

Aus der Klinik für Psychiatrie und Psychotherapie
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

**Messung verschiedener Formen der Verhaltenskontrolle
und deren Relevanz bei Alkoholabhängigkeit**

zur Erlangung des akademischen Grades
Doctor medicinae (Dr. med.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Jean Wendt

aus Berlin

Datum der Promotion: 04.06.2021

„Es ist ganz wahr, was die Philosophie sagt, dass das Leben rückwärts verstanden werden muss. Aber darüber vergisst man den andern Satz, dass vorwärts gelebt werden muss.“

(Kierkegaard, 1923)

Vorwort

Teilergebnisse der hier veröffentlichten Arbeit wurden bereits in folgendem Zeitschriftenartikel veröffentlicht:

Friedel, E., Koch, S.P., Wendt, J., Heinz, A., Deserno, L., Schlagenhaut, F., 2014. Devaluation and sequential decisions: linking goal-directed and model-based behavior. *Front Hum Neurosci* 8, 587. <https://doi.org/10.3389/fnhum.2014.00587>

Dies betrifft die Abschnitte 3.3., 3.4 und 3.5. Dementsprechend sind auch einige Abbildungen dieser Arbeit an jene des genannten Zeitschriftenartikels angelehnt und entsprechend gekennzeichnet. Darüber hinaus wurden in den genannten Teilabschnitten bereits veröffentlichte Daten um weitere Berechnungen ergänzt und die entsprechenden Ergebnisse aufgeführt (u.a. nicht devaluierte und neutrale Bedingung beim Devaluationsparadigma, Haupteffekt Transition, Korrelation Devaluations sensitivität und p-Belohnung, Korrelation Gewinn und p-Interaktion). Trotz der bereits vorhandenen Daten und Ergebnisse im Rahmen der genannten Publikation, wurde die statistische Auswertung für alle erhobenen Daten erneut meinerseits durchgeführt. Aufgrund dieser Berechnungen, welche auch zur einheitlichen Anpassung von Nachkommastellen durchgeführt wurden, waren einige Ergebnisse geringfügig abweichend von den bereits veröffentlichten Daten. Davon betroffen sind insbesondere die Ergebnisse in Bezug auf die Korrelationsanalysen.

Inhaltsverzeichnis

Vorwort	III
Inhaltsverzeichnis	IV
Tabellen- und Abbildungsverzeichnis	VII
Abkürzungsverzeichnis	VIII
Zusammenfassung	IX
Abstract	XI
1. Einleitung	1
1.1. Theoretischer Hintergrund	4
1.1.1. Klassische Konditionierung	5
1.1.2. Operante Konditionierung und instrumentelles Lernen	5
1.1.3. Extinktionslernen	7
1.2. Habituelles und zielgerichtetes Verhalten	7
1.2.1. Operationalisierung habitueller und zielgerichteter Verhaltenskontrolle	10
1.2.2. Neurobiologische Korrelate habituellen und zielgerichteten Verhaltens	12
1.3. Modellbasiertes und modellfreies Verhalten	14
1.3.1. Neurobiologische Korrelate modellbasierten und modellfreien Verhaltens	17
1.4. Die Rolle veränderter Lern- und Verhaltensmechanismen bei der Entstehung und Aufrechterhaltung der Alkoholabhängigkeit	19
1.5. Hypothesenbildung	25
2. Methoden	27
2.1. Datenerhebung	27
2.1.1. Pilotstichprobe für die selektive Devaluationsaufgabe	28
2.1.2. Probandenstichprobe für Hauptstudie	28
2.2. allgemeine Versuchsanordnung der Hauptstudie	28
2.3. Devaluationsparadigma	30

2.3.1.	instrumentelle Lernaufgabe	31
2.3.2.	Devaluation	34
2.3.3.	Extinktionstest	35
2.3.4.	Pilotierung des selektiven Devaluationsparadigmas	36
2.4.	<i>two-step</i> -Aufgabe	37
2.5.	Statistische Analysen	40
2.5.1.	Analyse der Verhaltensdaten und Ratings des Devaluationsparadigmas	40
2.5.2.	Analyse der Verhaltensdaten der <i>two-step</i> -Aufgabe	41
2.5.3.	Korrelationsanalyse zwischen den zwei Paradigmen	42
2.5.4.	Korrelationsanalysen zwischen den Ergebnissen der neuropsychologischen Testung und den Interaktionstermen der Paradigmen	44
3.	Ergebnisse	45
3.1.	Pilotstudie zur Etablierung des selektiven Devaluationsparadigmas	45
3.2.	Bewertung der Pilotdaten und daraus abgeleitete Modifikationen des Devaluationsparadigmas	47
3.3.	Devaluationsparadigma	48
3.3.1.	Trainingsphase/ instrumentelle Lernaufgabe	48
3.3.2.	Devaluation	48
3.3.3.	Extinktionsphase	49
3.4.	<i>two-step</i> -Aufgabe	50
3.5.	Konstruktvalidität: Korrelation zwischen den beiden Paradigmen	51
3.6.	Zusammenhang zwischen Entscheidungsverhalten und kognitiven Fähigkeiten	52
4.	Diskussion	54
4.1.	Interpretation der Ergebnisse mit Hinblick auf die Konstruktvalidität	54
4.2.	Limitationen der Studie	64

4.3.	Bedeutung der Ergebnisse für die Untersuchung veränderter Lern- und Verhaltensmechanismen bei Alkoholabhängigkeit	70
5.	Literaturverzeichnis	75
	Eidesstattliche Versicherung	87
	Anteilerklärung an etwaigen erfolgten Publikationen	89
	Lebenslauf	91
	Publikationsliste	93
	Danksagung	94

Tabellen- und Abbildungsverzeichnis

Abbildung 1: allgemeiner Aufbau eines Devaluationsparadigmas	11
Abbildung 2: allgemeine Versuchsanordnung der Hauptstudie	29
Abbildung 3: Devaluationsparadigma in Anlehnung an Valentin et al. (2007) und Friedel et al. (2014)	32
Abbildung 4: <i>two-step</i> -Aufgabe in Anlehnung an Daw et al. (2011) und Friedel et al. (2014)	38
Abbildung 5: Beispiele für Ergebnisdiagramme bei der <i>two-step</i> -Aufgabe	42
Abbildung 6: Auswahlverhalten für den hoch probabilistischen Stimulus	45
Abbildung 7: Rating für Hunger und Anreizwert beim Devaluationsparadigma in der Pilotstudie	46
Abbildung 8: Rating für Hunger und Anreizwert beim Devaluationsparadigma in der Hauptstudie (N = 14) in Anlehnung an Friedel et al. (2014)	49
Abbildung 9: Analyse der Verhaltensdaten in Anlehnung an Friedel et al. (2014)	50
Abbildung 10: Korrelation der Parameter beider Paradigmen (N = 12)	52
Tabelle 1: Ergebnis der neuropsychologischen Testung (N = 18)	52
Tabelle 2: Korrelationsanalysen zwischen Ergebnissen der neuropsychologischen Testung und den aus den Paradigmen ermittelten Interaktionstermen	53

Abkürzungsverzeichnis

BMI	Body-Maß-Index
CR	konditionierte Reaktion
CS	konditionierter Stimulus
dIPFC	dorsolateraler präfrontaler Kortex
F	F-Wert
fMRT	funktionelle Magnetresonanztomographie
g	Gramm
min	Minuten
ml	Milliliter
mm	Millimeter
mOFC	medialer orbitofrontaler Kortex
N	Fallzahl/ Stichprobengröße
p	Irrtumswahrscheinlichkeit, Signifikanzwert
r	Rangkorrelationskoeffizient
RPE	<i>reward prediction error</i> (Belohnungsvorhersagefehler)
s	Sekunden
SD	Standardabweichung
SEM	Standardfehler
SPE	<i>state prediction error</i> (Prädiktionsfehler)
T	T-Wert
UCR	unkonditionierte Reaktion
US	unkonditionierter Stimulus
vmPFC	ventromedialer präfrontaler Kortex

Zusammenfassung

Einleitung: Das Verständnis davon, wie Lern- und Verhaltensmechanismen sowie deren Anpassung an sich ändernde Umweltbedingungen bei psychischen Erkrankungen modifiziert sind und auf welche Weise sie so zur Entstehung und Aufrechterhaltung jener Erkrankungen beitragen, gewinnt in der Forschung zunehmend an Interesse. Diesbezüglich werden zwei dichotome Systeme unterschieden, wobei sich zwei unterschiedliche Konzepte mit jeweils eigenen Begrifflichkeiten und Operationalisierungen entwickelt haben: habituelles vs. zielgerichtetes und modellfreies vs. modellbasiertes Verhalten. In der Forschung werden beide Konzepte mit ihrer jeweiligen Terminologie oft synonym verwendet, ohne dass bisher die Konstruktvalidität der Operationalisierungen dieser Konzepte durch entsprechende Paradigmen überprüft wurde. Das heißt, bislang erfolgten keine Untersuchungen mit der Fragestellung, inwieweit diese Operationalisierungen möglicherweise dasselbe theoretische Konstrukt abbilden und ob demnach zielgerichtetes mit modellbasiertem und habituelles mit modellfreiem Verhalten gleichzusetzen ist. Ziel dieser Studie war es daher, erstmals die Konstruktvalidität zu überprüfen und mögliche veränderte Mechanismen bei Alkoholabhängigkeit herauszuarbeiten.

Methodik: Zunächst erfolgte eine Pilotstudie mit 21 Probanden. Dabei stellte sich heraus, dass die Rekrutierung abstinenten, alkoholabhängiger Patienten deutlich erschwert war. So wurden letztlich 18 gesunde Probanden in die Hauptstudie eingeschlossen. Diese absolvierten sowohl ein Devaluationsparadigma, als auch eine sequentielle Entscheidungsaufgabe. Für beide Aufgaben wurden jeweils Parameter ermittelt (Devaluationssensitivität, p-Interaktion, p-Belohnung), die das individuelle Ausmaß der Verhaltensstrategien eines jeden Probanden abbildeten. Mittels Korrelationsanalyse wurde getestet, ob eine positive Korrelation zwischen diesen Parametern besteht, was wiederum für das Vorliegen einer Konstruktvalidität sprechen würde.

Ergebnisse: Die Korrelationsanalysen ergaben einen positiven Zusammenhang zwischen dem individuellen Ausmaß von Devaluationssensitivität (Interaktionsterm Devaluationsparadigma) und modellbasiertem Verhalten (p-Interaktion) in der sequentiellen Entscheidungsaufgabe (Spearman Rangkorrelation $r = 0.698$, $p = 0.008$). Interessanterweise zeigte sich keine signifikante Korrelation ($p > 0,6$)

zwischen dem Ausmaß an modellfreiem Verhalten (Haupteffekt Belohnung bei sequentieller Entscheidungsaufgabe, p-Belohnung) und der Devaluationssensitivität.

Schlussfolgerung: Es konnte gezeigt werden, dass zielgerichtetes und modellbasiertes Verhalten miteinander assoziiert sind und folglich auf ähnlichen Mechanismen bzw. einem gemeinsamen Konstrukt zu beruhen scheinen. Zwischen dem Ausmaß an modellfreiem Verhalten und der Devaluationssensitivität konnte hingegen kein Zusammenhang gezeigt werden. Die vorliegende Arbeit war die erste Studie, die die Konstruktvalidität der beiden Konzepte überprüft hat. Folgestudien konnten die Ergebnisse an größeren Stichproben replizieren. Die Konstruktvalidität ist nach den bisherigen Erkenntnissen nicht vollständig bestätigt. Mit Hinblick auf die Limitationen dieser Arbeit, sollten weiteren Studie folgen, wobei die Paradigmen derart angepasst werden müssten, dass sie im experimentellen Design für klinische Populationen geeignet sind und habituelles bzw. modellfreies Verhalten besser erfassen. Zudem sollte eine möglichst große und heterogene Probandenstichprobe herangezogen werden.

Abstract

Introduction: The scientific understanding of learning and decision making and the adaption of behavior to altered conditions are gaining increasing interest – especially the influence of altered behavior mechanisms on the development and maintenance of psychiatric diseases. By now, there is a broad agreement that two dichotomous systems of possible decision-making strategies exist. Referring to this, there has been developed two different concepts with proper terms, definitions and paradigms: habitual vs. goal-directed behavior and model-free vs. model-based decision-making. In research, these different concepts have been so far treated almost equivalently, without even knowing whether these different measurements reflect the same theoretical construct. In the present study, we therefore addressed the question of construct validity and altered behavior mechanisms involved in alcohol dependence.

Methods: First, pilot study was undertaken. Since the recruitment of alcohol dependent patients turned out to be difficult, we included only healthy participants in the main study (N = 18). Here, the participants performed two paradigms: a devaluation paradigm and a sequential decision-making task. To assess the individual degree of the different decision-making strategies, interaction score from both paradigms were computed for each participant. To verify the construct validity, a correlation analysis was used to assess an assumed positive correlation between these interaction scores.

Results: Our analysis showed a significant correlation between the interaction score derived from the devaluation task (devaluation sensitivity) with the interaction score (p-interaction, degree model-based behavior) from the sequential decision-making task (Spearman's $\rho = 0.698$, $p = 0.008$). Interestingly, the degree of model-free behavior in the sequential decision-making task (main effect of reward, p-reward) did not correlate with the degree of devaluation sensitivity ($p > 0,6$).

Conclusion: We were able to prove that goal-directed and model-based strategies are associated with each other. Hence, these two strategies of decision-making might rely on similar mechanisms suggesting a single framework underlying both paradigms. Such an association was not found between the degree of model-free behavior and devaluation sensitivity. Therefore, construct validity is not completely verified, yet. The present study was the first one assessing the construct validity of the two different paradigms. These results are in line with later studies. With regard to

the limitation of this work, further studies should develop and use paradigms which are able to determine habitual and model-free behavior as well. The experimental design should also be more suitable for patients. Furthermore, a larger and more heterogeneous sample of subjects is recommended.

1. Einleitung

Lern- und Entscheidungsfindungsprozesse sind ein wesentlicher Bestandteil unseres alltäglichen (Über-) Lebens. Die Erforschung dieser Prozesse gewann daher im Laufe des letzten Jahrhunderts zunehmend an Bedeutung. Im Rahmen dieser Forschungsbestrebungen wurden verschiedene Theorien und Konzepte entwickelt, mit dem Ziel, ein näheres Verständnis von den zugrundeliegenden Mechanismen zu erhalten. Mit der Entwicklung bildgebender Verfahren, insbesondere der funktionellen Magnetresonanztomographie, wurde es zudem möglich, auch die neuronalen Korrelate dieser Prozesse zu identifizieren. Diese Untersuchungen gaben und geben Aufschluss darüber, wie die involvierten Strukturen und neuronalen Netzwerke innerhalb des Entscheidungsprozesses Informationen und Feedback verarbeiten und nutzen. Hierbei werden zwei parallele Systeme bzw. Strategien der Verhaltenskontrolle beim belohnungsabhängigen Lernen und bei Entscheidungsfindungsprozessen unterschieden: Einem planenden, prospektiven System steht hierbei ein retrospektiv orientiertes Kontrollsystem gegenüber.

Diesbezüglich haben sich historisch gesehen zwei ähnliche Konzepte mit jeweils eigenen Begrifflichkeiten entwickelt, die aus zwei unterschiedlich gewachsenen Bereichen entspringen. Dabei wird im Bereich der theoretischen Psychologie zwischen einem zielgerichteten System und einem gewohnheitsmäßigen, habituellen System der Verhaltenskontrolle unterschieden. Ein ähnliches Konstrukt stammt aus der computationalen Neurowissenschaft, also aus einem Bereich, der funktionelle Bestandteile des Nervensystems mit Hilfe mathematischer Modelle beschreibt, wobei hier zwischen modellbasiertem und modellfreiem Verhalten unterschieden wird.

In der Literatur werden die unterschiedlichen Begrifflichkeiten mitunter äquivalent verwendet. So fassen Dolan und Dayan in ihrer Übersichtsarbeit zielgerichtetes und modellbasiertes Verhalten unter dem Begriff des „reflektierten“ Lern- und Entscheidungsverhaltens zusammen (Dolan and Dayan, 2013). Reflektiert bedeutet in diesem Zusammenhang, dass das Verhalten an zukünftigen Handlungsoptionen und ihrer Konsequenzen orientiert ist, wobei diese Optionen mit Blick auf das gewünschte Ziel gegeneinander abgewogen und bewertet werden. Diese Strategie der Verhaltenskontrolle ist demnach prospektiv und beruht gewissermaßen auf einer vorausschauenden Abwägung. Im Gegensatz dazu basiert eine reflexive Verhaltensstrategie (also habituelles bzw. modellfreies Verhalten) auf zurückliegenden Lernerfahrungen, nutzt folglich erlebte positive und negative

Handlungskonsequenzen für den gegenwärtigen Entscheidungsprozess. Van der Meer et al. (2012) hingegen verwenden die Begriffe eines habituellen und deliberativen Systems der Verhaltenskontrolle zur Beschreibung der beiden Kontrollmechanismen von Verhalten: Dabei meint deliberativ, dass Entscheidungen „wohlüberlegt“ sind und auf dem Wissen über den Zusammenhang zwischen dem möglichen Verhalten und dessen Konsequenzen beruhen.

Es mag der Eindruck entstehen, dass es sich bei den genannten Konzepten (habituelle vs. zielgerichtete bzw. modellfreie vs. modellbasierte Strategien der Verhaltenskontrolle) in Bezug auf verschiedene Verhaltensmechanismen, unabhängig von der verwendeten Terminologie, um dichotome Konstrukte handeln würde. Die Grundidee dieser Dichotomie geht darauf zurück, dass für jede Entscheidungsoption zwei unabhängige zugrundeliegende Systeme der Verhaltenskontrolle existieren, die den Entscheidungsprozess dominieren und so einen Einfluss auf das gezeigte Verhalten haben. Tatsächlich wird aber vor allem in den computationalen Neurowissenschaften diskutiert, ob diese gegensätzlichen Systeme letztlich nur Pole eines kontinuierlichen Spektrums von Lern- und Entscheidungsverhalten darstellen (Dolan and Dayan, 2013). Dabei wird davon ausgegangen, dass die Fähigkeit, zwischen dem retrospektivem (habituell bzw. modellfrei) und prospektivem System (zielgerichtet bzw. modellbasiert) der Verhaltenskontrolle zu vermitteln und gewissermaßen eine Balance zu schaffen, eine notwendige Voraussetzung für eine erfolgreiche Verhaltensstrategie darstellt. Ist diese Fähigkeit eingeschränkt und kommt es zu einer Imbalance zwischen diesen Systemen, so resultiert daraus möglicherweise inflexibles, repetitives Verhalten, wie es auch bei verschiedenen psychiatrischen Erkrankungen beobachtet werden kann (Voon et al., 2017). Die Befunde aus der Untersuchung von Lern- und Entscheidungsmechanismen und den damit einhergehenden individuellen Strategien der Verhaltenskontrolle können folglich auch neue Erkenntnisse über diejenigen Prozesse liefern, die zur Entstehung und Aufrechterhaltung psychischer Erkrankungen beitragen und auf diese Weise auch die Entwicklung neuer therapeutischer Ansätze vorantreiben (Eberl et al., 2013).

In der experimentellen Psychologie wurden zur Operationalisierung der genannten Konstrukte unterschiedliche Paradigmen etabliert. In Hinblick auf zielgerichtetes vs. habituelles Verhalten werden vorrangig Devaluationsparadigmen verwendet (Tricomi et al., 2009; Valentin et al., 2007). Zur Erforschung modellbasierten und modellfreien

Verhaltens kommen dagegen im Allgemeinen sequentielle Entscheidungsaufgaben zum Einsatz, welche eine Quantifizierung dieser beiden Anteile ermöglichen (Daw et al., 2011; Gläscher et al., 2010). Es gibt allerdings eine anhaltende Diskussion darüber, ob diese unterschiedlichen Paradigmen letztlich die gleichen Aspekte instrumentellen Lernverhaltens abbilden (Dolan and Dayan, 2013; Doll et al., 2012). Es wird davon ausgegangen, dass die Paradigmen letztlich die gleichen theoretischen Konstrukte überprüfen und somit zwei ähnliche Mechanismen lediglich aus unterschiedlicher Perspektive beleuchten, die aus zwei verschiedenen Bereichen entstanden sind (nämlich der theoretischen Psychologie und den computationalen Neurowissenschaften).

