate correlation fit and 95% confidence bands.	70
3.41	Functioning bias of full Model 2 simulations plotted over corresponding t9 values for each session. Lines indicate correlation fit and 95% confidence bands.	71
3.42	Functioning bias of full Model 2 simulations plotted over corresponding t9 values for each session. t9 outliers are removed. Lines indicate correlation fit and 95% confidence bands.	72
3.43	Functioning bias of experimental data plotted over corresponding t9 values estimated by Model 2 for each session. Lines indicate correlation fit and 95% confidence bands.	73
3.44	Functioning bias of experimental data plotted over corresponding t9 values estimated by Model 2 for each session. t9 outliers are removed. Lines indicate correlation fit and 95% confidence bands.	74
3.45	Functioning bias of subject groups defined by median split of fitted k_{exp} based on Model 1 fits (a,b) and Model 2 fits (c,d). The groups' bias is plotted using the full simulation data (a,c) and the corresponding experimental data (b,d).	77
3.46	Functioning bias of full Model 1 simulations plotted over Model 1 k_{exp} fits for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.	79
3.47	Functioning bias of full Model 1 simulations plotted over Model 1 k_{exp} fits without outliers for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.	80
3.48	Functioning bias during experiments plotted over Model 1 k_{exp} fits for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.	81
3.49	Functioning bias during experiments plotted over Model 1 k_{exp} fits without outliers for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.	82
3.50	Functioning bias of full Model 2 simulations plotted over Model 2 k_{exp} fits for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.	83

3.51	Functioning bias of full Model 2 simulations plotted over Model 2 k_{exp} fits without outliers for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.	84
3.52	Functioning bias during experiments plotted over Model 2 k_{exp} fits for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.	85
3.53	Functioning bias during experiments plotted over Model 2 k_{exp} fits without outliers for each session as well as averaged over sessions. Lines indicate correlation fit and 95% confidence bands.	86
3.54	τ_p plotted over k_{exp} for each subject fitted by Model 1 (a) and Model 2 (b). Data are again plotted after removing outliers with respect to τ_p as well as outliers with respect to k_{exp} (c,d). Lines indicate correlation fit and 95% confidence bands.	88

Bibliography

- [1] Q. Wang, J. Bolhuis, C. Rothkopf, T. Kolling, M. Knopf, and J. Triesch. Infants in Control: Rapid Anticipation of Action Outcomes in a Gaze-Contingent Paradigm. *PLoS ONE*, 7, 2012.
- [2] H. Harlow. Learning and satiation of response in intrinsically motivated complex puzzle performance by monkeys. *Journal of Comparative and Physiological Psychology*, 43, 1950.
- [3] H. Harlow, M. Harlow, and D. Meyer. Learning motivated by a manipulation drive. *Journal of Experimental Psychology*, 40, 1950.
- [4] D. Berlyne. A theory of human curiosity. *British Journal of Psychology*, 45, 1954.
- [5] R. White. Motivation reconsidered: The concept of competence. *Psychological Review*, 66, 1959.
- [6] G. Baldassarre and M. Mirolli, editors. *Intrinsically Motivated Learning in Natural and Artificial Systems*. Springer, Berlin, 2013.
- [7] R. Ryan and E. Deci. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, 25, 2000.
- [8] M. Mirolli and G. Baldassarre. *Intrinsically Motivated Learning in Natural and Artificial Systems*, chapter Functions and mechanisms of intrinsic motivations: the knowledge versus competence distinction. Springer-Verlag, Berlin, 2013.
- [9] R. Ryan and E. Deci. Self-determination theory and the facilitation of intrinsic motivation. *American psychologist*, 55, 2000.
- [10] J. Schmidhuber. Formal theory of creativity, fun, and intrinsic motivation (1990-2010). *IEEE Transactions on Autonomous Mental Development*, 2, 2010.
- [11] J. Weng, J. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur, and E. Theelen. Autonomous mental development by robots and animals. *Science*, 291, 2001.
- [12] D. McFarland and T. Bösser. *Intelligent Behavior in Animals and Robots*. MIT, Cambridge, 1993.
- [13] C. Hull. *Principles of Behavior*. D. Appleton-Century, New York, 1943.
- [14] C. Hull. *Essentials of Behavior*. Yale University Press, New Haven, 1951.