Allerdings wurde bisher nicht überprüft, inwieweit zwischen den beiden Operationalisierungen ein Zusammenhang besteht und ob zielgerichtetes mit modellfreiem und habituelles mit modellfreiem Verhalten gleichzusetzen ist. In Anbetracht dessen, dass in den Neurowissenschaften die Konstrukte und deren Operationalisierungen beinahe äquivalent gebraucht werden und folglich auch entsprechende Rückschlüsse in Hinblick auf mutmaßlich identische Prozesse gezogen werden, ist es allerdings weiterhin von Bedeutung, die Konstruktvalidität der Paradigmen zu überprüfen. Um dies zu realisieren, wurden in der vorliegenden Arbeit zwei etablierte Paradigmen verwendet: eine selektive Devaluationsaufgabe (Schwabe and Wolf, 2009; Valentin et al., 2007) und eine sequentielle Entscheidungsaufgabe (Daw et al., 2011). Im Rahmen eines *within-subject*-Designs wurde das Ausmaß von zielgerichtetem und habituellem sowie von modellbasiertem und modellfreiem Verhalten für jeden Probanden erfasst. Ziel dieser Arbeit ist es, hierbei zu überprüfen, ob eine positive Korrelation zwischen den verschiedenen Parametern besteht. Dabei soll die Untersuchung sowohl an einer Stichprobe von gesunden Probanden als auch an einer klinischen Population stattfinden. Denn bisherige Studien lassen darauf schließen, dass Lern- und Verhaltensmechanismen bei psychischen Erkrankungen verändert sind (Heinz, 2017; Lee, 2013; Voon et al., 2017). Daher sollten auch alkoholabhängige Patienten in die Studie eingeschlossen werden. Auf diese Weise bot sich die Möglichkeit, mit Blick auf eine klinische Population, potentielle Schwierigkeiten bei der Anwendbarkeit der Paradigmen zu identifizieren, eventuelle Änderungen der Verhaltenskontrolle im Rahmen der Alkoholabhängigkeit zu erfassen und die Annahme zu überprüfen, dass auch bei veränderten Mechanismen die Konstruktvalidität gegeben ist.

Im Folgenden werden zunächst die Grundlagen der (behavioristischen) Lerntheorie dargestellt, auf denen die beiden oben genannten Konzepte beruhen. Im Anschluss sollen diese Konzepte mit ihren jeweiligen Strategien der Verhaltenskontrolle eingeführt werden. Dabei werden sowohl die jeweiligen Eigenschaften und Funktionsweisen dieser Strategien sowie deren Vor- bzw. Nachteile erläutert. Des Weiteren werden die entsprechenden Operationalisierungen und Paradigmen der beiden Konzepte eingeführt, wie sie auch zur Überprüfung der Konstruktvalidität im Rahmen dieser Arbeit verwendet werden. Auch wenn es sich hierbei um eine rein verhaltensexperimentelle Studie handelt, so soll auch auf die neurobiologischen Korrelate der jeweiligen Strategien der Verhaltenskontrolle eingegangen werden. Hintergrund hierfür ist, dass die funktionelle Bildgebung und die damit einhergehende Identifikation der den beiden Konzepten zugrundeliegenden neuronalen Mechanismen eine weitere Chance bietet, alternative und differenziertere Erklärungsansätze für das Zusammenspiel der entsprechenden Systeme der Verhaltenskontrolle zu entwickeln. Darüber hinaus kann auf diese Weise überprüft werden, ob die (bisher angenommene) Konstruktvalidität der Paradigmen der beiden Konzepte auch auf neuronaler Ebene auf ähnlichen Mechanismen beruhen. Die Motivation für die vorliegende Arbeit stammt insbesondere aus der Suchtforschung, da in diesem Bereich die Untersuchung grundlegender Lernmechanismen und derer Operationalisierungen in den letzten Jahren zu einem besseren Verständnis der Entstehung und Aufrechterhaltung von Suchterkrankungen beigetragen hat. Daher wird abschließend auch eine Übersicht über die bisherigen Forschungserkenntnisse zu veränderten (dysfunktionalen) Lern- und Verhaltensmechanismen bei psychischen Erkrankungen gegeben – mit dem Schwerpunkt auf Alkoholabhängigkeit.

1.1. Theoretischer Hintergrund

Die Fähigkeit, die Konsequenzen unseres eigenen Verhaltens zu erkennen und vorherzusagen sowie unsere Verhaltensmuster dementsprechend anzupassen, ist ein wichtiger Bestandteil unseres alltäglichen Lebens, insbesondere in Hinblick auf sich ständig verändernde Umweltbedingungen. Bereits Anfang des 20. Jahrhunderts postulierte Thorndike (1927) in seinem „law of effect“, dass die Auftretenswahrscheinlichkeit eines spezifischen Verhaltens wesentlich von den darauffolgenden Konsequenzen abhängt und auf diese Weise Verhaltensmuster etabliert werden können. Sowohl durch Belohnung (Annäherungslernen) als auch

durch Bestrafung (Vermeidungslernen) kann Verhaltensmodifikation stattfinden, mit dem Ziel, dass ein Organismus auf diese Weise Belohnung maximieren und Bestrafung minimieren kann (Sutton and Barto, 1998). Dieser Prozess wird auch als instrumentelles Lernen bezeichnet.

1.1.1. Klassische Konditionierung

Der Begriff der klassischen Konditionierung geht auf den russischen Physiologen Pawlow und die Interpretation seiner experimentellen Beobachtungen im Rahmen der behavioristischen Lerntheorie zurück (Pawlow, 1927). Sie wird daher oftmals auch als Pawlowsche Konditionierung bezeichnet. Klassische Konditionierung stellt letztlich das Erlernen von Assoziationen zwischen verschiedenen Reizen dar. Die Theorie besagt, dass ein unkonditionierter Reiz bzw. Stimulus (US) regelhaft bzw. reflexartig zu einer unkonditionierten Reaktion (UCR) führt. Unkonditioniert bedeutet in diesem Zusammenhang, dass sowohl Stimulus als auch Reaktion bereits vor dem Experiment vorhanden waren bzw. gewissermaßen angeboren sind. Bei UCRs handelt es sich demnach um physiologische Antworten im Sinne von angeborenen motorischen oder auch emotionalen Reaktionen (z.B. Speichelabsonderung, Schockstarre, aber auch Annäherung, Flucht, Angst), wobei diese in ihrem Spektrum begrenzt sind. Als US hingegen bezeichnet man jene Stimuli, die für einen Organismus eine angeborene Bedeutung haben z.B. Futter oder Fortpflanzung. Klassische Konditionierung findet dann statt, wenn ein US zusammen mit einem vormals neutralen Reiz gepaart wird, wobei letzterer anfangs üblicherweise keine spezifische Reaktion auslöst. Durch die wiederholte gleichzeitige Darbietung beider Reize erlangt der bislang neutrale Reiz zunehmend an Bedeutung. Schließlich kann dieser selbst, auch in Abwesenheit des US, die entsprechende Reaktion auslösen. Der vormals neutrale Reiz wird demnach zu einem konditionierten Reiz (CS), der eine konditionierte Reaktion (CR) hervorruft. Durch die Kopplung mit dem US hat der CS also erst eine motivationale Bedeutung erlangt. Klassische Konditionierung wird folglich auch als Assoziationslernen zwischen prädiktiven Schlüsselreizen und motivational relevanten Reizen bezeichnet (Stimulus-Stimulus-Assoziation).

1.1.2 Operante Konditionierung und instrumentelles Lernen

Klassische Konditionierung allein erklärt aber noch nicht, wie es zum Erlernen neuer Verhaltensweisen kommt oder was Organismen dazu befähigt, auch in einer neuen

Umgebung durch Anpassung zu überleben. Einen wesentlichen Beitrag zur Weiterentwicklung des Konzepts der klassischen Konditionierung trug Thorndike mit dem Prinzip des operanten Konditionierens als Reiz-Reaktions-Lernen bei (Thorndike, 1898). Das „law of effect“ besagt (Thorndike, 1927), dass die Auftretenswahrscheinlichkeit eines bestimmten Verhaltens oder Reaktionsmusters nachhaltig durch dessen angenehme oder unangenehme Konsequenzen verändert wird. So lernt der Organismus beispielsweise, nach der Präsentation eines bestimmten Reizes vermehrt dasjenige Verhalten zu zeigen, welches zuvor belohnt wurde. Der Organismus als ein lernfähiges Individuum modifiziert also spezifisches, zielorientiertes Verhalten anhand von Konsequenzen. Das eigene Verhalten wird folglich gezielt (als Instrument) eingesetzt, um ein gewünschtes Ziel zu erreichen, woraus der Begriff des instrumentellen Lernens entstanden ist. Jener Begriff wird oftmals mit demjenigen der operanten Konditionierung gleichgesetzt, wobei Letzterer die Modifikation spontanen Verhaltens durch Bestrafung oder Belohnung meint.

Ob das Auftreten eines bestimmten Verhaltens begünstigt oder erschwert wird, ist abhängig vom eingesetzten Verstärker, dessen Erhalt durch eine bestimmte instrumentelle Handlung erreicht werden soll. Es werden sowohl primäre als auch sekundäre Verstärker unterschieden. Primäre Verstärker sind solche, die mit einer angeborenen, positiven Bedeutung für den Organismus einhergehen wie z.B. Futter. Ihre verhaltensverstärkende Wirkung beruht demnach darauf, dass sie physiologische, überlebenswichtige Bedürfnisse stillen. Sekundäre Verstärker hingegen sind durch klassische Konditionierung aus primären Verstärkern hervorgegangen und erlangen ihre Bedeutung durch Kontingenz mit dem primären Verstärker (z.B. für Geld kann man Nahrung erwerben). Was letztlich als Verstärker fungiert ist vom Individuum abhängig, da dies oft auf motivationalen Beweggründen beruht (z.B. Hunger oder auch Bedürfnis nach Zuwendung).

Interessanterweise konnte gezeigt werden, dass instrumentelles Lernen nicht nur abhängig von Konsequenzen ist, sondern durchaus auch durch klassisch konditionierte Reize moduliert werden kann, selbst wenn diese per se nicht mit dem instrumentellen Verhalten in Zusammenhang stehen (Talmi et al., 2008). Ein häufiges Beispiel hierfür ist in Supermärkten eingesetzte Musik oder auch Düfte, welche Käufer in ihrem Entscheidungsverhalten beeinflussen sollen. Dabei stellen Musikstücke oder Düfte konditionierte Reize dar, welche vom Individuum aufgrund Pawlowscher Konditionierung mit positiven Eigenschaften oder angenehmen

Zuständen assoziiert sind (Stimulus-Stimulus-Assoziation). Werden diese positiv konditionierten Reize nun gleichzeitig mit einer Produktwerbung präsentiert, so können sie verstärkend auf das Einkaufsverhalten einwirken, mit der Folge, dass der Kunde zum Kauf des Produktes animiert wird (instrumentelles Verhalten). Dem Einkaufsverhalten selbst folgt also eigentlich eine andere Belohnung (nämlich der Besitz des Produktes) als jene, die im Rahmen der Pawlowschen Konditionierung zur Ausprägung der Stimulus-Stimulus-Assoziation beitragen hat (z.B. auf Musik folgt angenehmes Gefühl). Dennoch vermag diese positive Assoziation derart auf den Kunden einzuwirken, dass das angenehme Gefühl, welches er bei der Präsentation von Musik oder Düften verspürt, zum Produktkauf anregt, obwohl dem Produkt an sich zuvor keine positiven Eigenschaften zugesprochen worden waren. Dies wird als PIT-Effekt (Pawlowsch-instrumenteller Transfer) bezeichnet (Garbusow et al., 2016).

1.1.3. Extinktionslernen

Extinktion bezeichnet eine Methode des Verhaltensabbaus, wobei dies nicht nur durch bloßes Vergessen oder Verlernen eines vormals durch instrumentelles Lernen erworbene Reaktionsmusters hervorgerufen wird. Vielmehr steht Extinktion für eine weitere Form eines aktiven Lernprozesses. Dabei handelt es sich um ein vorübergehendes und kontextabhängiges Außerkräftsetzen der Wirkung eines Reizes und dem damit einhergehenden Antwortverhalten. Extinktion kann in Hinblick auf das instrumentelle Lernen durch den Wegfall des Verstärkers erreicht werden. Bei der klassischen Konditionierung hingegen wird ein CS so lange ohne den US präsentiert, bis die vormals CR ausbleibt. Während dieses Lernprozesses kommt es also zunehmend zu einer Abschwächung der gelernten Assoziation zwischen Stimulus und Stimulus bzw. Stimulus und Reaktion. Man spricht auch von inhibitorischem Lernen, da es nicht zu einer vermeintlichen Löschung der vormaligen Assoziation kommt, sondern vielmehr zu einer Hemmung ebenjener durch eine andere, neue Assoziation (Bouton, 2002).

1.2. Habituelles und zielgerichtetes Verhalten

Das Konzept der habituellen und zielgerichteten Verhaltensstrategien basiert auf der Grundannahme, dass zwei parallele Systeme existieren, auf denen instrumentelles Entscheidungsverhalten letztlich beruht: einem habituellen, also gewohnheitsmäßigen und einem zielgerichteten System (Dickinson, 1985).

Instrumentelles Verhalten wird dann als zielgerichtet betrachtet, wenn es zwei Kriterien erfüllt: Zum einen basiert es auf dem Wissen über den Zusammenhang zwischen dem eigenen Verhalten (*action* oder *response*) und den darauffolgenden Konsequenzen (*outcome*). Dieser Zusammenhang wird auch als *response-outcome*-Assoziation bezeichnet (R-O-Assoziation). Des Weiteren berücksichtigt zielgerichtetes Verhalten auch den motivationalen Charakter bzw. Anreizwert, welcher den Konsequenzen bzw. dem *outcome* zugesprochen wird. Nur wenn eine Konsequenz in dem Moment der Entscheidung auch motivational von Bedeutung ist, wird das entsprechende Verhalten ausgeführt. Im Gegensatz dazu ist habituelles Verhalten geprägt von der vorausgegangenen Belohnungserfahrung, beruht also auf der Assoziation zwischen einem bestimmten Schlüsselreiz (oder einer bestimmten Situation) und dem Antwortverhalten (*stimulus-response*-Assoziation, kurz S-R) und ist somit unabhängig von der Bewertung bzw. dem aktuellen Anreizwert einer assoziierten Konsequenz. Habituelle Verhaltensmechanismen sind folglich erlernte, automatisierte Handlungsabläufe mit spezifischen Merkmalen, die kontextabhängig auftreten und damit resistent gegenüber kurzfristigen Änderungen der Zielmotivation sind. Intermittierende Verstärkung kann dazu führen, dass habituelles Verhalten besonders gut ausgeprägt wird und stabil bleibt (Hull, 1943).

Letztlich spielen beide Verhaltensmechanismen in unserem alltäglichen Leben eine Rolle und haben sowohl Vor- als auch Nachteile. Insbesondere dem habituellen System und den damit einhergehenden Gewohnheiten lastet mitunter ein negativer Aspekt an, stellen Gewohnheiten doch starre, inflexible, geradezu rigide anmutende Verhaltensmuster dar, die in einer sich ständig verändernden Umwelt hinderlich scheinen. Dabei haben habituelle Verhaltensmechanismen einen hohen Stellenwert in unserem Alltag. Jeden Tag treffen wir zahlreiche Entscheidungen, die meisten davon werden uns nicht einmal bewusst oder erfordern nur einen geringen Aufwand an kognitiven Kapazitäten, da sie eben routinemäßig ablaufen. Untersuchungen zeigen, dass ca. 43% unseres Handelns auf alltäglichen und kontextabhängigen Verhaltensmustern beruhen und wir häufig nicht einmal über diese repetitiven Handlungen nachdenken (Wood et al., 2002). Eben hier liegt der Vorteil von Gewohnheiten: Sie sind effizient und schnell, einmal aktiviert, löst das habituelle System eine erlernte Standardreaktion aus, ohne dass der Organismus kognitive Ressourcen aufwenden muss, um immer wieder aufs Neue über mögliche Verhaltensoptionen nachzudenken. Gewohnheiten ermöglichen uns, selbst komplexe

Bewegungsabläufe durch häufiges Wiederholen derart einzuüben, dass sie schlussendlich ohne explizites Nachdenken ausgeführt werden können und so kognitive Kapazitäten für andere Aufgaben bereitstellen (z.B. während des Autofahrens auf die Schilder, Radfahrer und Ampeln achten). Mitunter werden die Begriffe habituelle Verhaltensmuster und Automatismen deshalb synonym verwendet. Durch bestimmte Trigger ausgelöstes habituelles Verhalten ist jedoch nicht bloß ein motorisches Programm, welches automatisch abläuft, sondern stellt vielmehr ein multimodales Antwortverhalten dar, welches verschiedene Aspekte wie z.B. Emotionen, Vorlieben und Erinnerungen umfasst (beispielsweise bei Essgewohnheiten). Doch gewohnheitsmäßiges Verhalten schränkt aufgrund seiner Inflexibilität auch das Handlungsspektrum eines Organismus ein und kann auf diese Weise die Möglichkeit verhindern, neue Erfahrungen zu machen. Es konnte gezeigt werden, dass wiederholtes, gewohnheitsmäßiges Antwortverhalten in einem bestimmten Kontext zu einer eingeschränkten kognitiven Flexibilität hinsichtlich möglicher Alternativen führt (Danner et al., 2007). Ein typisches Merkmal habitueller Verhaltensmechanismen ist zudem die Insensibilität gegenüber wechselnden Konsequenzen und einer möglichen Änderung des Anreizwertes (Dickinson, 1985). Handeln aus der Gewohnheit heraus kann demnach auch im Konflikt stehen mit derzeitigen Intentionen und Motivationen. Von solchen habituellen Verhaltensmustern wird das zielgerichtete Verhalten abgegrenzt. Zielgerichtetes Verhalten zeichnet sich durch das Wissen über den Zusammenhang zwischen dem eigenen Verhalten und den darauffolgenden Konsequenzen aus, sodass entsprechend des aktuellen Anreizwertes der Konsequenzen das Verhalten geplant und ausgeführt werden kann. Damit bieten zielgerichtete Verhaltensmechanismen eine höhere Flexibilität und damit einen entscheidenden Vorteil in komplexen und sich schnell ändernden Bedingungen. Sie ermöglichen eine Verhaltensadaptation an sich verändernde Umweltbedingungen und Kontingenzen, weil sie auch mögliche wechselnde Konsequenzen des Verhaltens und die augenblicklichen Motive des Organismus in den Entscheidungsfindungsprozess mit einbeziehen (Balleine and Dickinson, 1998). Auch in bisher unbekanntem Situationen kann so eine Verhaltensadaptation stattfinden, da mögliche Konsequenzen simuliert werden, ohne dass sie zuvor erlebt werden müssen. Jedoch erfordert solch ein System auch größere kognitive Ressourcen. Dementsprechend ist zielgerichtetes Verhalten auch von Faktoren wie der kognitiven Leistungsfähigkeit (Schad et al., 2014; Smittenaar et

al., 2013), dem Alter (de Wit et al., 2014), dem Arbeitsgedächtnis (Collins and Frank, 2012) und dem aktuellen Zustand des Organismus abhängig, also ob er beispielsweise gestresst ist (Schwabe and Wolf, 2009).

1.2.1. Operationalisierung habituellem und zielgerichteter Verhaltenskontrolle

Um zu operationalisieren, inwieweit ein Organismus eines der beiden genannten Systeme nutzt, wurden verschiedene Paradigmen entwickelt. Grundlage bildeten zunächst tierexperimentelle Untersuchungen (Dickinson, 1985), welche später erfolgreich für Studien am Menschen adaptiert wurden (Horstmann et al., 2015; Tricomi et al., 2009; Valentin et al., 2007). Ob ein Organismus zielgerichtetes oder habituelles Verhalten zeigt, wird vorrangig mittels so genannter Devaluationsparadigmen überprüft. Diese Paradigmen unterteilen sich üblicherweise in drei Phasen: Training, Devaluation und Test (Abbildung 1). Bei diesen Experimenten lernen die Probanden zunächst, auf einen bestimmten Stimulus hin ein Antwortverhalten zu zeigen, um zu einer Belohnung zu gelangen (instrumentelle Lernaufgabe). Im nächsten Schritt wird beispielsweise die Belohnung (wie Nahrung) so lange zur Verfügung gestellt, bis eine Sättigung eintritt und die vormalige Belohnung nicht mehr als angenehm empfunden wird (Devaluation bzw. Abwertung des Verstärkers). Die Belohnung wird folglich weniger erstrebenswert, sie gilt als devaluiert. Habituelles Verhalten nach der Devaluation zeichnet sich im Wesentlichen dadurch aus, dass weiterhin derjenige Stimulus ausgewählt wird, der mit der devaluierten Belohnung bzw. Rückmeldung assoziiert ist, obwohl dieses Auswahlverhalten ja nun zu einer eigentlich unerwünschten Verhaltenskonsequenz führt. Ursache hierfür ist, dass das System auf der Assoziation zwischen Stimulus und Antwortverhalten beruht und darüber hinaus den Wert der durch die Devaluation veränderten Handlungskonsequenzen nicht berücksichtigt. Ein Proband mit zielgerichtetem Verhalten hingegen wird den mit der devaluierten Belohnung assoziierten Stimulus meiden, weil er auch das Wissen um den aktuellen motivationalen Wert der Belohnung mit einbezieht (Adams and Dickinson, 1981).