- [15] C. Hull. *A Behavior System: An Introduction to Behavior Theory Concerning the Individual Organism*. Yale University Press, New Haven, 1952.
- [16] D. Berlyne. *Conflict, Arousal, Curiosity*. McGraw-Hill, New York, 1960.
- [17] D. Berlyne. *Aesthetics and Psychobiology*. Appleton-Century-Crofts, New York, 1971.
- [18] P.-Y. Oudeyer, F. Kaplan, and V. Hafner. Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, 11, 2007.
- [19] J. Schmidhuber. A possibility for implementing curiosity and boredom in model-building neural controllers. In *From animals to animats: Proceedings of the first international conference on simulation of adaptive behavior*, 1991.
- [20] A. Barto, S. Singh, and N. Chentanez. Intrinsically motivated learning of hierarchical collections of skills. In J. Triesch and T. Jebara, editors, *International Conference on Development and Learning (ICDL) 2004*, 2004.
- [21] S. Singh, A. Barto, and N. Chentanez. Intrinsically motivated reinforcement learning. In L. Saul, Y. Weiss, and L. Bottou, editors, *Advances in Neural Information Processing Systems 17: Proceedings of the 2004 Conference*, 2005.
- [22] A. Barto. *Intrinsically Motivated Learning in Natural and Artificial Systems*, chapter Intrinsic motivation and reinforcement learning. Springer-Verlag, Berlin, 2013.
- [23] J. Schmidhuber. *Anticipatory Behavior in Adaptive Learning Systems. From Psychological Theories to Artificial Cognitive Systems*, chapter Driven by compression progress: A simple principle explains essential aspects of subjective beauty, novelty, surprise, interestingness, attention, curiosity, creativity, art, science, music, jokes. Springer-Verlag, Berlin, 2009.
- [24] E. Dommett, V. Coizet, C. Blaha, J. Martindale, V. Lefebvre, N. Walton, J. Mayhew, P. Overton, and P. Redgrave. How visual stimuli activate dopaminergic neurons at short latency. *Science*, 307, 2005.
- [25] J. Horvitz. Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. *Neuroscience*, 96, 2000.
- [26] J. Lisman and A. Grace. The hippocampal-vta loop: Controlling the entry of information into long-term memory. *Neuron*, 46, 2005.
- [27] N. Otmakova, E. Duzel, A. Deutch, and J. Lisman. *Intrinsically Motivated Learning in Natural and Artificial Systems*, chapter The hippocampal-vta loop: The role of novelty and motivation in controlling the entry of information into long-term memory. Springer-Verlag, Berlin, 2012.

- [28] P. Redgrave and K. Gurney. The short-latency dopamine signal: a role in discovering novel actions? *Nature reviews neuroscience*, 7, 2006.
- [29] K. Gurney, N. Leopatra, A. Shah, A. Koene, and P. Redgrave. *Intrinsically Motivated Learning in Natural and Artificial Systems*, chapter Action discovery and intrinsic motivation: A biologically constrained formalisation. Springer-Verlag, Berlin, 2012.
- [30] P. Redgrave, K. Gurney, T. Stafford, M. Thirkettle, and J. Lewis. *Intrinsically Motivated Learning in Natural and Artificial Systems*, chapter The role of the basal ganglia in discovering novel actions. Springer-Verlag, Berlin, 2012.
- [31] C. Rovee-Collier, M. Sullivan, M. Enright, D. Lucas, and J. Fagen. Reactivation of infant memory. *Science*, 208, 1980.
- [32] I. Kalnins and J. Bruner. The coordination of visual observation and instrumental behavior in early infancy. *Perception*, 2, 1973.
- [33] A. DeCasper and W. Fifer. Of human bonding: Newborns prefer their mothers' voices. *Science*, 208, 1980.
- [34] P. Rochat and T. Striano. Perceived self in infancy. *Infant Behavior and Development*, 23, 2000.
- [35] C. Rovee-Collier, H. Hayne, and M. Colombo. *The development of implicit and explicit memory*. John Benjamins Publishing Company, Amsterdam, 2001.
- [36] B. Kenward. 10-month-olds visually anticipate an outcome contingent on their own action. *Infancy*, 15, 2010.
- [37] G. Bronson. Changes in infants' visual scanning across the 2- to 14-week age period. *Journal of Experimental Child Psychology*, 49, 1990.
- [38] M. Johnson, M. Posner, and M. Rothbart. Components of visual orienting in early infancy: Contingency learning, anticipatory looking, and disengaging. *Journal of Cognitive Neuroscience*, 3, 1991.
- [39] R. Fantz. Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science*, 146, 1964.
- [40] L. Cohen. Attention-getting and attention-holding processes of infant visual preferences. *Child Development*, 43, 1972.
- [41] R. Baillargeon, E. Spelke, and S. Wassermann. Object permanence in five-month-old infants. *Cognition*, 20, 1985.
- [42] M. Haith, C. Hazan, and G. Goodman. Expectation and anticipation of dynamic visual events by 3.5-month-old babies. *Child Development*, 59, 1988.