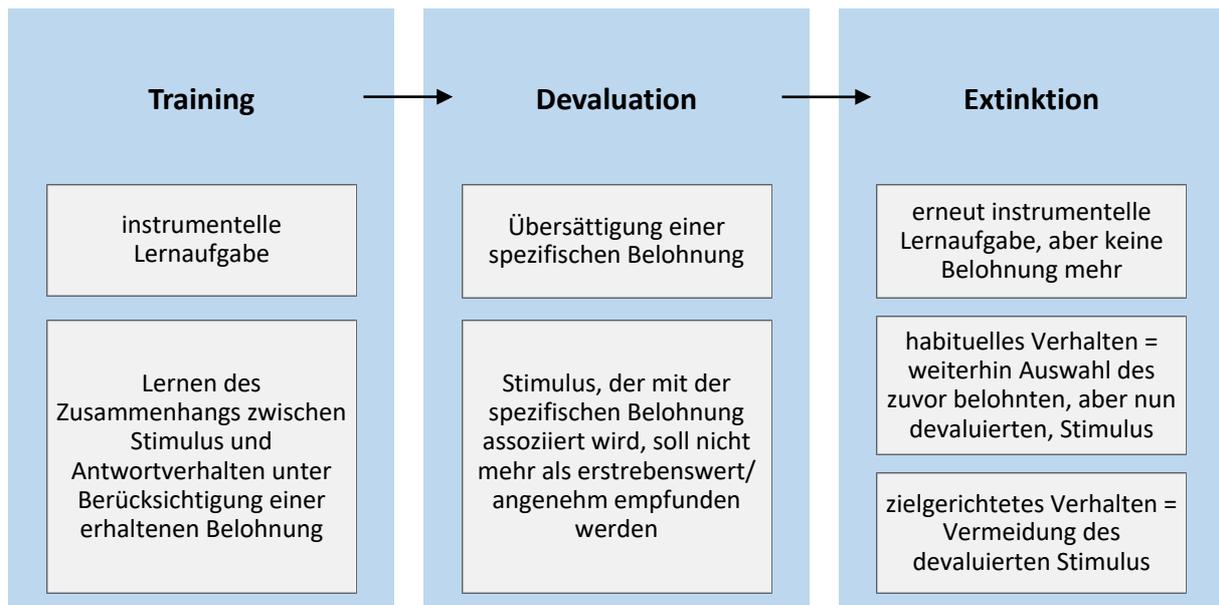


Abbildung 1: allgemeiner Aufbau eines Devaluationsparadigmas

unabhängig vom genauen experimentellen Design, durchlaufen Probanden bei einem Devaluationsparadigma stets drei Abschnitte

Ein anderer Ansatz zur Operationalisierung beruht auf dem sog. Kontingenzabbau (Balleine and Dickinson, 1998; Liljeholm et al., 2011). Hierbei wird die Kontingenz, also der Zusammenhang bzw. das Häufungsverhältnis zwischen Handlung und Belohnung modifiziert, welche in den Entscheidungsprozess des zielgerichteten Systems mit einbezogen wird. Man stelle sich eine Ratte vor, die im Rahmen einer instrumentellen Lernaufgabe durch Drücken eines Hebels Belohnung in Form von Futter erhält. Ändern sich im Verlauf der Aufgabe die Bedingungen und die Ratte erhält beispielsweise Futter unabhängig davon, ob sie den Hebel drückt oder nicht, so wird zielgerichtetes Verhalten durch eine reduzierte Frequenz des Hebeldrückens angezeigt. Ein Überwiegen des habituellen Systems würde hingegen in diesem Fall zu weiterem Drücken des Hebels führen (Dickinson and Balleine, 1994). Es konnte gezeigt werden, dass habituelles Verhalten – gemessen entweder durch Devaluationsparadigmen oder Kontingenzabbau – durch extensives, also übermäßiges Training im Verlauf der instrumentellen Lernaufgabe induziert werden kann (Adams and Dickinson, 1981). Ein entscheidender Nachteil der Paradigmen ist jedoch, dass sie häufig nur ein oder wenige Merkmale des jeweiligen Verhaltenssystems erfassen und abbilden können. Darüber hinaus existieren aber auch weitere Variablen und Störfaktoren, die die jeweiligen Strategien der Verhaltenskontrolle beeinflussen und durch die entsprechenden

Operationalisierungen nicht unbedingt erfasst werden. Letztere bilden folglich also nur einige Aspekte eines eigentlich komplexen Systems ab, was bei der Interpretation der Ergebnisse solcher Verhaltensexperimente berücksichtigt werden sollte.

1.2.2. Neurobiologische Korrelate habituellen und zielgerichteten Verhaltens

Da experimentelle Verhaltensstudien und theoretische Überlegungen zu habituellem und zielgerichtetem Verhalten allein keine ausreichende Erklärung der zugrundeliegenden Mechanismen boten, rückte in den letzten Jahrzehnten vor allem die Untersuchung der neuronalen Netzwerke, auf denen diese Verhaltensmuster beruhen, in den Fokus der Neurowissenschaften. Ein Großteil des Verständnisses über die neurobiologischen Korrelate von habituellem und zielgerichtetem Verhalten stammt aus tierexperimentellen Untersuchungen, v.a. an Ratten, bei denen mithilfe von induzierten Läsionen, Inaktivierungen und pharmakologischer Manipulation verschiedene neuronale Areale untersucht worden sind. Dabei rückte vor allem das Striatum, welches als subkortikales Kerngebiet innerhalb des Großhirns liegt, in den Fokus des Forschungsinteresses: Seinen distinkten Subregionen wird eine differentielle Rolle bei der Vermittlung der neurobiologischen Systeme von habitueller und zielgerichteter Verhaltenskontrolle zugeschrieben. Dabei konnte in mehreren Studien an Ratten nachgewiesen werden, dass innerhalb des dorsomedialen Striatums ein Netzwerk zu existieren scheint, welches zielgerichtetes Verhalten fördert (Balleine and Dickinson, 1998; Corbit and Balleine, 2005; Gremel and Costa, 2013; Yin et al., 2005). Habituelle Verhaltensweisen werden dagegen dem dorsolateralen Striatum zugeschrieben (Balleine, 2005; Balleine and Dickinson, 1998; Corbit and Balleine, 2005; Yin et al., 2004). Das dorsomediale Striatum steht überwiegend mit kortikalen Assoziationsarealen in Verbindung, wohingegen das dorsolaterale Striatum eng verbunden ist mit dem sensorischen und motorischen Kortex (Swanson, 2000). Des Weiteren konnte bei Ratten gezeigt werden, dass Läsionen im dorsomedialen Striatum und im prälimbischen Kortex zu einer Blockade zielgerichteten Verhalten führt (Balleine and Dickinson, 1998; Yin et al., 2005; Yin and Knowlton, 2006). Im Gegensatz dazu resultieren Läsionen im dorsolateralen Striatum oder eine Blockade seines dopaminergen Inputs in einer Aufrechterhaltung von zielgerichtetem Verhalten selbst bei extensivem Training (Balleine and Dickinson, 1998; Faure et al., 2005; Killcross and Coutureau, 2003; Yin et al., 2004).

Interessanterweise gibt es auch Hinweise darauf, dass es eine Wechselwirkung zwischen beiden Systemen zu geben scheint: In Untersuchungen konnte belegt werden, dass es beispielsweise im Laufe des Trainings von einer anfänglich starken Aktivierung des dorsomedialen Striatums hin zu einem Überwiegen seiner dorsolateralen Anteile kommen kann – als Ausdruck für das allmähliche Überwiegen von habituellen über zielgerichtete Verhaltensmuster (Thorn et al., 2010; Yin et al., 2009). So sehr einige Untersuchungen also den Eindruck vermitteln, dass diesen beiden Strategien der Verhaltenskontrolle auch zwei voneinander getrennte neuronale Netzwerke zugrunde liegen, eine strikte Unterteilung kann und sollte nicht vorgenommen werden.

Die Ergebnisse von Tierstudien an Ratten sind selbstverständlich nicht uneingeschränkt auf Menschen übertragbar. So wird das dorsale Striatum bei Menschen und Primaten von zwei Strukturen gebildet: dem Putamen und dem Nucleus caudatus. Jene beiden Strukturen werden durch eine interne Kapsel voneinander getrennt. Solch eine Unterteilung liegt bei Ratten nicht vor. Dennoch herrscht mittlerweile Konsens darüber, dass homologe Areale existieren: So entspricht das dorsolaterale Striatum bei Ratten dem Putamen. Das dorsomediale Striatum hingegen ist mit dem Nucleus caudatus beim Menschen gleichzusetzen (Balleine and O'Doherty, 2010). Dementsprechend zeigten sich vergleichbare Befunde auch in Humanstudien. Hier konnte durch funktionell bildgebende Verfahren während entsprechender Lernaufgaben gezeigt werden, dass bei Menschen zielgerichtetes Verhalten vor allem durch den ventromedialen präfrontalen Kortex (PFC) und den Nucleus caudatus (de Wit et al., 2012) sowie durch den medialen orbitofrontalen Kortex (OFC), welcher ein Subareal des präfrontalen Kortex darstellt, vermittelt wird (Valentin et al., 2007). Habituelle Lernmechanismen hingegen sind mit einer höheren Aktivität des posterioren Putamens (Tricomi et al., 2009, 2009) und erhöhter Konnektivität zwischen dem sensomotorischen Kortex und dem Putamen (Horga et al., 2015) assoziiert. Zudem korreliert der Grad der Aktivierung des posterioren Putamens (Tricomi et al., 2009) sowie dessen Konnektivität mit dem prämotorischen Kortex mit dem individuellen Grad habituellen Verhaltens (de Wit et al., 2012). Eine ausführliche Übersicht über die bisherigen Erkenntnisse neurobiologischer Untersuchungen bieten Dolan und Dayan (2013) sowie Daw und O'Doherty (2014).

1.3. Modellbasiertes und modellfreies Verhalten

Dieses Konzept stammt aus dem Bereich der computationalen Neurowissenschaften und ist teilweise parallel, aber auch als Reaktion auf die Befunde der klassischen, theoretischen Psychologie entwickelt worden. Das Prinzip der computationalen Neurowissenschaften liegt darin, Bestandteile des Nervensystems mittels mathematischer Modelle zu beschreiben, um so ein genaueres Verständnis der informationsverarbeitenden Prozesse zu erhalten. Im Falle der Lern- und Verhaltenspsychologie wird also versucht, das beobachtbare Lernverhalten mithilfe eines mathematischen Modells nachzubilden. Diese Modelle beschreiben demnach eine Vorstellung davon, wie Lern- und Entscheidungsprozesse ablaufen, und müssen daraufhin geprüft werden, inwieweit sie die beobachteten Prozesse – nämlich das letztlich gezeigte Verhalten – tatsächlich abbilden. Diese Modelle können immer nur eine Annäherung an die Komplexität der Wirklichkeit sein.

Historisch gesehen geht das Konzept modellbasierten und modellfreien Verhaltens auf Tolmans Theorie der kognitiven Karten zurück (Tolman, 1948). Einer der wichtigsten Arbeiten Tolmans war die Erforschung des Lernens räumlicher Zusammenhänge (sog. Ortslernen) bei Ratten in Labyrinthen. Dabei stellte er fest, dass die Ratten mit zunehmender Erfahrung immer besser dazu in der Lage waren, ein Labyrinth zu durchqueren. Sie machten zunehmend weniger Fehler und benötigten auch sukzessiv weniger Zeit für die Labyrinthaufgabe. Im Gegensatz zu Thorndike (1927) schloss Tolman daraus, dass das Verhalten der Ratten nicht nur mit einem reinen Reiz-Reaktionsmuster erklärt werden könne. Vielmehr postulierte Tolman (1948), dass bei der Suche nach dem optimalen Weg innere, kognitive Prozesse durchlaufen werden müssen und folglich der Aufbau einer „inneren Landkarte“ stattfindet. Mit Hinblick auf die Verhaltenskontrolle bietet solch eine kognitive Karte gewissermaßen eine gedankliche Vorlage, die es dem Tier ermöglicht, der Situation entsprechend das bestmögliche Antwortverhalten zu finden. Darüber hinaus werden beim Aufbau dieser Karte auch Erwartungen über mögliche Verhaltenskonsequenzen mit einbezogen und überprüft z.B. ob nach dem Abbiegen in eine Gasse eine Mauer oder ein weiterer Gang zu erwarten ist.

Tolman konnte allerdings aufgrund fehlenden mathematischen Verständnisses von Informationsverarbeitungsprozessen und Algorithmen keine umfassende Erklärung dafür liefern, wie dieses komplexe System der inneren Landkarte letztlich entsteht, aufgebaut ist und arbeitet. Dies erschwerte auch die Etablierung geeigneter

Experimente, da kaum quantifiziert werden konnte, wie komplex und herausfordernd die jeweiligen Aufgaben für die Tiere waren. Die Lösung für dieses Problem lieferten schließlich dynamische Programmierungstechniken (Bellman, 1957) und entsprechende Lernmodelle (Sutton and Barto, 1998). Indem die zugrundeliegenden Lernprozesse anhand computationaler Mechanismen dargestellt wurden, konnten mithilfe der computationalen Neurowissenschaft demnach Entscheidungsprozesse präzisiert und analysiert werden. So entstand auch die Theorie des *reinforcement learning* (Theorie vom bestärkenden Lernen), welche eine computationale Beschreibung von instrumentellem Lernen sowie optimaler Verhaltenskontrolle darstellt (Sutton and Barto, 1998). Sie geht der Frage nach, wie Systeme lernen können, durch Verhaltensmodifikation Belohnung zu maximieren und Bestrafung zu minimieren. Bei einer so genannten modellbasierten Verhaltensstrategie erschafft sich der Agent (so wird in der Theorie der Organismus genannt) ein Modell der kausalen Strukturen der Welt. Dabei werden bei der Entwicklung jenes Modells vorherige Lernerfahrungen, deren Konsequenzen und letztlich deren Bewertung genutzt, um ein möglichst differenziertes und umfangreiches Modell der Welt zu konzipieren. Der Agent nutzt dieses innere Modell, um die möglichen Konsequenzen des eigenen Verhaltens unter gegebenen Bedingungen vorauszusagen, diese zu bewerten und daraufhin zu überprüfen, ob sie mit seinen derzeitigen Wünschen und Bedürfnissen im Einklang stehen. Der Agent muss also in der Lage sein, sich selbst in der Zukunft zu imaginieren und fortwährend seine Strategie, und somit auch sein Verhalten, anzupassen, um so langfristig gesehen den für ihn idealen Zustand zu erreichen. Der Vorteil bei dieser Form des Entscheidungsprozesses ist, dass es den Agenten dazu befähigt, Entscheidungen mitunter auch in für ihn unbekanntem oder sich schnell ändernden Situationen zu treffen, für die bisher kein etabliertes Lernmuster besteht. Die Grundlage für eine Entscheidungsfindung in solchen Situationen bilden neben den aktuellen motivationalen Beweggründen letztlich jene Annahmen und theoretischen Überlegungen, die auf das vom Agenten erschaffene Modell der Welt zurückgehen. Die fortlaufende Anpassung und Aktualisierung dieses Modells ermöglicht es dem Agenten, in unbekanntem und veränderlichen Situationen die für ihn notwendigen Informationen in seinen Entscheidungsprozess zu integrieren, sich so schnell und flexibel an die Situation zu adaptieren und letztlich das entsprechende Verhalten zu zeigen. Diese Fähigkeit erscheint auch deshalb so wichtig, weil es einem Organismus kaum gegeben ist, jedwede Lernerfahrungen zu

machen, die notwendig wären, um sich in einer neuen oder veränderten Umwelt zu bewegen.

Auf Grundlage dieser Überlegungen zum *reinforcement learning* entstand schließlich das Konzept von modellbasiertem und modellfreiem Verhalten, wobei der Begriff „Modell“ hier nicht im computationalen, sondern mentalen Sinne zu verstehen ist (Daw et al., 2005). Diese beiden Strategien der Verhaltenskontrolle stellen letztlich Endpunkte des großen Spektrums des *reinforcement learning* dar. Eine modellbasierte Strategie zeichnet sich in diesem Konzept dadurch aus, dass im Rahmen des Entscheidungsfindungsprozesses ein inneres Modell der Umwelt aufgebaut und so vorausschauendes Planen ermöglicht wird. Vereinfacht gesehen kann dieses innere Modell beispielsweise als Entscheidungsbaum konzipiert werden, der regelmäßig aktualisiert wird und dessen Ausgangspunkt den aktuellen Zustand des Organismus darstellt. Von diesem Ausgangspunkt gehen viele Verzweigungen ab, die für alle möglichen Verhaltens- und Entscheidungsoptionen stehen, die sich wiederum in mögliche Konsequenzen gliedern. Die Endpunkte eines jeden möglichen Pfades werden schlussendlich evaluiert und gegeneinander abgewogen. Ähnlich wie beim Konzept des zielgerichteten Verhaltens, beinhaltet auch modellbasiertes Verhalten folglich die Kenntnis über die kausalen Zusammenhänge von Verhalten und Konsequenzen. Dadurch ist dieses System sehr flexibel, bedeutet aber auch mit zunehmender Größe des Entscheidungsbaumes eine enorme Prozessleistung, die auf mentaler Ebene ein hohes Maß an Aufmerksamkeit und Arbeitsgedächtnis voraussetzt. Im Gegensatz dazu basiert modellfreies Verhalten nicht auf einem solch komplexen Modell, wie es zuvor beschrieben wurde. Die Basis der Entscheidungsfindung besteht hier in gespeicherten Belohnungswerten für die möglichen Handlungsoptionen, die auf den bisherigen Lernerfahrungen beruhen. Zukünftiges Verhalten wird daran orientiert und bewertet, wie der Agent in der Vergangenheit für ein bestimmtes Verhalten belohnt oder bestraft wurde. Kommt es möglicherweise in der aktuellen Situation zu Änderungen der Verhaltenskonsequenzen oder der individuellen motivationalen Faktoren, so werden diese nicht im Entscheidungsprozess berücksichtigt. Diese Änderungen müssen erst erfahren werden. Eine schnelle Verhaltensanpassung aufgrund von vorausschauendem Planen wie bei modellbasierten Strategien kann nicht stattfinden. Daher ist eine modellfreie Strategie ein retrospektiver Prozess. Modellfreies Lernen kann durch mathematische Modelle, die auf Lernen von einfachen so genannten

Belohnungsvorhersagefehlern (*reward prediction error, RPE*) beruhen, beschrieben werden. Ein RPE kodiert einen Abgleich zwischen dem, was der Agent infolge seines Verhaltens erwartet, und dem, was er am Ende tatsächlich erhält (Sutton, 1988). Je kleiner der Vorhersagefehler, desto größer ist die Übereinstimmung zwischen den erwarteten und erfahrenen Verhaltenskonsequenzen. Der Vorteil der modellfreien Strategie liegt darin, dass sie eher implizit und computational effizient ist. Sie ermöglicht dem Agenten in stabilen Situationen schnell zu handeln, allerdings auf Kosten einer größeren Inflexibilität. In Situationen mit sich schnell verändernden Bedingungen ist ein modellfreier Lerner im Nachteil, da es ihm nur gelingt, sich durch zurückliegende Erfahrungen von Belohnung und Bestrafung an die neue Situation anzupassen – und nicht durch vorausschauendes Planen.

In der Theorie des *reinforcement learning* wird zunehmend davon ausgegangen, dass diese beiden Strategien der Verhaltenskontrolle nicht durch vollständig voneinander getrennt arbeitende neuronale Systeme realisiert werden, sondern diese neurobiologischen Systeme durchaus synergistisch arbeiten (Cooper et al., 2014; Daw et al., 2005). Um die Anteile von modellfreiem und modellbasiertem Verhalten zu quantifizieren wurden sequentielle Entscheidungsaufgaben wie der *decision-tree* (Huys et al., 2012) und die *two-step*-Aufgabe (Daw et al., 2011) entwickelt. Bei diesen Paradigmen müssen die Probanden eine Folge von Entscheidungen treffen um letztlich zu einer Belohnung zu gelangen (Markow-Entscheidungsprozess).