- [43] P. Quinn, P. Eimas, and S. Rosenkrantz. Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception*, 22, 1993.
- [44] G. Csibra, G. Gergely, S. Bíró, O. Koós, and M. Brockbank. Goal attribution without agency cues: the perception of 'pure reason' in infancy. *Cognition*, 72, 1999.
- [45] W. Smith, S. Johnson, and E. Spelke. Motion and edge sensitivity in perception of object unity. *Cognitive Psychology*, 46, 2003.
- [46] B. McMurray and R. Aslin. Anticipatory eye movements reveal infants' auditory and visual categories. *Infancy*, 6, 2004.
- [47] R. Aslin. What's in a look? *Developmental Science*, 10, 2007.
- [48] S. Johnson, D. Amso, and J. Slemmer. Development of object concepts in infancy: Evidence for early learning in an eye-tracking paradigm. *PNAS*, 100, 2003.
- [49] G. Gredebäck, S. Johnson, and C. von Hofsten. Eye tracking in infancy research. *Developmental Neuropsychology*, 35, 2009.
- [50] L. Oakes. Advances in eye tracking in infancy research. *Infancy*, 17, 2012.
- [51] R. Aslin. Infant eyes: A window on cognitive development. *Infancy*, 17, 2012.
- [52] S. Reader. On-line monitoring of eye position signals in contingent and noncontingent paradigms. *Behaviour Research Methods & Instrumentation*, 5, 1973.
- [53] A. Duchowski. A breadth-first survey of eye-tracking applications. *Behavior Research Methods*, 34, 2002.
- [54] K. Holmboe, R. Fearon, G. Csibra, L. Tucker, and M. Johnson. Freeze-frame: A new infant inhibition task and its relation to frontal cortex tasks during infancy and early childhood. *Journal of Experimental Child Psychology*, 100, 2008.
- [55] F. Deligianna, A. Senju, G. Gergely, and G. Csibra. Automated gaze-contingent objects elicit orientation following in 8-month-old infants. *Developmental Psychology*, 47, 2011.
- [56] S. Wass, K. Porayska-Pomsta, and M. Johnson. Training attentional control in infancy. *Current Biology*, 18, 2011.
- [57] K. Tummeltshammer, D. Mareschal, and N. Kirkham. Infants' selective attention to reliable visual cues in the presence of salient distractors. *Child Development*, 85, 2014.
- [58] M. Miyazaki, H. Takahashi, M. Rolf, H. Okada, and T. Omori. The image-scratch paradigm: a new paradigm for evaluating infants' motivated gaze control. *Scientific reports*, 4, 2014.

- [59] R. Bolado-Gomez and K. Gurney. A biologically plausible embodied model of action discovery. *frontiers in Neurorobotics*, 7, 2013.
- [60] N. Hansen. *Towards a new evolutionary computation*, chapter The CMA evolution strategy: a comparing review. Springer Berlin Heidelberg, 2006.
- [61] R. Sutton and A. Barto. *Reinforcement Learning*. The MIT Press, 1998.
- [62] R. Munroe. The Difference. <https://xkcd.com/242/>, March 2007.
- [63] E. Deci and R. Ryan. *Intrinsic Motivation and Self-Determination in Human Behavior*. Plenum, New York, 1985.
- [64] C. Cofer and M. Appley. *Motivation: Theory and Research*. Wiley, New York, 1964.
- [65] W. Cannon. *The Wisdom of the Body*. W.W. Norton, New York, 1932.
- [66] D. Berlyne. Curiosity and exploration. *Science*, 143, 1966.
- [67] G. Baldassarre and M. Mirolli. *Intrinsically Motivated Learning in Natural and Artificial Systems*, chapter Intrinsically motivated learning systems: an overview. Springer-Verlag, Berlin, 2013.
- [68] S. Marsland, U. Nehmzow, and J. Shapiro. On-line novelty detection for autonomous mobile robots. *Robotics and Autonomous Systems*, 51, 2005.
- [69] U. Nehmzow, Y. Gatsoulis, E. Kerr, J. Condell, N. Siddique, and M. McGinnity. *Intrinsically Motivated Learning in Natural and Artificial Systems*, chapter Novelty detection as an intrinsic motivation for cumulative learning robots. Springer-Verlag, Berlin, 2013.
- [70] M. Schembri, M. Mirolli, and G. Baldassarre. Evolving internal reinforcers for an intrinsically motivated reinforcement-learning robot. In Y. Demiris, D. Mareschal, B. Scassellati, and J. Weng, editors, *Proceedings of the 6th International Conference on Development and Learning*, 2007.
- [71] V. Santucci, G. Baldassarre, and M. Mirolli. Which is the best intrinsic motivation signal for learning multiple skills? *Frontiers in Neurorobotics*, 7, 2013.
- [72] A. Baranes and P. Oudeyer. Active learning of inverse models with intrinsically motivated goal exploration in robots. *Robotics and Autonomous Systems*, 61, 2013.
- [73] J. Mink. The basal ganglia: Focused selection and inhibition of competing motor programs. *Progress in Neurobiology*, 50, 1996.
- [74] P. Redgrave. Basal ganglia. *Scholarpedia*, 2, 2007.
- [75] A. Barto. *Models of Information Processing in the Basal Ganglia*, chapter Adaptive critics and the basal ganglia. MIT, Cambridge, 1995.

- [76] K. Doya. Complementary roles of basal ganglia and cerebellum in learning and motor control. *Current Opinions in Neurobiology*, 10, 2000.
- [77] J. Houk, J. Adams, and A. Barto. *Models of Information Processing in the Basal Ganglia*, chapter A model of how the basal ganglia generates and uses neural signals that predict reinforcement. MIT, Cambridge, 1995.
- [78] D. Joel, Y. Niv, and E. Ruppin. Actor-critic models of the basal ganglia: New anatomical and computational perspectives. *Neural Networks*, 15, 2002.
- [79] P. Redgrave, N. Vautrelle, and J. Reynolds. Functional properties of the basal ganglia's reentrant loop architecture: selection and reinforcement. *Neuroscience*, 198, 2011.
- [80] W. Schultz, P. Dayan, and P. Montague. A neural substrate of prediction and reward. *Science*, 275, 1997.
- [81] W. Schultz and A. Dickinson. Neuronal coding of prediction errors. *Annual Review of Neuroscience*, 23, 2000.
- [82] G. Baldassarre, T. Stafford, M. Mirolli, P. Redgrave, R. Ryan, and A. Barto. Intrinsic motivations and open-ended development in animals, humans, and robots: an overview. *frontiers in Psychology*, 5, 2014.
- [83] V. Bush. Science: The endless frontier. *Transactions of the Kansas Academy of Science (1903-)*, 48, 1945.
- [84] E. Siqueland and C. De Lucia. Visual reinforcement of nonnutritive sucking in human infants. *Science*, 12, 1969.
- [85] P. Rochat and T. Striano. Emerging self-exploration by two-month-old infants. *Science*, 2, 1999.
- [86] L. Bahrick and J. Watson. Detection of intermodal proprioceptive-visual contingency as a potential basis of self-perception in infancy. *Developmental Psychology*, 21, 1985.
- [87] D. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. Chichester, John Wiley & Sons, 1992.
- [88] T. Kolling, J. Hohmann, M. Murakami, J. Triesch, and M. Knopf. Gaze-contingent learning and retention in 8-month-old infants. submitted.
- [89] A. Gancarz, M. San George, L. Ashrafioun, and J. Richards. Locomotor activity in a novel environment predicts both responding for a visual stimulus and self-administration of a low dose of methamphetamine in rats. *Behavioral Processes*, 86, 2011.