1.3.1. Neurobiologische Korrelate modellbasierten und modellfreien Verhaltens

Eine entscheidende Rolle beim Verständnis der neuronalen Mechanismen von modellbasiertem und modellfreiem Verhalten ist die neuronale Repräsentation des Prädiktionsfehlers, auch Vorhersagefehler genannt (*prediction error*). Dieser bezieht sich im Fall einer modellfreien Strategie auf die mögliche Belohnung, stellt also die Diskrepanz zwischen erwarteter und erhaltener Belohnung dar (*reward prediction error, RPE*, Belohnungsvorhersagefehler). In erster Linie wurden subkortikale Strukturen wie das ventrale Striatum als jene Bereiche detektiert, die den RPE neuronal repräsentieren und somit ein neurobiologisches Korrelat modellfreier Strategien darstellen (Gläscher et al., 2010; O'Doherty et al., 2003). Bei modellbasierten Verhaltensweisen spielt dieser Prädiktionsfehler nur indirekt eine Rolle, vielmehr werden auch Überlegungen zu dem erwarteten Modell von der

Umwelt oder Aufgabenstruktur mit einbezogen. Dabei stellt der so genannte *state prediction error* (SPE) die Diskrepanz zwischen dem erwarteten und dem letztlich erfahrenen Zustand der Umwelt dar, wobei die Erwartung auf einem Modell der Umwelt basiert. Untersuchungen konnten zeigen, dass der SPE vor allem im dorsolateralen präfrontalen Kortex (dlPFC) und im intraparietalen Sulcus repräsentiert wird (Gläscher et al., 2010). Die Inaktivierung des dlPFC mittels transkranieller Magnetsimulation führte zudem zu einer Abschwächung modellbasierten Verhaltens in einer sequentiellen Entscheidungsaufgabe (Smittenaar et al., 2013). Darüber hinaus scheinen auch der mediale orbitofrontale Kortex (mOFC) und der ventromediale präfrontale Kortex (vmPFC) als weitere Subareale des präfrontalen Kortex bei der modellbasierten Kontrolle von Entscheidungsverhalten von Bedeutung zu sein, wie strukturelle Bildgebungsbefunde nahelegen (Voon et al., 2015). Dabei dürften der vmPFC und der mOFC allerdings jeweils unterschiedliche Aspekte modellbasierter Verhaltensweisen repräsentieren (Voon et al., 2017): Der vmPFC ist insbesondere beim Lernen über den Zusammenhang zwischen Antwortverhalten und Belohnung (R-O-Assoziation), aber auch bei sequentiellen Entscheidungsprozessen involviert. Der mOFC hingegen repräsentiert zum einen solche kognitiven Prozesse, die wichtig sind für die interindividuelle Abwägung zwischen den verschiedenen Verhaltensstrategien. Andererseits ist der mOFC bedeutsam für die Integration notwendiger Informationen in den Entscheidungsvorgang, wie beispielsweise dem Wert einer – unter Berücksichtigung der Aufgabenstruktur – erwarteten Konsequenzen bzw. Belohnung. In Bezug auf die neurobiologischen Strukturen, die modellfreiem und modellbasiertem Verhalten unterliegen, scheint es keine eindeutige Trennung zwischen beiden Systemen zu geben: So wurde wiederholt eine räumliche Überlagerung der neuronalen Repräsentationen beider oben genannter Prädiktionsfehler beschrieben (L. Deserno et al., 2015; Lorenz Deserno et al., 2015), beispielsweise im ventralen Striatum und medialen präfrontalen Kortex (Daw et al., 2011).

Aus pharmakologischer Sicht wurde vor allem die Vermittlung neuronaler Signale zwischen den beteiligten Hirnstrukturen mittels dem modulierendem Neurotransmitter Dopamin sowohl für modellfreies als auch für modellbasiertes Verhalten intensiv untersucht (Cools, 2011; Montague et al., 1996; Schultz et al., 1997). Dopamin gilt in den Neurowissenschaften als bedeutsamer Botenstoff in der Modulation von Lern-

und Entscheidungsprozessen bei der Verhaltenskontrolle. So führte beispielsweise die Einnahme von L-Dopa bei Probanden zu einer verstärkten Einbeziehung modellbasierter Strategien in den Entscheidungsprozess (Wunderlich et al., 2012). Ein weiteres wichtiges neurobiologisches Korrelat für das Verständnis beider Systeme ist zudem der Hippocampus, welcher vor allem an Ratten gut untersucht ist. Der Hippocampus gilt als neuronale Repräsentation der sogenannten kognitiven Landkarte. Eine der bekanntesten Funde des Hippocampus ist die Existenz von „Orts-Nervenzellen“ bei Ratten, welche eine interne, neuronale Abbildung für eine Karte des umgebenden Raumes darstellen und somit die räumliche Orientierung und Planung ermöglichen (O’Keefe and Nadel, 1978). Sich an vergangene Erfahrungen zu erinnern (episodisches Gedächtnis), sich eine mögliche Zukunft zu imaginieren sowie über die Gegenwart hinaus zu planen und mögliche Konsequenzen mit einzubeziehen (auch wenn ich sie bisher nicht selbst erfahren habe), sind grundlegende Fähigkeiten modellbasierten Verhaltens. Es konnte gezeigt werden, dass hippocampale Läsionen bei Menschen zu Einschränkungen eben jener Fähigkeiten führen (Addis and Schacter, 2011; Cohen and Eichenbaum, 1995). Demzufolge gilt – neben dem dlPFC, dem mOFC und dem vmPFC – auch der Hippocampus als eine bedeutsame Struktur im Prozess der modellbasierten Entscheidungsfindung.

1.4. Die Rolle veränderter Lern- und Verhaltensmechanismen bei der Entstehung und Aufrechterhaltung der Alkoholabhängigkeit

In den letzten Jahrzehnten wurden viele verschiedene Konzepte für das Verständnis von der Entstehung psychischer Erkrankungen entwickelt, wobei sich das biopsychosoziale Modell hier am meisten etabliert hat (Wittchen and Hoyer, 2011). Ein neuerer Untersuchungsansatz basiert darauf, psychische Erkrankungen als erlernte Verhaltensmuster zu verstehen (Garbusow et al., 2013; Heinz, 2017; Heinz et al., 2013; Wittchen and Hoyer, 2011). Dabei wird davon ausgegangen, dass sich genetische und neurobiologische Faktoren derart auf Lernprozesse (einschließlich basaler Lernmechanismen wie die Pawlowsche und operante Konditionierung) auswirken, dass es zu einer systematischen Veränderung von Verhaltensweisen kommt, die sowohl zur Entstehung als auch zur Aufrechterhaltung von psychischen Erkrankungen beitragen. Solch ein Ansatz mag kritisch betrachtet eine vereinfachte Darstellung der wirklichen zu Grunde liegenden Mechanismen sein und deren

Komplexität nicht gerecht werden. Aber er bietet auch die Möglichkeit, eine Brücke zu schaffen zwischen den wissenschaftlich nachweisbaren (neuro-)biologischen Veränderungen und den zu beobachtenden klinischen Störungen. Ein wesentlicher Vorteil der Fokussierung auf Lernmechanismen als ein entscheidender Einflussfaktor bei der Entstehung und Aufrechterhaltung psychischer Erkrankungen ist zudem die Betrachtung der vielschichtigen individuellen Symptomausprägungen, die gerade bei solchen Erkrankungen deutlich werden und deren Varianz womöglich auf veränderten basalen Lernmechanismen als Folge des Einflusses persönlicher Erfahrungen beruhen. Die bisherige Einteilung psychiatrischer Diagnosen in verschiedene Kategorien spiegelt eben nicht unbedingt deren zugrundeliegenden genetischen, psychologischen und neurobiologischen Korrelate sowie Dysfunktionen wieder. Vielmehr sind sie eine deskriptive Beschreibung von psychiatrischen Klassifizierungen und beruhen weniger auf ätiologischen Phänomenen. Dies ist allein daran zu erkennen, dass es häufig hohe Überschneidungen von Symptomen zwischen den einzelnen Diagnosen gibt. Mitunter können aber auch zwei Patienten mit der gleichen Hauptdiagnose sehr unterschiedliche Symptome aufweisen (z.B. Schizophrenie). In Zusammenhang mit dem hier beschriebenen Untersuchungsansatz sind vor allem jene psychischen Erkrankungen erforscht worden, die mit zwanghaften Verhaltensweisen einhergehen. Unter der Symptomdimension „Zwanghaftigkeit“ versteht man ein theoretisches Konstrukt, welches bei vielen psychischen Störungen zu finden ist, und unter anderem einhergeht mit einem Kontrollverlust bezüglich repetitiver, letztlich auf lange Sicht gesehen selbstschädigender, Verhaltensmuster (Dalley et al., 2011; Gillan et al., 2016). Diese Symptomdimension wird insbesondere Abhängigkeits- und Zwangserkrankungen zugeschrieben (Everitt and Robbins, 2005; Gillan and Robbins, 2014). Der zwanghafte Charakter dieser Erkrankungen wird von Forschern mit einem Ungleichgewicht zwischen zwei verschiedenen Strategien der Verhaltenskontrolle (also zielgerichtet vs. habituell bzw. modellbasiert vs. modellfrei) assoziiert (Everitt and Robbins, 2005), welches wiederum einen Vulnerabilitätsfaktor für die Entwicklung rigider Verhaltensweisen darstellt. Defizite in zielgerichtetem bzw. modellbasiertem Verhalten lassen sich allerdings auch bei solchen psychischen Erkrankungen finden, die nicht mit einem zwanghaften Charakter assoziiert sind, u.a. Schizophrenie (Morris et al., 2015), soziale Phobie (Alvares et al., 2014) und Depression (Watkins and Nolen-Hoeksema, 2014). Somit scheint eine Verlagerung

der Verhaltensweisen hin zu habituellen bzw. modellfreien Entscheidungsprozessen nicht unbedingt spezifisch für das Symptom „Zwanghaftigkeit“ zu sein. Eine entscheidende Voraussetzung für die Nutzung modellbasierter Strategien ist zudem die Fähigkeit, ein inneres Modell der Umwelt und möglicher Handlungskonsequenzen zu repräsentieren, um folglich unter Berücksichtigung motivationaler Prozesse eine Entscheidung zu treffen. Es gibt Hinweise darauf, dass diese Fähigkeit bei Abhängigkeits- und Zwangserkrankungen sowie der Binge-Eating-Störung eingeschränkt ist (Gillan et al., 2014; Reiter et al., 2017, 2016). Demnach scheinen von solchen Erkrankungen betroffene Personen derart in ihrem Handlungsspektrum beeinflusst, dass sie potentiell schädigende Verhaltensweisen wie Substanzkonsum, Essattacken oder eine permanente Wiederholung von Handlungen nur schwer unterlassen können. Obwohl jene Patienten also durchaus die Motivation besitzen, derartige Verhaltensmuster aufgrund der langfristig negativen Konsequenzen abzubauen, überwiegen anscheinend andere Aspekte der Entscheidungsfindung, die diesen Ambitionen entgegenwirken.

Interessanterweise konnte im Rahmen eines transdiagnostischen Untersuchungsansatzes gezeigt werden, dass auch bei gesunden Probanden in Abhängigkeit von bestimmten Ausprägungsmerkmalen eine Verschiebung innerhalb des Systems der Verhaltenskontrolle vorliegen kann. So war ein Ergebnis dieser onlinebasierten Studie mit großer Probandenstichprobe (N = 1413), dass der Schweregrad selbstberichteter pathologischer Phänomene von Essverhalten assoziiert ist mit einer Verminderung von modellbasierten Strategien in der *two-step*-Aufgabe (Gillan et al., 2016). Reduziertes modellbasiertes Verhalten ging außerdem einher mit den transdiagnostischen Symptomdimensionen „Zwanghaftigkeit“ (im Sinne von zwanghaften Verhaltensmustern) und „intrusive Gedanken“ (Gillan et al., 2016).

Die Annahme, dass auch gelernte Verhaltensmuster zur Entstehung und Aufrechterhaltung psychischer Erkrankungen beitragen, erlaubt außerdem einen anderen Blickwinkel in der Behandlung ebenjener Erkrankungen. Dieses Konzept bietet betroffenen Personen die Möglichkeit, individuelle, schädliche Verhaltensmuster zu erkennen und zumindest graduell zu verändern, eine Idee, auf der auch die kognitive Verhaltenstherapie beruht. Das genauere Verständnis von veränderten Lern- und Verhaltensmechanismen bei psychischen Erkrankungen birgt also auch die Möglichkeit zur Entwicklung neuer therapeutischer Ansätze.

Innerhalb dieses Konzepts wird insbesondere ein Prozess als wesentlicher Einflussfaktor für die Entwicklung bestimmter psychischer Erkrankungen angesehen: der Übergang von zielgerichtetem zum habituellen Verhalten bzw. die Entstehung einer Imbalance zwischen diesen beiden Systemen der Verhaltenskontrolle. Anfänglich zeigen z.B. von einer Abhängigkeitserkrankung betroffene Personen häufig Lernverhalten, was durchaus zielgerichtet erscheinen mag, an dem Wert der Konsequenzen orientiert ist und durch entsprechende Stimuli verstärkt wird. Sie lernen beispielsweise auf einen bestimmten Stimulus, z.B. in Form eines unangenehmen Gefühls wie Traurigkeit oder Einsamkeit, auf dem Hintergrund ihrer persönlichen Erfahrungen sowie neurobiologischen und körperlichen Gegebenheiten mit einem entsprechenden Verhalten zu reagieren (z.B. Substanz konsumieren). Dies führt wiederum zu positiven Konsequenzen im Sinne einer Belohnung (z.B. besser schlafen können, sich angenehm berauscht fühlen) bzw. einer Vermeidung beispielsweise negativer Emotionen oder aversiv besetzter Situationen (z.B. weniger Angst und Sorgen haben). Jenes ursprünglich operant verstärkte Verhalten wird zunehmend automatisiert und gewohnheitsmäßig, also habituell ausgeführt trotz verstärkter negativer Konsequenzen wie sozialer Rückzug, finanzieller Folgen oder körperlicher Beeinträchtigungen. Die betroffenen Personen scheinen nicht mehr in der Lage dazu, eine Anpassung ihres anfänglichen Verhaltens (z.B. Alkohol trinken, um sich weniger traurig zu fühlen) herbeizuführen, obwohl dies vermehrt zu einer eigentlich unerwünschten Verhaltenskonsequenz führt. Diese Verschiebung von einer initial zielgerichteten Verhaltenskontrolle hin zu überwiegend habituellen Entscheidungsprozessen kann in diversen psychischen Erkrankungen und insbesondere bei deren Entstehung beobachtet werden (Heinz, 2017; Lee, 2013).

Mit Hinblick auf den großen volkswirtschaftlichen und gesundheitlichen Schaden, den Abhängigkeitserkrankungen und insbesondere die Alkoholabhängigkeit verursachen (Effertz, 2020), ist in den letzten Jahren ein zunehmendes Interesse am Verständnis der gestörten Lern- und Verhaltensmechanismen, die zur Entstehung und Aufrechterhaltung von Alkoholabhängigkeit beitragen, zu verzeichnen. Denn auch Abhängigkeitserkrankungen können als ein Störungsbild mit veränderten Lernvorgängen konzeptualisiert werden (Belin et al., 2009), bei denen im Verlauf der Erkrankung zunehmend habituelle Verhaltensweisen in den Vordergrund rücken. Eine entscheidende Rolle scheinen hierbei zudem Pawlowsche, konditionierte Reize einzunehmen, welche zunehmend automatisierte Handlungsketten auslösen und

somit vor allem zur Aufrechterhaltung des Störungsmusters beitragen. Diese habituellen Verhaltensmechanismen persistieren demnach ungeachtet der eigentlich negativen Konsequenzen, die der weitere Suchtmittelkonsum mit sich bringt (Carter and Tiffany, 1999; Flagel et al., 2009). Das Entscheidungsverhalten wird folglich im Verlauf immer weniger durch die Konsequenzen gesteuert, ist also nicht mehr an das eigentlich antizipierte Ergebnis der Handlung gekoppelt. Vielmehr erfolgt die Steuerung durch spezifische, suchtmittelassoziierte Reize, die eine „automatische“, emotional eher neutrale habituelle Konsumreaktion auslösen (in den Supermarkt gehen, Bier einkaufen, Flasche öffnen). Dabei können ehemals neutrale Umweltreize unterschiedlicher Natur (z.B. Licht einer Kneipe, Supermarktregal mit alkoholischen Getränken, Zigarettengeruch) oder auch interne Reize (z.B. Einsamkeit, Erleben von Stress) durch die Assoziation mit der zunächst angenehmen Wirkung des Suchtmittels zu konditionierten, also suchtmittelassoziierten Schlüsselreizen werden, die zu einer konditionierten Reaktion (Suchtmittelkonsum) führen. Allein die Wahrnehmung dieser Schlüsselreize und die selektive Fokussierung auf ebenjene können Betroffene dazu motivieren, das Suchtmittel zu konsumieren, obwohl möglicherweise eine angenehme Wirkung ausbleibt (Schoenmakers et al., 2010). Daher scheint es nur folgerichtig, dass alkoholabhängige Patienten gegenüber gesunden Probanden in entsprechenden Untersuchungen signifikant größere Annäherungstendenzen in Bezug auf suchtmittelassoziierte Reize zeigen (Wiers et al., 2014). Interessanterweise werden alkoholasoziierte Reize von den Betroffenen selbst häufig als subjektiv aversiv erlebt und können dysphorische Zustände auslösen (Childress et al., 1999). Dies zeigt einmal mehr, inwieweit erlernte Verhaltensmechanismen trotz scheinbarer Motivation zur Abstinenz und Ablehnung gegenüber dem Suchtmittel bereits bei alltäglichen Entscheidungen das Handeln beeinflussen und so zur Aufrechterhaltung der Abhängigkeitserkrankung beitragen können. Es konnte sogar gezeigt werden, dass eine neuronale Aktivierung selbst dann durch suchtmittelassoziierte Reize bei betroffenen Personen hervorgerufen wird, wenn diese unterhalb der bewussten Wahrnehmungsschwelle präsentiert werden (Childress et al., 2008).

Zudem wird angenommen, dass sich bei Personen mit Abhängigkeitserkrankungen eine verminderte Reaktivität gegenüber natürlichen Verstärkern zeigt (Wrase et al., 2007). Folglich kommt es zu einer Verschiebung der Wertigkeit von Belohnungen hinzu suchtmittelassoziierten Verstärker. Potentiell belohnende Situationen wie eine

erfüllende Partnerschaft, soziale Kontakte oder Hobbys rücken somit zunehmend in den Hintergrund und scheinen nur noch einen verminderten motivationalen Charakter zu besitzen. Suchtmittelbezogene Reize gewinnen im Verlauf der Erkrankung zunehmend an Salienz (Konzept der *incentive sensitization*) und werden somit zu einem wichtigen Modulator des Motivations- und Belohnungssystems (Berridge and Robinson, 2016).

Ebenso stellt die individuelle Wirksamkeitserwartung in Bezug auf das Suchtmittel vermutlich einen modulierenden Faktor bei Abhängigkeitserkrankungen dar: So waren in einer Follow-Up-Studie jene alkoholabhängige Patienten häufiger rückfällig, die wenige Tage nach einer Entzugsbehandlung im entsprechenden Fragebogen (AEQ-G) hohe Alkoholwirkungserwartungen angegeben hatten (Sebold et al., 2017). Ein weiterer, zu erwähnender Aspekt ist, dass viele Suchtmittel das dopaminerge Belohnungssystem beeinflussen, welches wiederum bei der Modulation von Lern- und Entscheidungsprozessen, welche sich letztlich auf das individuelle Verhalten auswirken, miteinbezogen wird. Für eine umfassendere Übersicht diesbezüglich ist auf die Arbeiten von Garbusow et al. (2013) sowie Heinz et al. (2012, 2017) zu verweisen.

Letztlich führen all die genannten Veränderungen bei alkoholabhängigen Patienten gewissermaßen zu einer Dysregulation innerhalb des Systems der Verhaltenskontrolle: So konnte mithilfe einer Variante des Devaluationsparadigmas nachgewiesen werden, dass es zu einem Überwiegen habituellen Verhaltens bei gleichzeitiger Abnahme von zielgerichteten Verhaltensstrategien kommt (Sjoerds et al., 2013). Weitere Studien mit alkoholabhängigen Patienten konnten zudem aufzeigen, dass ein Defizit modellbasierter Strategien vorliegt (Sebold et al., 2014; Voon et al., 2015).

Die Erfassung jener Mechanismen, die eine Verlagerung hin zu habituellen bzw. modellfreien Verhaltensmustern bei Abhängigkeitserkrankungen aufzeigen, erfolgte bisher durch verschiedene Paradigmen, ungeachtet dessen, ob wirklich eine Konstruktvalidität zwischen diesen vorliegt. Hinzu kommt, dass Devaluationsparadigmen vor allem tierexperimentell gut untersucht worden sind, jedoch weniger am Menschen, was wohl überwiegend an der Umsetzung des experimentellen Aufbaus liegen mag. Sequentielle Entscheidungsaufgaben hingegen können bei Menschen besser genutzt werden und wurden dementsprechend in Humanstudien gut untersucht.

Vor diesem Hintergrund stellt sich nun die Frage, inwieweit die in der Studie verwendeten Paradigmen diese gewohnheitsmäßigen Verhaltensmuster erfassen und wenn ja, ob sie ähnliche Aspekte dieser Muster abbilden oder ob den theoretischen Konstrukten von habitueller und modellfreier bzw. zielgerichteter und modellbasierter Verhaltenskontrolle letztlich unterschiedliche Mechanismen zugrunde liegen.

Zusammenfassend lässt sich sagen, dass Defizite in zielgerichteten und modellbasierten Strategien der Verhaltenskontrolle zwar möglicherweise ein transdiagnostischer Marker psychischer Erkrankungen darstellen, insbesondere aber bei Abhängigkeitserkrankungen anscheinend eine wichtige Rolle in der Entstehung und Aufrechterhaltung dieser Erkrankung einnehmen. Das Verständnis der zugrundeliegenden dysfunktionalen psychologischen und neuronalen Mechanismen ermöglicht auf diese Weise ein besseres Verständnis der individuellen Ausprägung der Abhängigkeitserkrankung. Gleichzeitig bietet es so die Möglichkeit auch transdiagnostisch geeignete pharmakologische und therapeutische Interventionen in der Behandlung der Alkoholabhängigkeit zu entwickeln und auszubauen.

1.5. Hypothesenbildung

Ziel dieser Arbeit ist es, die Konstruktvalidität zweier etablierter Paradigmen zu überprüfen, die sich konzeptionell aus verschiedenen Bereichen der experimentellen Psychologie entwickelt haben. Es stellt sich die Frage, inwieweit die konzeptionell unterschiedlich gewachsenen Begrifflichkeiten in Hinblick auf ein dichotomes System von Strategien der Verhaltenskontrolle letztlich auf gleiche oder zumindest überlappende theoretische Konstrukte zurückgehen. Bislang wurde zwar konstatiert, dass beide Formen der Operationalisierung dieser Konstrukte ähnliche zugrundeliegende Mechanismen abbilden, dies wurde allerdings bisher noch nie direkt in Studien am Menschen überprüft. Zur Erfassung von zielgerichtetem vs. habituellem Verhalten wurde eine selektive Devaluationsaufgabe eingesetzt. Mittels einer sequentiellen Entscheidungsaufgabe sollten die Anteile modellbasierten und modellfreien Verhaltens erfasst werden.

Folgende Hypothesen sollen hierbei überprüft werden:

- (1) Gesunde Probanden sind zielgerichtet in einer selektiven Devaluationsaufgabe, d.h. sie wählen nach einer Devaluation eines spezifischen Verstärkers den devaluierten im Vergleich zum nicht devaluierten Stimulus weniger häufig aus.
- (2) Gesunde Probanden zeigen eine Mischung aus modellfreier und modellbasierter Verhaltenskontrolle während einer sequentiellen Entscheidungsaufgabe. Folglich beruht ihre Entscheidungsfindung nicht nur auf dem gegenwärtigen Erhalt oder Ausbleiben einer Belohnung, sondern berücksichtigt auch die übergeordnete Aufgabenstruktur.
- (3) Das Ausmaß der zielgerichteten Verhaltenskontrolle in der selektiven Devaluationsaufgabe (Devaluationssensitivität) ist positiv assoziiert mit dem Ausmaß der modellbasierten Verhaltenskontrolle (p -Interaktion) in der sequentiellen Entscheidungsaufgabe.
- (4) Das Ausmaß der zielgerichteten Verhaltenskontrolle in der selektiven Devaluationsaufgabe (Devaluationssensitivität) ist negativ assoziiert mit der modellfreien Verhaltenskontrolle (Haupteffekt der Belohnung) in der sequentiellen Entscheidungsaufgabe.

2. Methoden

Bevor die eigentliche Hauptuntersuchung stattfand, wurde zunächst das Devaluationsparadigma an einer separaten Probandenstichprobe getestet. Ziel dieser Pilotierungsphase war es, mögliche Schwierigkeiten in der Konzeptualisierung sowie dem Aufbau und der Durchführung des Experimentes zu identifizieren, um so entsprechende Anpassungen für die Hauptstudie vorzunehmen zu können. Die in der vorliegenden Arbeit verwendete sequentielle Entscheidungsaufgabe (*two-step*-Aufgabe) war innerhalb der Forschergruppe bereits in früheren Studien (Garbusow et al. 2014, Sebold et al. 2014) sowohl an gesunden als auch an alkoholabhängigen Patienten gut untersucht worden. Folglich handelte es sich um ein etabliertes Paradigma, sodass diesbezüglich keine Pilotierung erfolgte. Sowohl die Rekrutierung, als auch die Durchführung der beiden Studienphasen erfolgten durch dieselbe Versuchsleiterin.

2.1. Datenerhebung

Mit jedem der Teilnehmer wurde zunächst ein telefonisches Screening durchgeführt, welches sich an den Screening-Fragen des SKID-I-Interviews (Wittchen et al., 1997) orientierte. Hierdurch sollten bestehende oder chronische psychiatrische Störungen (mit Ausnahme von Alkoholabhängigkeit in der klinischen Stichprobe) sowie neurologische und schwere körperliche Erkrankungen ausgeschlossen werden. Da für das Devaluationsparadigma Lebensmittel als primäre Verstärker verwendet wurden, waren eine bestehende Lebensmittelunverträglichkeit sowie eine zurückliegende, aktuelle oder geplante Diät ein weiteres Ausschlusskriterium. Linkshänder wurden ebenso ausgeschlossen. Um sicherzustellen, dass keiner der Probanden eine Essstörung aufwies, wurde die deutsche Version des Eating Attitudes Tests (EAT-26) ausgefüllt (Garner et al., 1982). Ein notwendiges Einschlusskriterium für die Teilnahme war, dass die Probanden die im Devaluationsparadigma als natürliche Verstärker verwendeten Getränke und Lebensmittel als angenehm einstufen. Darüber hinaus wurden demographische Daten wie Geschlecht, Schulbildung, Beruf und Substanzmittelkonsum etc. erfasst. Die Studie wurde von der zuständigen Ethikkommission der Charité genehmigt. Die Programmierung der Experimente, welche am Computer durchgeführt wurden, sowie die Erhebung der entsprechenden Rohdaten erfolgte mittels MATLAB Version R2007a.

2.1.1. Pilotstichprobe für die selektive Devaluationsaufgabe

An der Pilotierungsstudie nahmen 21 Probanden teil (Rekrutierung an der Universitätsklinik Charité Mitte, Klinik für Psychiatrie und Psychotherapie, AG Lernen und Kognition). Für die Teilnahme erhielten die Probanden eine feste Aufwandsentschädigung von 10€. Zwei der insgesamt 21 Probanden waren alkoholabhängige Patienten nach einer Entgiftungstherapie, deren letzter Konsumtag max. 21 Tage zurücklag. Da bisherige Untersuchungen durchaus veränderte Mechanismen der Verhaltenskontrolle bei Alkoholabhängigkeit nachweisen konnten, sollte der Einschluss alkoholabhängiger Patienten in die Studie eine Möglichkeit bieten, diese veränderten Verhaltensstrategien mit zu erfassen. Schlussendlich konnten allerdings keine weiteren alkoholabhängigen Patienten mit den entsprechenden Einschlusskriterien für die Studie rekrutiert werden. Grund hierfür war vor allem, dass viele Patienten den Geschmack von Schokolade als nicht ausreichend angenehm ein. Die anfänglichen Bestrebungen, neben gesunden Probanden auch alkoholabhängige Patienten einzuschließen, musste demnach verworfen werden – sowohl für die Pilotstudie, als auch für die Hauptstudie. Daher wurden letztlich nur gesunde Probanden in die Analysen der Pilotstudie eingeschlossen (N = 19).

2.1.2. Probandenstichprobe für Hauptstudie

An der Hauptstudie nahmen 18 gesunde, männliche, rechtshändige Probanden im Alter von durchschnittlich 43.8 Jahren (SD = 8.19) teil. Die Probanden wurden aus der Datenbank des Max-Planck-Instituts für Kognitions- und Neurowissenschaften Leipzig rekrutiert. Für ihre Teilnahme erhielten die Probanden pro Stunde 7€, hinzu kam der erzielte Gewinn aus der sequentiellen Entscheidungsaufgabe.

2.2. allgemeine Versuchsanordnung der Hauptstudie

Zunächst gaben alle Probanden ihre schriftliche Einverständniserklärung zur Teilnahme an der Studie ab. Während in der Pilotstudie lediglich das Devaluationsparadigma durchgeführt wurde, absolvierten die Probanden in der Hauptstudie zudem die *two-step*-Aufgabe und wurden zusätzlich einer neuropsychologischen Testung unterzogen (Abbildung 2).

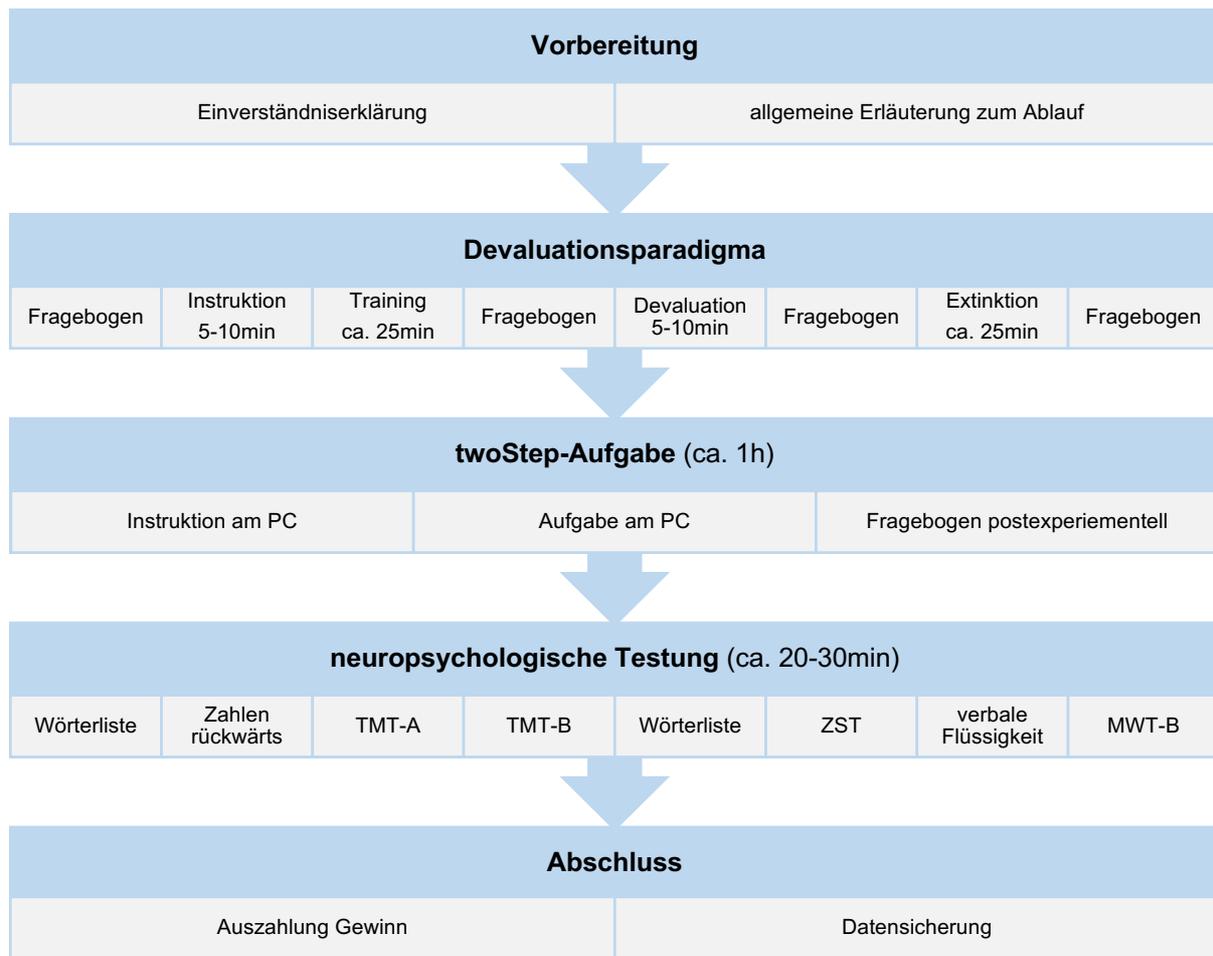


Abbildung 2: allgemeine Versuchsanordnung der Hauptstudie

Bei jedem Probanden erfolgte die Untersuchung gemäß einer festgelegten Versuchsanordnung. Für jede Untersuchung wurde ein Zeitfenster von drei Stunden angesetzt. Zwischen den einzelnen Abschnitten wurde abhängig vom Probanden eine Pause von 5 bis 10 Minuten gemacht (dicke Pfeile).

Hintergrund für die zusätzliche Durchführung einer neuropsychologischen Testung war, dass in Untersuchungen bereits ein Zusammenhang zwischen kognitiven Fähigkeiten und der individuellen Ausprägung der verschiedenen Strategien der Verhaltenskontrolle nachgewiesen werden konnte. Insbesondere das Ausmaß an kognitiver Verarbeitungsgeschwindigkeit scheint einen Einfluss auf die Ausprägung modellbasierter Strategien zu haben. Diesbezüglich konnte vor allem ein Zusammenhang zwischen den Leistungen im Zahlen-Symbol-Test und dem Grad modellbasierten Verhaltens in der *two-step*-Aufgabe belegt werden (Schad et al., 2014; Sebold et al., 2014; Sjoerds et al., 2016).

Für die neuropsychologische Testung wurde folgende Testbatterie verwendet: der Mehrwortschatztest (MWT) zur Messung der verbalen Intelligenz (Schmidt and Metzler, 1992), der Zahlen-Symbol-Test (ZST) als Parameter für die kognitive Verarbeitungsgeschwindigkeit (Wechsler, 1981), Zahlennachsprechen rückwärts zur

Überprüfung des verbalen Arbeitsgedächtnisses bzw. der Merkfähigkeit (Wechsler, 1981), der Trail Making Test (TMT A/B) zur Messung exekutiver Funktionen (Tombaugh, 2004), die Wörterliste zur Überprüfung des verbalen Gedächtnisses sowie der semantische Wortflüssigkeitstest zur Messung der verbalen Flüssigkeit entsprechend der CERAD-Liste (Morris et al., 1989).

2.3. Devaluationsparadigma

Bei dem Experiment handelte es sich um eine modifizierte Version eines etablierten selektiven Devaluationsparadigmas von Valentin et al. (2007). Hiermit sollte ermittelt werden, ob die Probanden eher habituelles oder zielgerichtetes Verhalten zeigen. Als Verstärker erhielten die Probanden in der Hauptstudie Schokoladenmilch und Tomatensaft. Die Entscheidung, Flüssigkeiten als Verstärker für das Experiment zu verwenden, beruhte auf folgenden Überlegungen: Zum einen stellen Nahrungsmittel primäre Verstärker dar, welche einen angeborenen, positiven Belohnungseffekt besitzen. Darüber hinaus sollte eine unmittelbare Belohnung auf das Auswahlverhalten der Probanden erfolgen, mit dem Ziel, dass der Zusammenhang zwischen einem bestimmten Schlüsselreiz bzw. Stimulus und dem darauffolgenden Antwortverhalten (S-R-Assoziation) auch gut gelernt wird. Mithilfe der sensorischen Erfahrung sollte dieser Lernprozess zusätzlich unterstützt werden. Zudem hatte sich in vorherigen Untersuchungen (Valentin et al., 2007) die Verwendung von Flüssigkeiten im fMRT als durchaus praktikabel in der Handhabung erwiesen. Ferner konnte durch dieses Procedere ein direktes, sensorisches Signal gemessen werden, was eine differenziertere Betrachtung der zugrundeliegenden neurobiologischen Prozesse ermöglichte. Auch wenn die vorliegende Studie eine rein verhaltensexperimentelle Untersuchung darstellte, so war es dennoch ein Anliegen, das experimentelle Design so zu gestalten, dass dieses für Folgestudien, auch im fMRT, verwendbar wäre. Eine weitere Überlegung geschah in Hinblick auf die Untersuchung alkoholabhängiger Patienten: So kommt die Verabreichung eines Verstärkers in flüssiger Form der Einnahme des Suchtmittels Alkohol doch sehr nahe.

Eine wichtige Voraussetzung für die erfolgreiche Devaluation in diesem Experiment war ein bestehendes Hungergefühl, da Lebensmittel als natürliche Verstärker gewählt worden waren. Um dies sicher zu stellen, wurden die Probanden gebeten, nüchtern zur Untersuchung zu erscheinen, also sechs Stunden lang vor dem

Experiment weder etwas zu essen, noch koffeinhaltige Getränke oder schwarzen Tee zu sich zu nehmen. Durch Letzteres sollte insbesondere ein möglicher Einfluss von Stimulantien auf die kognitive Leistungsfähigkeit in der neuropsychologischen Untersuchung vermieden werden. Wasser zu trinken wurde den Testpersonen erlaubt. Sie wurden dazu angehalten, ausreichend Wasser vor der Testung zu trinken um einem möglichen Einfluss eines Durstgefühls auf das Entscheidungsverhalten vorzubeugen. Die Uhrzeit, zu der die Untersuchung stattfand, war abhängig davon, ob die Probanden für gewöhnlich ihre erste Mahlzeit am Morgen oder erst später gegen Mittag zu sich nahmen. Jemand, der also für gewöhnlich nicht frühstückte, wurde demnach erst zum Nachmittag bzw. Abend hin einbestellt. So sollte sichergestellt werden, dass die Probanden auch wirklich hungrig waren. Bevor mit dem eigentlichen Experiment begonnen wurde, füllten die Probanden zunächst am Computer Fragebögen aus, bei denen sie mittels visueller Analogskalen angeben sollten, wie hungrig sie sich augenblicklich einschätzten (von „überhaupt kein Hunger“ bis „extrem hungrig“) und wie angenehm sie den Geschmack der später dargebotenen Getränke empfanden (Einschätzung des Anreizwertes mittels visueller Analogskala von „unangenehm“ bis „angenehm“). Die Devaluationsaufgabe bestand aus drei Abschnitten (instrumentelles Lernen, Devaluation und Extinktion), welche im Folgenden beschrieben werden.

2.3.1. instrumentelle Lernaufgabe

Diese Aufgabe beinhaltete drei verschiedene Bedingungen (Schokoladenmilch und Tomatensaft sowie Wasser als neutrale Kondition), welche jeweils 50-mal gezeigt wurden und deren Reihenfolge über das Experiment randomisiert waren (Abbildung 3). Dementsprechend beinhaltete dieser Teil des Paradigmas insgesamt 150 Durchgänge und dauerte ca. 25 Minuten.

Den Probanden wurde mitgeteilt, dass im folgenden Experiment Paare aus abstrakten Bildern gezeigt würden und sie nun immer eines der beiden gezeigten Bilder auszuwählen hätten. Sie wurden weiterhin darüber aufgeklärt, dass sie nach der Auswahl eines Bildes mit einer festgelegten Wahrscheinlichkeit 0,75ml Schokoladenmilch, Tomatensaft, Früchtetee oder Wasser erhalten würden. Die Probanden erhielten jedoch keine Information darüber, welches Bild zur Ausgabe des jeweiligen Getränkes führte. Allerdings bekamen sie den Hinweis, dass es bei jedem Paar immer ein Bild gäbe, welches mit einer höheren Wahrscheinlichkeit zur

Ausgabe eines angenehmen Getränkes, also Schokoladenmilch oder Tomatensaft, führen würde. Die Probanden wurden gebeten herauszufinden, welche Bildauswahl sie treffen müssen um eines der angenehmen Getränke zu erhalten.