- [90] R. Bolado-Gomez. *The basal ganglia and the dopaminergic system: a computational study of the ‘sensory prediction error’ interpretation*. PhD thesis, University of Sheffield, 2014.
- [91] K. Gurney, M. Humphries, and P. Redgrave. A new framework for cortico-striatal plasticity: behavioural theory meets in vitro data at the reinforcement-action interfacey. *PLoS Biology*, 13, 2015.
- [92] J. Schmidhuber. Curious model-building control systems. In *Proceedings of the International Joint Conference on Neural Networks*, 1991.
- [93] C. Moulin-Frier and P.-Y. Oudeyer. Exploration strategies in developmental robotics: a unified probabilistic framework. In *IEEE Third Joint International Conference on Developmental and Learning and Epigenetic Robotics (ICDL)*, 2013.
- [94] P. Redgrave, K. Gurney, and J. Reynolds. What is reinforced by phasic dopamine signals? *Brain research reviews*, 58, 2008.
- [95] E. Kandel, J. Schwartz, T. Jessell, S. Siegelbaum, and A. Hudspeth, editors. *Principles of Neural Science*, chapter 43. McGraw-Hill, 2013.
- [96] W. Shen, M. Flajolet, P. Greengard, and D. Surmeier. Dichotomous dopaminergic control of striatal synaptic plasticity. *Science*, 321, 2008.
- [97] K. Gurney, M. Humphries, and P. Redgrave. Cortico-striatal plasticity for action-outcome learning using spike timing dependent eligibility. *BMC Neuroscience*, 10, 2009.
- [98] N. Hansen. The CMA Evolution Strategy: A Tutorial, 2011.
- [99] S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 1979.
- [100] J. Mink and W. Thach. Basal ganglia intrinsic circuits and their role in behavior. *Current Opinions in Neurobiology*, 3, 1993.
- [101] K. Doya. What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Networks*, 12, 1999.
- [102] P. Redgrave, T. Prescott, and K. Gurney. The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience*, 89, 1999.
- [103] J. Houk, C. Bastianen, D. Fansler, A. Fishbach, D. Fraser, P. Reber, S. Roy, and L. Simo. Action selection and refinement in subcortical loops through basal ganglia and cerebellum. *Philosophical Transactions of the Royal Society B Biological Sciences*, 362, 2007.
- [104] K. Gurney, N. Lepora, A. Shah, A. Koene, and P. Redgrave. *Action Discovery and Intrinsic Motivation: A Biologically Constrained Formalisation*. Berlin: Springer Verlag, 2013.

- [105] R. Marraffa, V. Sperati, D. Caligiore, J. Triesch, and G. Baldassarre. A bio-inspired attention model of anticipation in gaze-contingency experiments with infants. In *International Conference for Development and Learning and Epigenetic Robotics, ICDL-Epirob 2012*.
- [106] N. Bruce and J. Tsotsos. Attention based on information maximization. *Journal of Vision*, 7, 2007.



Max Murakami

Personal Information

Address Herborner Straße 62
60439 Frankfurt am Main

Telephone 01575 4868775

E-mail murakami@fias.uni-frankfurt.de

Marital married

Status

Nationality German

Birthday/place 14.5.1988 in Tokyo

Education

PhD

since 2014 **Physics**, *Goethe University*, Frankfurt am Main.