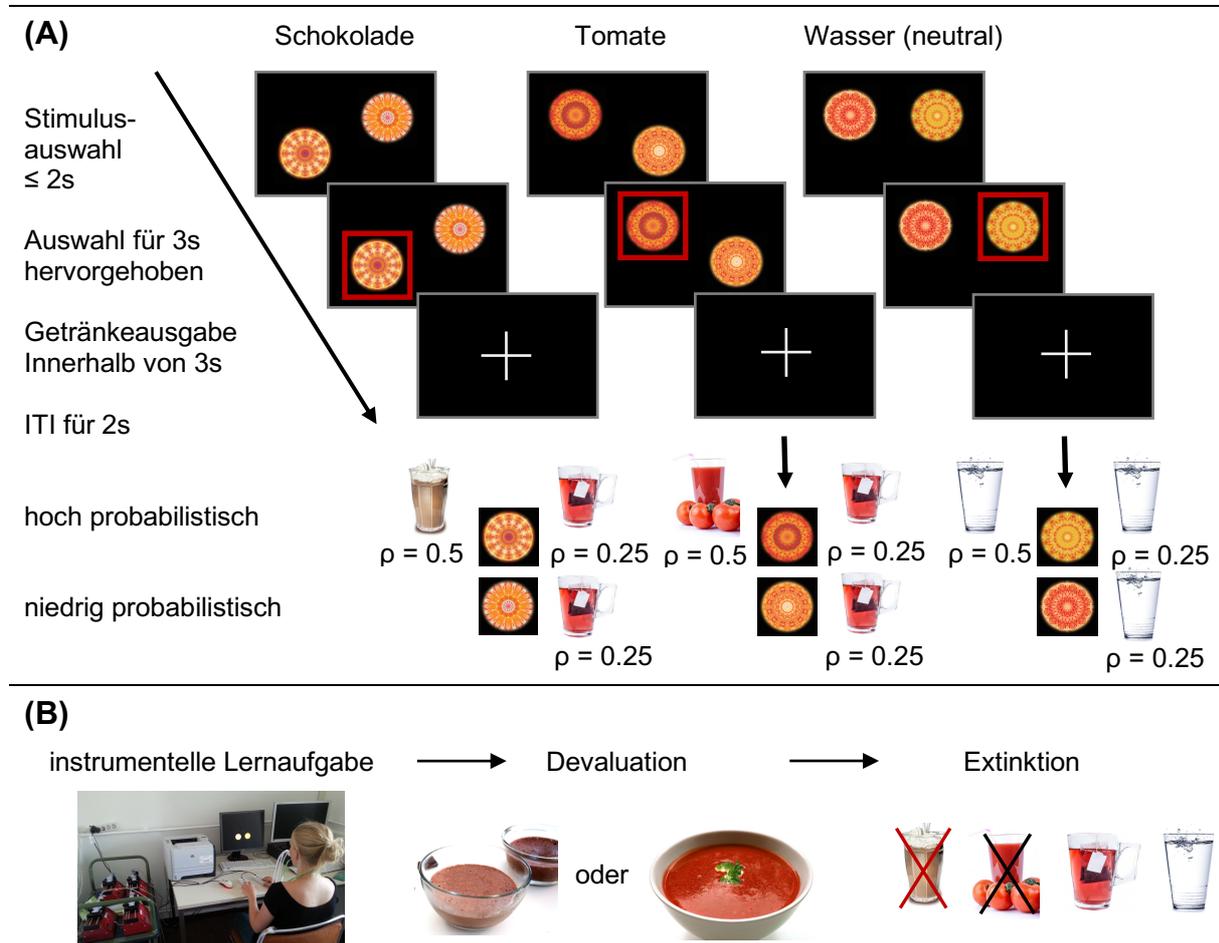


Abbildung 3: Devaluationsparadigma in Anlehnung an Valentin et al. (2007) und Friedel et al. (2014)

(A) Aufgabenstruktur und Belohnungswahrscheinlichkeiten für jede Bedingung, Stimuli und Position wurden über alle Probanden randomisiert

(B) experimentelles Design und Prozedere

Jeder Durchgang startete mit der Präsentation eines Paares aus zwei abstrakten Stimuli. Jedem der drei oben genannten Bedingungen war ein spezifisches Stimuluspaar zugeordnet. Der Proband musste nun innerhalb von zwei Sekunden einen der beiden präsentierten Stimuli auswählen. Die Auswahl der Bilder erfolgte, indem der Proband eine von vier Tasten eines Bedienfeldes entsprechend der Position des Bildes (oben links, oben, rechts, unten links, unten rechts) drückte. Jedem Bildpaar der jeweiligen Bedingung wurde anfänglich randomisiert eine bestimmte Position zugeordnet, welche über das ganze Experiment hinweg konstant blieb. Dies sollte es den Testpersonen erleichtern zu lernen, welches der Bilder zu

welcher Getränkeausgabe führte (Assoziation zwischen Stimulus und Belohnung bzw. Konsequenz). In jeder der drei Bedingungen erhielt der hoch probabilistische Stimulus jeweils eine bestimmte räumliche Zuordnung. Diese Zuordnung wiederum wurde innerhalb der Stichprobe ausbalanciert.

Geschah die Auswahl eines Stimulus nicht innerhalb von zwei Sekunden, so wurde der Proband durch eine Fehlermeldung („zu langsam“) darauf aufmerksam gemacht und der entsprechende Durchgang abgebrochen. Hatte der Proband jedoch innerhalb der zwei Sekunden einen der beiden Stimuli ausgewählt, dann wurde dieser mittels eines roten Rahmens für drei Sekunden hervorgehoben. Dabei unterschieden sich die beiden Stimuli darin, mit welcher Wahrscheinlichkeit es anschließend zur Ausgabe eines Getränkes kam. Die Auswahl eines der beiden Stimuli führte mit einer Wahrscheinlichkeit von $p = 0.75$ zur Getränkeausgabe (hoch probabilistischer Stimulus), wohingegen dies bei dem anderen Stimulus nur zu einer Wahrscheinlichkeit von $p = 0.25$ geschah (niedrig probabilistischer Stimulus). Sowohl in der Schokoladen- als auch in der Tomatenbedingung erhielten die Probanden nach der Wahl des hoch probabilistischen Stimulus mit einer Wahrscheinlichkeit von $p = 0.5$ die entsprechende Belohnung, also Schokoladenmilch oder Tomatensaft, und mit einer Wahrscheinlichkeit von $p = 0.25$ eine Flüssigkeit ohne Verstärkereffekt, in diesem Experiment Früchtetee. Die jeweilige Belohnung und der Früchtetee wurden nie innerhalb des gleichen Durchgangs ausgegeben. Durch die probabilistische Zuordnung sollte ein kontinuierliches Lernen ermöglicht werden. Die Wahl des niedrig probabilistischen Stimulus hingegen war zu keiner Zeit mit einer Belohnung im Sinne von Schokoladenmilch oder Tomatensaft assoziiert, sondern führte lediglich mit einer Wahrscheinlichkeit von $p = 0.25$ zu der gemeinsamen Rückmeldung in Form von Früchtetee.

In den Durchgängen der neutralen Kondition wurde stilles Mineralwasser mit einer Wahrscheinlichkeit von entweder $p = 0.75$ (hoch probabilistischer Stimulus) oder $p = 0.25$ (niedrig probabilistischer Stimulus) dargeboten. Diese neutrale Kondition diente als Kontrollbedingung um die Effekte der jeweiligen Belohnung auf das Auswahlverhalten des Probanden abschätzen zu können.

Je nachdem, welche Auswahl der Proband getroffen hatte, erhielt er daraufhin – probabilistisch entsprechend der jeweiligen Bedingung – entweder 0,75ml eines Getränkes oder gar nichts. Um sicherzustellen, dass die Probanden nicht sehen, sondern lediglich schmecken konnten, welches Getränk sie erhielten, wurde

zwischen Untersuchungstisch und den Pumpen eine spanische Wand installiert bzw. die Pumpen hinter dem Probanden platziert. Nach der Getränkeausgabe (drei Sekunden) wurde für zwei Sekunden ein Fixierkreuz auf dem Bildschirm präsentiert, ehe der nächste Durchgang begann (Inter-Trial-Intervall, ITI). Nachdem die Probanden also gelernt hatten, bevorzugt denjenigen Stimulus eines jeden Paares auszuwählen, welcher mit einer höheren Wahrscheinlichkeit mit einem appetitiven Getränk bzw. Verstärker verbunden war, sollten sie erneut ihren Hunger und die subjektive Wahrnehmung eines jeden dargebotenen Getränkes als angenehm oder unangenehm beurteilen.

2.3.2. Devaluation

Nachdem die Probanden die instrumentelle Lernaufgabe absolviert hatten, wurden sie dazu aufgefordert, entweder Schokoladenpudding oder Tomatensuppe (durchschnittlich 357.1g, SD = 196.0) zu essen, bis einerseits das subjektive Gefühl der Sättigung eingetreten war und die Probanden nicht mehr weiter essen wollten. Zum anderen sollte solange gegessen werden, bis das entsprechende, vormals als Belohnung fungierende, Getränk als nun wenig appetitiver Verstärker eingestuft wurde (verminderter Anreizwert). Dieses selektive Devaluationsprozedere diente dazu, den Wert eines spezifischen, vormals positiv assoziierten, also mit einer Belohnung verknüpften, Stimulus derart zu mindern, dass jener als weniger erstrebenswert empfunden wurde (selektive Devaluation). Hatte sich der Proband beispielsweise an Schokoladenpudding „übersättigt“, war Schokoladenmilch nun ein devaluierter (abgewerteter) Verstärker und sollte als weniger angenehm empfunden werden. Dahingegen sollte das nicht devaluierte Getränk (in diesem Beispiel Tomatensaft) als genauso bzw. nur unweit geringer angenehm empfunden werden. Welches der beiden Lebensmittel für die Devaluation verwendet wurde (Schokoladenpudding oder Tomatensuppe), wurde über die Probandenstichprobe ausbalanciert. Im Anschluss an die Devaluation wurden wieder die entsprechenden Ratings für Hunger und Geschmack der Getränke abgegeben. Zeigte sich hier, dass der Proband noch nicht ausreichend gesättigt war oder das devaluierte Getränk weiterhin als angenehm eingestuft wurde, so wurde der Proband dazu animiert, noch weiter zu essen, bis es zu einer ausreichenden Sättigung kam.

2.3.3. Extinktionstest

Nach der Devaluation sollten die Probanden erneut 50 Durchgänge für jede der drei Bedingungen in randomisierter Form durchlaufen. Dabei wurden ihnen die gleichen abstrakten Stimuluspaare präsentiert, welche im instrumentellen Training mit den verschiedenen Belohnungen assoziiert waren. Abermals mussten sich die Probanden für einen der beiden Stimuli entscheiden, woraufhin der ausgewählte Stimulus mit einem Rahmen hervorgehoben wurde. Allerdings erhielten die Probanden nun im Anschluss daran keine Belohnung mehr, d.h. während der gesamten Extinktionsphase wurden weder Schokoladenmilch noch Tomatensaft verabreicht. Um jedoch ein gewisses Maß an Antwortverhalten aufrechtzuerhalten ohne ein gleichzeitiges Lernen zu provozieren, führte die Auswahl eines Stimulus, unabhängig davon ob es sich um den hoch oder niedrig probabilistischen handelte, sowohl in der Schokoladen- als auch in der Tomatenbedingung mit einer Wahrscheinlichkeit von $p = 0.25$ zu einer Flüssigkeit ohne Verstärkereffekt (Früchtetee). Dies entsprach genau der Wahrscheinlichkeit, mit der der Früchtetee bereits während des instrumentellen Trainings präsentiert wurde. In der neutralen Kondition führte beiden Auswahlmöglichkeiten des spezifischen Stimuluspaares zur Ausgabe von stillem Mineralwasser – ebenfalls mit einer Wahrscheinlichkeit von $p = 0.25$. Dieses Extinktionsprozedere gewährleistete, dass die Probanden Informationen lediglich aus der vorhergehenden instrumentellen Lernaufgabe nutzten, also ihre Entscheidung bei der Stimulusauswahl auf diejenige Assoziation zwischen einem abstrakten Stimulus und der Ausgabe eines bestimmten Getränkes basierten, die sie zuvor gelernt hatten. Andernfalls, wenn also Tomatensaft und Schokoladenmilch weiterhin verabreicht worden wären, hätten die Probanden möglicherweise eine neue Assoziation gelernt. Dies wiederum wäre eine Störvariable in Bezug auf das Nutzen der zuvor erlernten Assoziation zwischen Stimulus und Getränk bzw. Antwortverhalten und Getränk gewesen.

Wie schon durch Valentin et al. (2007) beschrieben, zeichnete sich zielgerichtetes Verhalten dadurch aus, dass es in der Extinktionsphase zu einer Abwertung jenes Verstärkers – und folglich auch des damit assoziierten Stimulus – kam, welcher zuvor in der Devaluationsphase übersättigt worden war. Wurde der Proband beispielsweise mit Schokopudding übersättigt, so vermied er in der Extinktionsphase denjenigen Stimulus, welcher zur Ausgabe von Schokoladenmilch führte (devaluierter Verstärker). Fand also eine selektive Devaluation statt, bevorzugte der Proband in

der devaluierten Bedingung statt dem hoch probabilistischen, den niedrig probabilistischen Stimulus, um so den Erhalt des übersättigen Verstärkers (Schokoladenmilch) zu vermeiden. Kam es hingegen zur Präsentation eines Stimuluspaares der nicht devaluierten Bedingung, so wählte der Proband weiterhin den hoch probabilistischen Stimulus aus, also denjenigen, der zur Ausgabe des nicht devaluierten Verstärkers führte (in diesem Beispiel Tomatensaft). Kam es jedoch nicht zu einer Änderung im Auswahlverhalten, wurde also auch in der devaluierten Bedingung der hoch probabilistische Stimulus annähernd genauso häufig wie im vorhergehenden Training ausgewählt, sprach dies für eine habituelle Strategie der Verhaltenskontrolle.

Nach Abschluss dieser Phase des Experimentes gaben die Probanden letztmalig ihre Ratings für Hunger und Geschmack der Getränke ab. Ferner wurde nach den Strategien gefragt, welche sowohl während der instrumentellen Lernaufgabe als auch während der Extinktion verfolgt worden waren.

2.3.4. Pilotierung des selektiven Devaluationsparadigmas

In der Pilotierungsphase sollte zum einen erörtert werden, wie gut die Zuordnung zwischen einem Stimulus und dem dazugehörigen Getränk gelernt wurde. Zum anderen war die Frage, inwieweit die Probanden mit der Versuchsanordnung, also sowohl mit den Instruktionen als auch mit der Ausgabe der Getränke über Plastikschräuche, zurechtkommen würden. Die im Experiment verwendeten abstrakten Bilder wurden im Vorfeld der Pilotierungsphase in einer Stichprobe von 50 Personen auf gleiche Werte in Bezug auf Validität und Arousal aus einer Auswahl von zwölf Bildern geprüft. In der Pilotierungsphase waren in Anlehnung an Schwabe und Wolf (2009) zunächst Schokoladenmilch und Orangensaft als Belohnung bzw. appetitive Verstärker verabreicht worden. Allerdings ergab sich dann als Resultat der statistischen Auswertung der Pilotdaten, dass die Verwendung von Orangensaft unter mehreren Gesichtspunkten, welche im Ergebnisteil ausführlicher erläutert werden, als nicht geeignet erschien. So wurde in der Hauptstudie letztlich Tomaten- statt Orangensaft eingesetzt. Zudem mussten die Probanden der Pilotstudie die erwähnten Fragebögen mit visueller Analogskala zu Hungergefühl und Anreizwert der Lebensmittel händisch ausfüllen. Diese wurden später in der Hauptstudie in das Computereperiment integriert.

2.4. two-step-Aufgabe

Im Anschluss an das Devaluationsparadigma absolvierten die Probanden eine Aufgabe zur sequentiellen Entscheidungsfindung die sog. *two-step*-Aufgabe (Abbildung 4), adaptiert nach Daw et al. (2011). Vor Beginn des Experimentes wurden die Probanden eingehend bezüglich der Aufgabenstruktur instruiert, wobei eine möglichst exakte deutsche Übersetzung der englischen Instruktionen von Daw et al. (2011) genutzt wurde.

Die Aufgabe bestand darin, durch zwei aufeinanderfolgende Entscheidungen zwischen zwei abstrakten Reizen (geometrische Figuren) einen möglichst großen Geldgewinn zu erzielen. Jeder Durchgang bestand demnach aus zwei Ebenen (daher der Name *two-step*). Auf der ersten Stufe mussten sich die Probanden zwischen zwei verschiedenen Stimuli entscheiden, woraufhin ihnen eines von zwei weiteren möglichen Stimuluspaaren präsentiert wurde. Das Erscheinen des Stimuluspaares auf der zweiten Ebene war mit einer definierten Wahrscheinlichkeit an die Auswahl auf erster Ebene geknüpft: Jede der zwei Optionen auf der ersten Ebene führte mit einer Wahrscheinlichkeit von 70% zu einem bestimmten von zwei weiteren Stimuluspaaren (so genannte häufige Transition) und mit einer Wahrscheinlichkeit von 30% zu dem anderen der zwei beiden Stimuluspaare (seltene Transition). Die Wahrscheinlichkeitsverteilung für die Transition wurde den Probanden mitgeteilt. Des Weiteren wurde ihnen erläutert, dass dieser Zusammenhang zwischen einem bestimmten Stimulus auf erster Ebene und dem darauf häufiger folgenden Stimuluspaar auf zweiter Ebene immer gleichbleibe. Hierbei könne es jedoch zufällig passieren, dass die Stimulusauswahl auf erster Ebene nicht immer zu diesem bestimmten Stimuluspaar auf zweiter Ebene führe, sondern in seltenen Fällen auch zu einem anderen Stimuluspaar. Auf der zweiten Ebene sollten die Probanden wiederum eine Entscheidung treffen bezüglich der gezeigten Stimuli, wobei diesmal auf die Auswahl ein Geldgewinn folgte oder auch nicht. Der Geldgewinn wurde in Form einer 20-Cent-Münze auf dem Bildschirm angezeigt. Den Probanden wurde mitgeteilt, dass sie den gesamten Gewinn nach Abschluss des Experimentes ausgezahlt bekommen würden. Welches der letztlich vier möglichen Stimuli auf der zweiten Ebene häufiger zu einem Geldgewinn führte, veränderte sich im Verlauf des Experimentes langsam um einen anhaltenden Lernprozess zu ermöglichen. Das bedeutet, dass die Wahrscheinlichkeit, mit der die Auswahl eines bestimmten farbigen Stimulus auf der zweiten Ebene zu einer

Gewinnauszahlung führte, sich im Verlauf langsam ändern würde. Diesbezüglich wurde den Probanden folgendes Beispiel erläutert: Ein Stimulus auf zweiter Ebene, welcher anfänglich häufiger zu einer Geldauszahlung führte, würde im Verlauf möglicherweise weniger Gewinn einbringen. Stattdessen würde dann eine andere Stimulusauswahl nun häufiger in einer Gewinnauszahlung resultieren. Die Wahrscheinlichkeit für eine Gewinnauszahlung schwankte für jeden Stimulus auf zweiter Ebene dementsprechend zwischen 25% und 75%. Diese Veränderung der Wahrscheinlichkeit wird auch als Gauß'scher Random Walk bezeichnet, welcher exemplarisch in Abbildung 4 zu sehen ist.

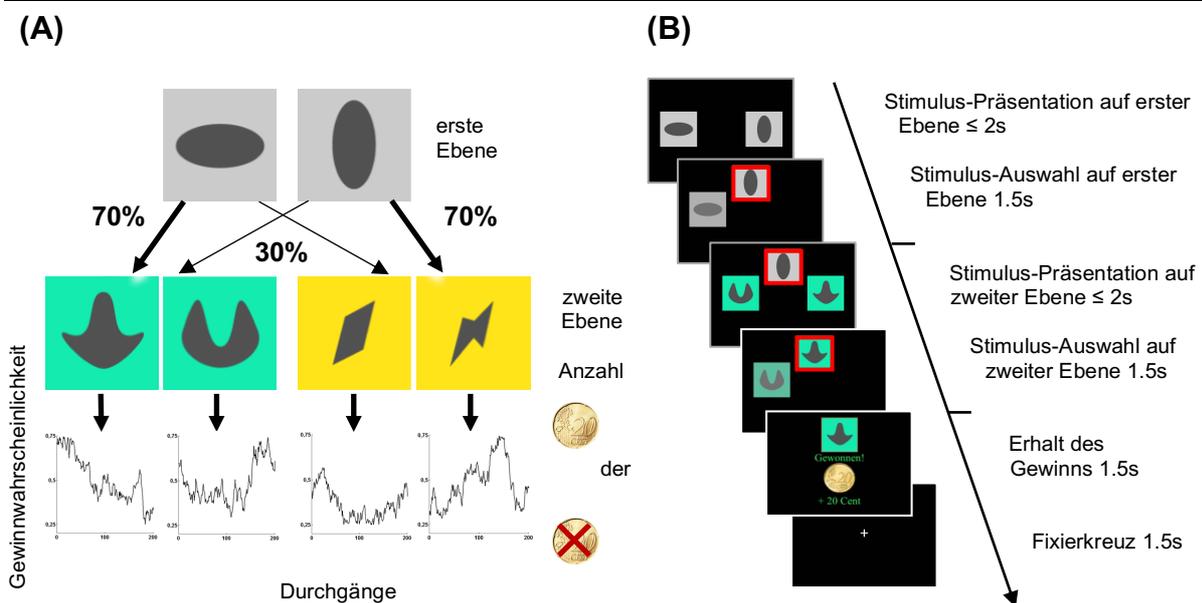


Abbildung 4: two-step-Aufgabe in Anlehnung an Daw et al. (2011) und Friedel et al. (2014)

(A) Darstellung der Transitionsstruktur (70% häufige Transition und 30% seltene Transition) und der sich langsam verändernden Gewinnwahrscheinlichkeiten (Gauß'scher Random Walk)

(B) Ablauf der two-step-Aufgabe am Computer

Ziel der Probanden sollte es sein, einen möglichst hohen Gewinn über den gesamten Verlauf des Experimentes zu erzielen. Die Gewinnausschüttung lag bei mindestens drei Euro (auch wenn der Proband größtenteils Verluste erzielt). Hinzu kamen die in der Aufgabe erzielten Gewinne. Die maximale Gewinnauszahlung war begrenzt auf acht Euro. Die eigentliche Aufgabe bestand aus insgesamt 201 Durchgängen, welche jeweils mit einer Pause von ein bis sieben Sekunden voneinander getrennt waren (exponentiell verteiltes inter-trial-Intervall, ITI). Sowohl auf der ersten als auch auf der zweiten Ebene hatten die Probanden zwei Sekunden Zeit, sich für eine der beiden gezeigten Stimuli zu entscheiden. Schafften sie es nicht, in der gegebenen

Zeit eine Auswahl zu treffen, erschien auf dem Bildschirm der Hinweis „zu langsam“, woraufhin der Durchgang abgebrochen wurde. Die Position der Stimuli (links oder rechts) auf der ersten und zweiten Ebenen waren von Durchgang zu Durchgang randomisiert.

Der wesentliche Indikator dafür, ob ein Proband bei dieser Aufgabe eher modellbasiertes oder modellfreies Verhalten aufwies, war vor allem die Auswahl des Stimulus auf erster Ebene, wobei insbesondere das Verhalten bei ausbleibendem Gewinn von besonderem Interesse ist. Hierin besteht die eigentliche Rationale der Entscheidungsaufgabe: In jenen Fällen, bei denen eine Entscheidung auf erster Ebene, die bisher häufig zu einem bestimmten Stimuluspaar auf zweiter Ebene führte, überraschenderweise zur Präsentation des anderen Stimuluspaares führt, zeigte sich, ob der Proband die Aufgabenstruktur der Transitionsmatrix benutzte. Nehmen wir beispielsweise an, dass der Proband im vorherigen Durchgang für einen bestimmten Stimulus auf der zweiten Ebene keinen Gewinn erhalten hatte. In den Entscheidungsprozess für die Stimulusauswahl im nächsten Durchgang würde er die Überlegung mit einbeziehen, ob der zuvor gewählte Stimulus auf der zweiten Ebene demjenigen Stimuluspaar zugehörig war, welches der getroffenen Auswahl auf erster Ebene häufiger (häufige Transition) oder seltener (seltene Transition) folgte. War der zuvor gewählte Stimulus Folge einer seltenen Transition und wurde auch nicht belohnt, dann würde der modellbasierte Lerner trotz des Misserfolges seine Auswahl auf erster Ebene wiederholen, da diese in 70% zu dem Stimuluspaar mit höherer Belohnungswahrscheinlichkeit führte. Führte die seltene Transition hingegen zu einer Belohnung, so würde der Proband trotz des Gewinnes sein Verhalten anpassen und seine Auswahl auf erster Ebene wechseln. Er würde also diejenige Option auf erster Ebene auswählen, welche häufiger zu dem Stimuluspaar führte, welches belohnt wurde. Ein Proband mit einer modellbasierten Strategie würde demnach in seinen Entscheidungsprozess nicht nur den Erhalt oder das Ausbleiben einer Belohnung im vorherigen Durchgang einbeziehen, sondern auch die Aufgabenstruktur und die damit einhergehenden Auftretenswahrscheinlichkeiten der Stimuluspaare von der ersten auf die zweite Ebene (Transitionsmatrix). Er hatte also ein „Modell“ davon, wie die Aufgabe aufgebaut war, ähnlich einem Entscheidungsbaum.

Eine modellfreie Strategie hingegen war dadurch charakterisiert, dass das Entscheidungsverhalten auf der ersten Ebene rein von vorangegangenen Gewinnen oder Verlusten abhing, ein Proband somit ungeachtet der

Transitionswahrscheinlichkeiten handelte. Folglich würde eine Person mit modellfreiem Entscheidungsverhalten nach Belohnung einer im vorherigen Durchgang getroffenen Auswahl beim nächsten Durchgang auf erster Ebene wieder denjenigen Stimulus wählen, für welchen er sich zuvor entschieden hatte. War sein Entscheidungsverhalten im vorherigen Durchgang nicht belohnt worden, so würde er auf erster Ebene seine Auswahl wechseln. Bei dieser Strategie agierte der Proband also unabhängig von der eigentlichen Transitionsmatrix und orientierte sein Verhalten ausschließlich an dem Erhalt oder Ausbleiben der Belohnung auf zweiter Ebene. Belohnungen führten also zur Wiederholung der vorherigen Auswahl und Misserfolge zu einem Wechsel der Auswahl auf erster Ebene. Das bedeutet, auch wenn das auf zweiter Ebene präsentierte Stimuluspaar eigentlich Folge einer seltenen Transition war, würde der Proband seine Auswahl auf erster Ebene wiederholen, wenn er zuvor belohnt worden war.

2.5. Statistische Analysen

2.5.1. Analyse der Verhaltensdaten und Ratings des Devaluationsparadigmas

Zunächst sollte mit einem Einstichproben-t-Test überprüft werden, ob das Lernkriterium erreicht wurde, die Probanden also im Verlauf des instrumentellen Trainings lernten, signifikant häufiger den mit einer höheren Wahrscheinlichkeit belohnten Stimulus auszuwählen. Hierzu wurde das Auswahlverhalten für jede einzelne Bedingung in letzten zehn Durchgänge der Trainingsphase betrachtet und ermittelt, ob sich die Auswahlhäufigkeit des hoch probabilistischen Stimulus gegenüber einer Wahrscheinlichkeit von 50%, also einer zufälligen Auswahl, signifikant unterschied (Testung gegen den Wert 50).

Um einschätzen zu können, in welchem Ausmaß die Probanden zielgerichtetes oder habituelles Verhalten nach der Devaluation zeigten, wurde die Anzahl der getroffenen Auswahlmöglichkeiten mittels einer 2 (Zeit: vor/nach Devaluation) x 2 (Wert des ausgewählten Stimulus: nicht devaluiert/ devaluiert) Varianzanalyse mit Messwiederholungen (ANOVA) ausgewertet. Diese Analysen wurden sowohl für die Pilotstudie als auch für die Hauptstudie durchgeführt. Um die Quellen eines möglicherweise signifikanten Interaktionseffektes zu überprüfen, sollten zudem post-hoc-Tests durchgeführt werden.

Darüber hinaus wurde mittels t-Test für abhängige Stichproben überprüft, ob es nach der Devaluation zu einer signifikanten Änderung in Bezug auf das Hungergefühl und den Anreizwert der dargebotenen Getränke bzw. Lebensmittel kam.

In der Pilotstudie musste ein Proband von den Analysen ausgeschlossen werden, da er das Lernkriterium im letzten Block der instrumentellen Lernaufgabe nicht erfüllt hatte (N = 18 für Pilotstudie). Auch in der Hauptstudie erreichten zwei Probanden das Lernkriterium nicht. Bei zwei weiteren Teilnehmern konnte kein ausreichender Sättigungseffekt erzielt werden, sodass letztlich 14 Probanden in die Analyse der Hauptstudie eingeschlossen wurden.

2.5.2. Analyse der Verhaltensdaten der *two-step*-Aufgabe

Anhand des Entscheidungsverhaltens im *two-step*-Paradigma sollte zunächst auf Grundlage von Rohdatenanalysen des Entscheidungsverhaltens auf der ersten Ebene ermittelt werden, ob ein Proband eher eine modellbasierte oder modellfreie Strategie verfolgte. Das Entscheidungsverhalten auf der ersten Ebene, d.h. ob derselbe Stimulus im nächsten Durchgang wieder ausgewählt oder ob zum andern Stimulus gewechselt wird, lässt sich folglich als eine Funktion der Belohnung (Gewinn/ kein Gewinn) und der Transitionsbedingung (häufige Transition/ seltene Transition) darstellen. Mit Hinblick auf diese Überlegungen wurden für diese Analyse der *two-step*-Aufgabe für jeden Probanden individuelle Scores ermittelt, welche angaben, wie hoch die Wahrscheinlichkeit war, dass der Proband sein Entscheidungsverhalten auf erster Ebene wiederholte. Modellfreie Strategien basieren lediglich auf dem Erhalt oder Ausbleiben einer Belohnung im vorherigen Durchgang, eine Entscheidung würde also ungeachtet der Transitionsmatrix wiederholt werden, sodass die Wiederholungswahrscheinlichkeiten sowohl für die seltene als auch für die häufige Transition annähernd gleich hoch wären. Folglich würde sich nur ein Unterschied in Hinblick auf den erhaltenden oder ausbleibenden Gewinn ergeben. Im Gegensatz dazu berücksichtigt eine modellbasierte Strategie die Interaktion zwischen Belohnung und Transitionsbedingung. Eine hypothetische Darstellung eines Probanden mit rein modellfreier bzw. rein modellbasierter Strategie ist in Abbildung 5 zu sehen. Die individuellen Wiederholungswahrscheinlichkeiten wurden einer Varianzanalyse mit Messwiederholungen (ANOVA) unterzogen, wobei die Belohnung (Gewinn/ kein Gewinn) und die Transitionsbedingung (häufige Transition/ seltene Transition) als *within-subject*-Faktoren einfließen.

Bei einem der Probanden kam es nach der Hälfte der Durchgänge zu einem Programmabbruch, sodass dieser von den Berechnungen ausgeschlossen werden musste (N = 17).

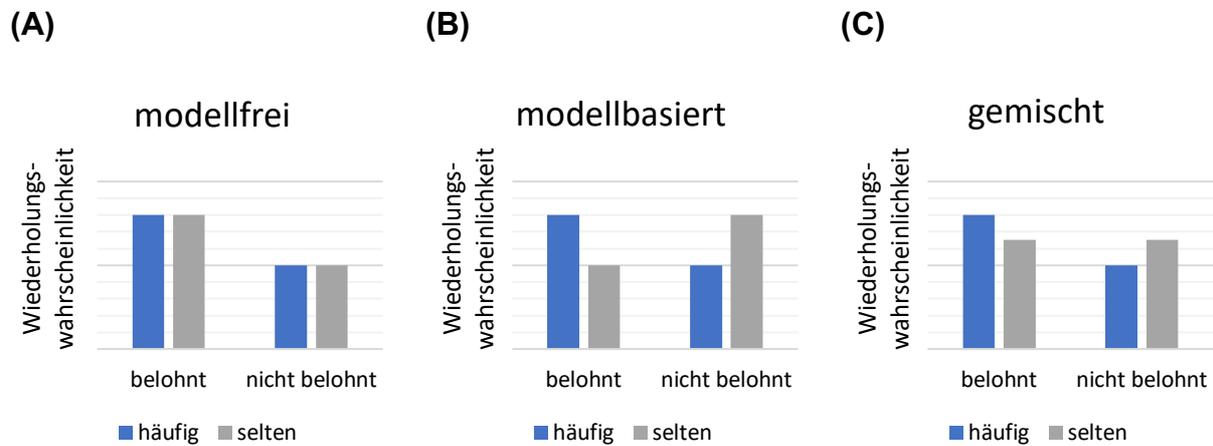


Abbildung 5: Beispiele für Ergebnisdiagramme bei der two-step-Aufgabe

(A) Ein modellfreier Lerner würde sich lediglich an einer erhaltenen Belohnung orientieren und seine vorherige Auswahl wiederholen, ohne Rücksicht auf die Transitionsstruktur. (B) Ein Proband mit modellbasierter Strategie hingegen berücksichtigt auch die Transitionsstruktur, sodass er bei einer Belohnung als Folge einer seltenen Transition seine Auswahl auf erster Ebene wechseln würde (erkennbar an dem niedrigeren grauen Balken). (C) Gesunde Probanden nutzen oftmals eine Mischung aus beiden Strategien, was sich am Diagramm im Unterschied der Wiederholungswahrscheinlichkeit sowohl zwischen, als auch innerhalb der verschiedenen Transitionsbedingungen (häufig/ selten) ablesen lässt.

2.5.3. Korrelationsanalyse zwischen den zwei Paradigmen

Um für jeden Probanden das individuelle Ausmaß der genutzten Strategien der Verhaltenskontrolle zu bestimmen, wurden jeweils Interaktionsterme ermittelt, welche sich auf die Anzahl der getroffenen Entscheidungen bezogen. Für die selektive Devaluationsaufgabe wurde der Interaktionsterm folgendermaßen ermittelt: $(\text{Auswahlhäufigkeit des devaluierten Stimulus vor der Devaluation} - \text{Auswahlhäufigkeit des devaluierten Stimulus nach der Devaluation}) - (\text{Auswahlhäufigkeit des nicht devaluierten Stimulus vor der Devaluation} - \text{Auswahlhäufigkeit des nicht devaluierten Stimulus nach der Devaluation})$. Ein hoher Wert bedeutete ein höheres Ausmaß an zielgerichtetem Verhalten. Das heißt, dass der Proband nach der Devaluation den nicht devaluierten Stimulus gegenüber dem devaluierten Stimulus bevorzugte bzw. häufiger auswählte. Habituelles Verhalten hingegen zeichnete sich bei diesem Paradigma dadurch aus, dass derjenige Stimulus, der mit dem devaluierten Getränk assoziiert war, nach der Devaluation

weiterhin genauso häufig wie zuvor ausgewählt wurde, obwohl er nicht mehr mit positiven Konsequenzen verknüpft war. Entsprechend hätte sich dieses habituelle Verhaltensmuster in einem niedrigen Interaktionsterm widergespiegelt. Der Interaktionsterm des Devaluationsparadigmas gab demnach Auskunft darüber, wie sensitiv ein Proband gegenüber dem Devaluationsprozedere war. Im Folgenden wird daher der Begriff der Devaluationssensitivität als Synonym für den individuellen Interaktionsterm der Devaluationsaufgabe verwendet.

Ein ähnlicher Interaktionsterm wurde auch für die *two-step*-Aufgabe errechnet (p -Interaktion), wobei hier die Wahrscheinlichkeit, einen zuvor gewählten Stimulus auf erster Ebene erneut auszuwählen, herangezogen wurde: (*Auswahlhäufigkeit eines belohnten Stimulus nach einer häufigen Transition* – *Auswahlhäufigkeit eines belohnten Stimulus nach einer seltenen Transition*) – (*Auswahlhäufigkeit eines unbelohnten Stimulus nach einer häufigen Transition* – *Auswahlhäufigkeit eines unbelohnten Stimulus nach einer seltenen Transition*). Ein Proband mit überwiegend modellbasierter Strategie wiederholte demnach auf erster Ebene seine Auswahl, verblieb also bei seiner Entscheidung, wenn er zuvor einen Gewinn nach einer häufigen Transition bzw. keinen Gewinn nach einer seltenen Transition erfahren hatte (Interaktion Belohnung x Transition). Der beschriebene Interaktionsterm (p -Interaktion) entsprach also dem individuellen Ausmaß an modellbasiertem Verhalten. Ein Maß für rein modellfreies Verhalten pro Proband, also eine ausschließliche Orientierung am erhaltenen oder ausbleibenden Gewinn unabhängig von der Transition, konnte durch das Ausmaß (*Auswahlhäufigkeit eines belohnten Stimulus* – *Auswahlhäufigkeit eines nicht belohnten Stimulus*) beschrieben werden (Haupteffekt der Belohnung, p -Belohnung).

Um den vermuteten, positiven Zusammenhang zwischen modellfreier (p -Belohnung) und habitueller (geringe Devaluationssensitivität) sowie modellbasierter (p -Interaktion) und zielgerichteter Strategie der Verhaltenskontrolle (hohe Devaluationssensitivität) zu überprüfen, wurde schließlich ein Korrelationstest durchgeführt. Hierbei flossen für jeden Probanden die genannten Interaktionsterme beider Paradigmen mit ein ($N = 13$). Aufgrund der kleinen Fallzahl und der Annahme, dass die ermittelten Parameter nicht einer Normalverteilung entsprechen, verwendeten wir den Spearman-Korrelationskoeffizienten. Außerdem zeigt sich dieser Korrelationskoeffizient robuster gegenüber Ausreißern. Die Analysen wurde mittels IBM SPSS Statistics 26.

2.5.4. Korrelationsanalysen zwischen den Ergebnissen der neuropsychologischen Testung und den Interaktionstermen der Paradigmen

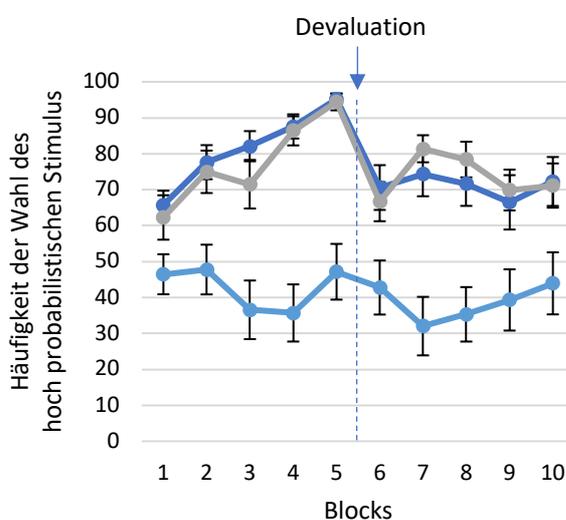
Um den – durch frühere Studien bereits nachgewiesenen – Zusammenhang zwischen kognitiven Fähigkeiten und dem individuellen Ausmaß der verschiedenen Strategien der Verhaltenskontrolle zu überprüfen, erfolgten auch hier Korrelationsanalysen (*two-step*: N = 17, Devaluation: N = 14). Bei diesen Analysen wurde ebenfalls der Spearman-Korrelationskoeffizient angewandt. Da insgesamt sechs verschiedene Parameter aus der neuropsychologischen Testung betrachtet wurden, erfolgte zudem eine Korrektur für multiples Testen (Bonferroni-Korrektur).

3. Ergebnisse

3.1. Pilotstudie zur Etablierung des selektiven Devaluationsparadigmas

Im Gegensatz zur Hauptstudie wurden in der Pilotierungsphase zunächst Schokoladenmilch und Orangensaft als appetitive Verstärker verabreicht (N = 18 nach Ausschluss der zwei alkoholabhängigen Probanden und eines Probanden aufgrund des nicht erfüllten Lernkriteriums). In der Auswertung mittels Einstichproben-t-Test zeigte sich, dass die Testpersonen im Verlauf des instrumentellen Trainings sowohl für die devaluierte als auch die nicht devaluierte Bedingung gelernt hatten, den hoch probabilistischen Stimulus signifikant häufiger auszuwählen (Abbildung 6: devaluiert: $T = 29.750$, $p < 0,001$; nicht devaluiert: $T = 18.297$, $p < 0,001$), wobei dies nicht für die neutrale Kondition mit Wasser galt (Abbildung 6: $T = 0.356$, $p = 0.726$). Das Lernverhalten zeigte sich insgesamt über das Training hinweg für die beiden entscheidenden Bedingungen (nicht devaluiert/ devaluiert) auf einem ähnlichen Niveau (siehe Lernkurven Abbildung 6).

(A) Pilotstudie



(B) Hauptstudie

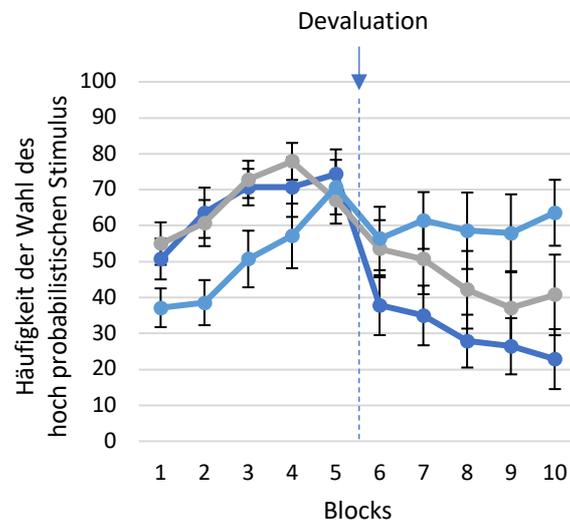


Abbildung 6: Auswahlverhalten für den hoch probabilistischen Stimulus

(A) Die Probanden (N = 18) lernten in der Pilotstudie für die devaluierte und nicht devaluierte Bedingung signifikant häufiger den hoch probabilistischen Stimulus auszuwählen (Anstieg Lernkurve), nicht jedoch für die neutrale Bedingung. Nach der Devaluation zeigte sich eine signifikante Abnahme in der Wahl des hoch probabilistischen Stimulus sowohl in der devaluierten als auch in der nicht devaluierten Bedingung, allerdings nicht für die neutrale Bedingung. (B) In der Hauptstudie (N = 14) zeigte sich eine etwas schlechtere Lernrate, dafür lernten die Probanden für alle drei Bedingungen annähernd gleich gut. Nach der Devaluation zeigte sich eine signifikante Abnahme der Wahl des hoch probabilistischen Stimulus in der devaluierten Bedingung, nicht jedoch in den anderen beiden Bedingungen (in Anlehnung an Friedel et al., 2014).