- Research area: Computational Neuroscience
- „A Neuronal Model of Gaze-Contingent Learning“
- Supervisor: Prof. Dr. Jochen Triesch

Studies

2012-2014 **Physics M.Sc.**, *Goethe University*, Frankfurt am Main.

- Minor: Japanese Studies
- Specialization: Computational Neuroscience
- Master thesis *Listen and Babble: A Model of Vowel Acquisition Based on Imitation Learning* with Prof. Dr. Jochen Triesch, grade 1,0
- Final grade: 1,0 with distinction

2008-2012 **Physics B.Sc.**, *Goethe University*, Frankfurt am Main.

- Minor: Japanese Studies
- Bachelor thesis *Effects of Transcranial Magnetic Stimulation on different-sized neurons* with Prof. Dr. Jochen Triesch, grade 1,0
- Final grade: 1,6

School Education

1998-2007 **Abitur**, *Reichsstadt-Gymnasium*, Rothenburg ob der Tauber.

- Advanced subjects: Mathematics and Physics
- Graduation thesis: *Principles of Data Compression*
- Final grade: 1,7

Job Experience

- since 2014 **Frankfurt Institute for Advanced Studies**, *Scientific Assistant*, Frankfurt am Main.
Research for DFG project *Gaze-Contingent Learning in Infants* and for EU project *Goal-based Open-ended Autonomous Learning Robots*.
- 2013-2014 **Frankfurt Institute for Advanced Studies**, *Webdesigner*, Frankfurt am Main.
Website maintenance of the Triesch lab.
- 2008 **C.F. Maier**, *Internship*, Schillingsfürst.
Quality assurance.
- 2007-2008 **Verbundklinikum Landkreis Ansbach**, *Community service*, Rothenburg ob der Tauber.
Nursing and transporting surgical patients.
- 2005-2008 **TSV 2000 Rothenburg**, *Trainer*, Rothenburg ob der Tauber.
Badminton training for high school students and adults.

Teaching Experience

- seit 2016 **seminar: Principles of Neural Science**,
Frankfurt Institute for Advanced Studies, Frankfurt am Main.
- 2014-2015 **tutorial: Theory of Special Relativity and Quantum Mechanics**,
Institute for Theoretical Physics, Goethe University, Frankfurt am Main.
- 2014 **tutorial: Electrodynamics**,
Institute for Theoretical Physics, Goethe University, Frankfurt am Main.
- 2012-2013 **tutorial: Mathematical Methods in Physics**,
Institute for Theoretical Physics, Goethe University, Frankfurt am Main.
- 2011 **tutorial: Classical Mechanics**,
Institute for Theoretical Physics, Goethe University, Frankfurt am Main.
- 2010-2011 **lab Classes**,
Institute for Applied Physics, Goethe University, Frankfurt am Main.

Publications

- 2015 **Seeing [u] aids vocal learning: babbling and imitation of vowels using a 3D vocal tract model, reinforcement learning, and reservoir computing**,
M. Murakami, B. Kröger, P. Birkholz, J. Triesch, IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL), 2015.
Conference paper
- 2014 **A Model of TMS-induced I-waves in Motor Cortex**,
C. Rusu, M. Murakami, U. Ziemann, J. Triesch, Brain Stimulation 7(3), 2014.

Conference Contributions

- 2016 **Look and Learn: A Computational Model of Gaze-Contingent Learning,**
M. Murakami, J. Bolhuis, T. Kolling, M. Knopf, J. Triesch, IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL), 2016.
Poster
- 2015 **Seeing [u] aids vocal learning: babbling and imitation of vowels using a 3D vocal tract model, reinforcement learning, and reservoir computing,**
M. Murakami, B. Kröger, P. Birkholz, J. Triesch, IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL), 2015.
Talk

IT and Programming

PYTHON, C/C++, JAVA, Matlab, L^AT_EX, HTML, NEURON, Maple, Microsoft Office, Linux, Microsoft Windows...

Languages

German	native speaker
English	business fluent
Japanese	advanced (JLPT 3 kyū)
French	advanced
Korean	basic

Awards

- 2007 Von Staudt Award for best mathematics A level.