Mittels t-test für abhängige Stichproben konnte ermittelt werden, dass die Probanden nach der Devaluation ein deutlich vermindertes Hungergefühl verspürten (Abbildung 7: $T = 3.510$, $p = 0.003$). Zudem stuften sie das devaluierte Lebensmittel als signifikant weniger angenehm ein (Abbildung 7: $T = 3.115$, $p = 0.006$). Der Anreizwert der anderen dargebotenen appetitiven Verstärker nahm im Gegensatz dazu nicht ab (Abbildung 7: nicht devaluiert: $T = 1.298$, $p = 0.212$; Wasser: $T = 0.276$, $p = 0.766$), was für eine erfolgreiche, selektive Devaluation gesprochen hätte.

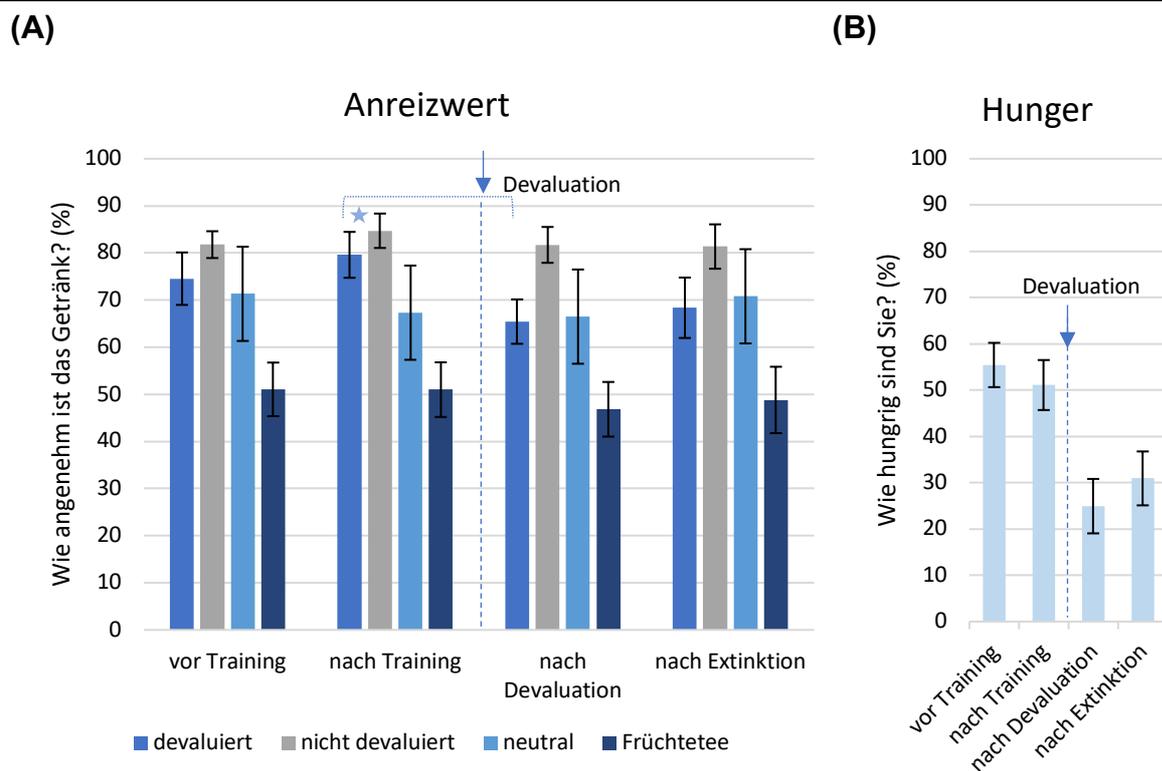


Abbildung 7: Rating für Hunger und Anreizwert beim Devaluationsparadigma in der Pilotstudie

In der Pilotstudie ($N = 18$) kam es nach der Devaluation zu einer signifikanten Abnahme des Anreizwertes des devaluierten Getränkes (A, hier mit einem Sternchen markiert) und einer signifikanten Verminderung des Hungergefühls (B). Für das Rating nach Extinktion gilt $N = 16$, da von zwei Probanden die entsprechenden Datenpunkte nicht erfasst wurden.

Diese Änderung der Anreizwerte spiegelte sich allerdings nicht in dem instrumentellen Verhalten nach der Devaluation wider. In den Varianzanalysen mit Messwiederholungen (ANOVA) zeigte sich bisweilen keine signifikante Interaktion zwischen den Parametern Zeit und Bedingung ($F < 1$, $p = 0.497$). Allerdings ergab sich in Bezug auf den Parameter Zeit über beide Bedingungen hinweg (devaluiert/nicht devaluiert) eine signifikante Abnahme des Auswahlverhaltens des hoch

probabilistischen Stimulus nach der Devaluation im Vergleich zum Ende des instrumentellen Trainings ($F(17) = 21.707, p < 0.001$). In den post-hoc-Tests zeigte sich, dass dieser Effekt sowohl auf einer signifikanten Abnahme innerhalb der devaluierten ($T = 3.756, p = 0.002$) als auch der nicht devaluierten Bedingung beruhte ($T = 5.000, p < 0,001$). Unabhängig davon also, mit welchem Lebensmittel zuvor übersättigt worden war, vermieden die Probanden demnach in der Extinktionsphase sowohl diejenigen Bilder, die zur Ausgabe von Schokoladenmilch führten, als auch jene, die die Ausgabe von Orangensaft zur Folge hatten.

3.2. Bewertung der Pilotdaten und daraus abgeleitete Modifikationen des Devaluationsparadigmas

In Zusammenschau der Ergebnisse zeigte sich, dass keine selektive Devaluation stattgefunden hatte. Die Ursache dafür lag möglicherweise in der sensorischen Ähnlichkeit der verwendeten Lebensmittel, weshalb kein spezifischer Sättigungseffekt (zumindest im Verhalten) erzielt werden konnte. Wir entschieden uns daher letztlich für die Verwendung von Tomaten- anstatt Orangensaft in der Hauptstudie um eine größtmögliche Unterscheidbarkeit in Geschmack und Konsistenz zu erzielen und so einen sensorisch spezifischen Sättigungseffekt zu erhalten. Tomatensaft wiederum erwies sich als schwierig in der Rekrutierung von Probanden, da solche gefunden werden mussten, die sowohl Tomatensaft als auch Schokoladenmilch als angenehm empfanden. Zur weiteren Verminderung von Heterogenität in der Stichprobe wurden zudem nur männliche Probanden rekrutiert, da Studien zeigen, dass Frauen insgesamt häufiger ein restriktives Essverhalten zeigen (Wardle et al., 2004), auch wenn sie derzeit keine Diät machen. Ein Ausschluss in Hinblick auf den BMI erfolgte nicht, er wurde aber in der Hauptstudie miterfasst.

In Bezug auf die Hungerratings fiel auf, dass diese vor und nach dem Training innerhalb der Probandenstichprobe divergierten. Somit schien bei einigen Probanden allein durch das Training der Hunger gesättigt, trotz der vorherigen Nahrungskarenz von mindestens sechs Stunden. Daher legten wir für die Hauptstudie folgende Einschlussbedingungen fest: Das Hungerlevel sollte initial hoch liegen und sich vor und nach dem Training nicht wesentlich unterscheiden. Zudem sollte es nach der Devaluation zu einem deutlichen Abfall des Hungerlevels kommen.

3.3. Devaluationsparadigma

In Hinblick auf das Devaluationsparadigma mussten in der Hauptstudie letztlich vier Probanden ausgeschlossen werden. Zwei der Probanden erfüllten das Lernkriterium nicht und zwei weitere Personen weigerten sich mehr zu essen, obwohl sie bei den Ratings das devaluierte Lebensmittel nicht als ausreichend weniger angenehm angaben.

3.3.1. Trainingsphase/ instrumentelle Lernaufgabe

Im Verlauf des Trainings trafen die Probanden (N = 14) sukzessiv häufiger ihre Auswahl zugunsten desjenigen Stimulus, welcher mit einer höheren Wahrscheinlichkeit dazu führte, eines der angenehmen Getränke zu erhalten. Dies zeigte sich im Einstichproben-t-Test daran, dass die Probanden in den letzten zehn Durchgängen der Trainingsphase den hoch probabilistischen Stimulus in allen drei Bedingungen signifikant häufiger auswählten (Testung gegen den Wert 50) als den niedrig probabilistischen, welcher nur zu Früchtetee oder Wasser führte (Abbildung 6: devaluiert: $T = 3.504$, $p = 0.004$; nicht devaluiert: $T = 2.604$, $p = 0.022$; neutral: $T = 2.728$, $p = 0.017$). Dementsprechend war instrumentelles Lernen erfolgreich induziert worden.

3.3.2. Devaluation

Nach der Devaluation mussten die Probanden mittels visueller Analogskala einschätzen, wie hungrig sie waren und als wie angenehm sie die dargebotenen appetitiven Verstärker in diesem Augenblick empfanden (Anreizwert). Anhand von t-Tests für abhängige Stichproben konnte ermittelt werden, dass die Probanden das devaluierte Lebensmittel (also entweder Schokolade oder Tomate) als signifikant weniger angenehm einstufen als diejenigen der anderen beiden Bedingungen, also nicht devaluiert und neutral (Abbildung 8: devaluiert: $T = 2.671$, $p = 0.019$; nicht devaluiert: $T = 1.965$, $p = 0.071$; Wasser: $T = 0.518$, $p = 0.613$). Zudem gaben die Probanden ein signifikant reduziertes Hungerlevel nach der Devaluation an (Abbildung 8: $T = 3.139$, $p = 0.008$). Im Gegensatz zur Pilotstudie hatte folglich in der Hauptstudie eine selektive Devaluation bezüglich des devaluierten Getränkes stattgefunden (zumindest in den subjektiven Angaben der Probanden).

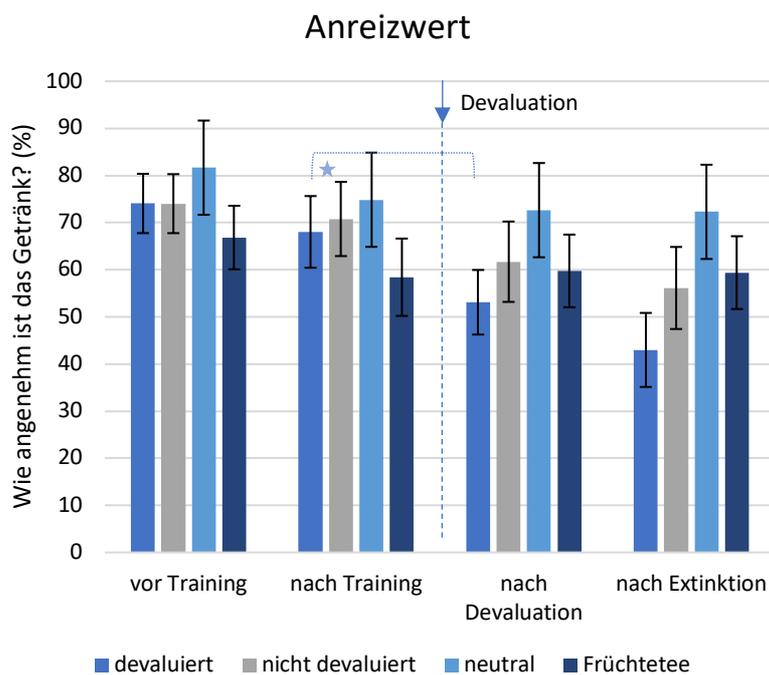
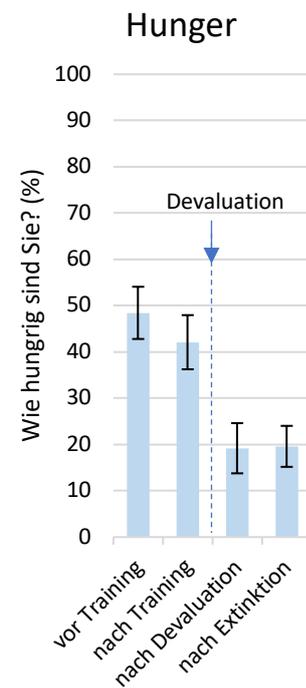
(A)**(B)**

Abbildung 8: Rating für Hunger und Anreizwert beim Devaluationsparadigma in der Hauptstudie (N = 14) in Anlehnung an Friedel et al. (2014)

(A) Subjektive Einstufung, wie angenehm die jeweiligen dargebotenen Reize (devaluiert: Schoko oder Tomate, neutral: Wasser, Flüssigkeit ohne Verstärkereffekt: Fruchttetee) zu vier unterschiedlichen Zeitpunkten während des Experimentes eingestuft wurden. Nach der Devaluation wurde das devaluierte Getränk als signifikant weniger angenehm angegeben (in der Grafik markiert durch ein Sternchen) im Vergleich zu der neutralen Bedingung und dem nicht devaluierten Getränk. (B) Signifikante Verminderung des Hungers nach der Devaluation

3.3.3. Extinktionsphase

Um das Entscheidungsverhalten in der Extinktionsphase einschätzen zu können und herauszufinden, ob die Probanden (N = 14) eher habituelles oder zielgerichtetes Verhalten zeigten, wurde mittels Varianzanalyse mit Messwiederholungen (2x2 ANOVA) die Interaktion zwischen Zeit (vor und nach Devaluation) und Bedingung (devaluiert/ nicht devaluiert) als *within-subject*-Faktoren betrachtet.

Es zeigte sich eine signifikante Interaktion zwischen den Parametern Zeit und Bedingung (Abbildung 9: $F(13) = 5.200$, $p = 0.040$). Im post-hoc-t-Test wurde deutlich, dass dieser Effekt vor allem darauf beruhte, dass es im Vergleich zwischen den ersten zehn Durchgängen der Extinktionsphase und den letzten zehn Durchgängen der Trainingsphase in den Verhaltensanalysen zu einer signifikanten Abnahme der Auswahlhäufigkeit des hoch probabilistischen Stimulus in der devaluierten Bedingung kam (Abbildung 6: $T = 3.151$, $p = 0.008$). Dies war in den

beiden anderen Bedingungen nicht gegeben (Abbildung 6: nicht devaluiert: $T = 1.479$, $p = 0.163$, neutral: $T = 1.251$, $p = 0.233$). Alle Testpersonen waren demnach sensitiv gegenüber der Devaluation und zeigten klare Anzeichen für zielgerichtetes Verhalten. Sie waren folglich in der Lage, ihr Entscheidungsverhalten nicht nur von ihrer gelernten Assoziation zwischen Stimulus und Getränk abhängig zu machen, sondern auch den aktuellen Anreizwert eines Getränkes mit einzubeziehen und dementsprechend ihr Antwortverhalten anzupassen.

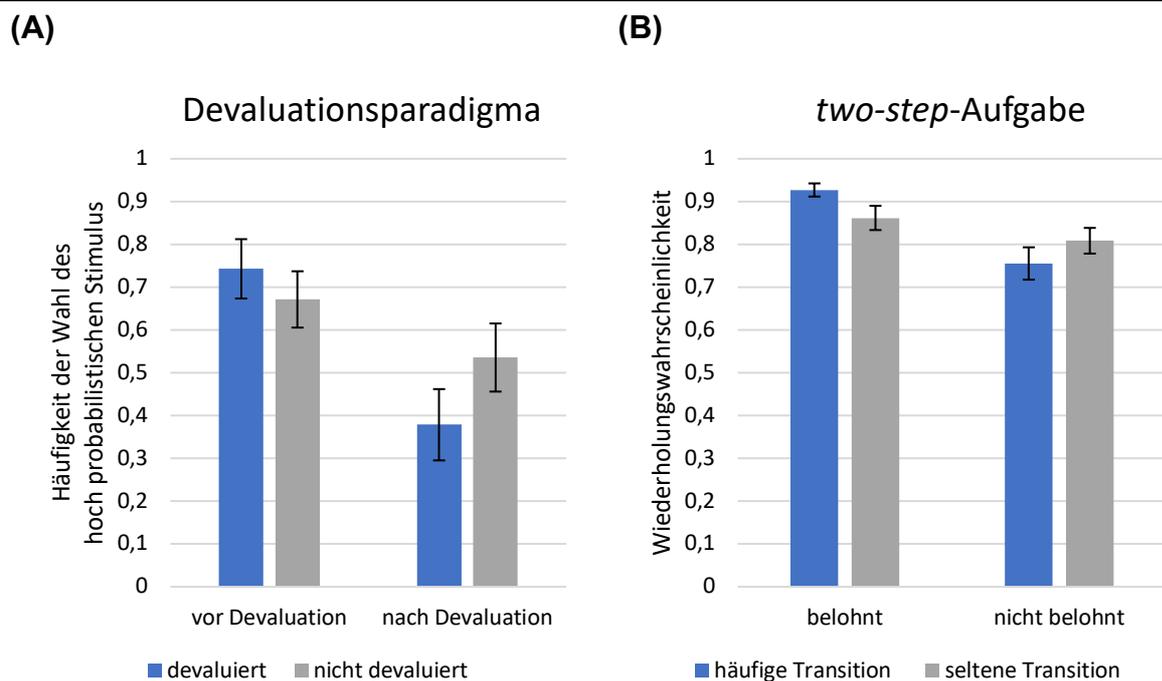


Abbildung 9: Analyse der Verhaltensdaten in Anlehnung an Friedel et al. (2014)

(A) Während der Extinktionsphase, also nach der Devaluation, kam es zu einer signifikanten Abnahme der Wahl des hoch probabilistischen Stimulus, der mit dem devaluierten Lebensmittel assoziiert war ($N = 14$) (B) Für die *two-step*-Aufgabe zeigte sich für alle Probanden ($N = 17$) ein Verhaltensmuster, welches sich aus modellfreien und modellbasierten Anteilen zusammensetzte.

3.4. *two-step*-Aufgabe

In Hinblick auf die Analysen der Verhaltensdaten mittels 2x2 ANOVA für die *two-step*-Aufgabe zeigte sich, dass das Entscheidungsverhalten der Probanden weder auf rein modellfreien noch auf rein modellbasierten Strategien basierte. Die Probanden nutzten also vielmehr eine Mischung aus beiden Strategien der Verhaltenskontrolle (Abbildung 9). Dies zeigte sich daran, dass sowohl ein signifikanter Haupteffekt für Belohnung (p -Belohnung) als Indikator für modellfreies Verhalten ($F(16) = 14.410$, $p = 0.002$) als auch eine signifikante Interaktion von

Belohnung und Transitionsbedingung (p-Interaktion) als Indikator für modellbasiertes Verhalten nachgewiesen werden konnte ($F(16) = 6.048, p = 0.026$). Der Haupteffekt Transition war nicht signifikant ($F < 1, p = 0.631$).

Im Durchschnitt erzielten die Probanden einen Gewinn von 6,58€ (Intervall zwischen 5,80€ und 8,00€, SD: 0.63). Aufgabe der Probanden war es, ihren Gewinn durch entsprechendes Entscheidungsverhalten zu maximieren. Um dies zu erreichen, musste das Verhalten nicht nur an der Belohnung im vorherigen Durchgang, sondern auch an der Transitionsmatrix der Aufgabe orientiert werden. Folglich müsste ein Proband umso mehr Gewinn erzielen, je höher der Anteil der modellbasierten Strategien an seinem Verhalten ist. Letztlich zeigte sich in den Korrelationsanalysen zwar von der Richtung ein moderater, positiver Zusammenhang zwischen dem erzielten Gewinn und dem Ausmaß modellbasierten Verhaltens (p-Interaktion) in der *two-step*-Aufgabe, dieser war allerdings statistisch nicht signifikant (Spearman Rangkorrelation $r = 0.223, p = 0.390$).

3.5. Konstruktvalidität: Korrelation zwischen den beiden Paradigmen

Insgesamt wurden 13 Probanden in die finalen Analysen eingeschlossen. Von den vormals 18 Testpersonen mussten in Hinblick auf das Devaluationsparadigma zwei wegen des nicht erfüllten Lernkriteriums und zwei weitere wegen fehlender Sättigungseffekte von den Analysen ausgeschlossen werden. Bei einer weiteren Person kam es während der *two-step*-Aufgabe zu einem Programmabbruch. Letztlich konnte in den Analysen eine signifikante, positive Korrelation zwischen der Devaluationssensitivität, welche über den Interaktionsterm der Devaluationsaufgabe ermittelt wurde, und dem Interaktionsterm p-Interaktion – als Gradmesser für das individuelle Ausmaß modellbasierten Verhaltens in der *two-step*-Aufgabe – nachgewiesen werden (Spearman Rangkorrelation $r = 0.698, p = 0.008$, zweiseitig). Auch nach Ausschluss eines Ausreißers (3 Standardabweichungen größer als das Mittel) war dieser Effekt in der Stichprobe von 12 Probanden signifikant (Abbildung 10: Spearman Rangkorrelation $r = 0.735, p = 0.006$, zweiseitig). Eine signifikante, negative Korrelation zwischen dem Haupteffekt für Belohnung (p-Belohnung) und der Devaluationssensitivität zeigte sich hingegen nicht (Abbildung 10: Spearman Rangkorrelation $r = 0.163, p = 0.612$, zweiseitig).

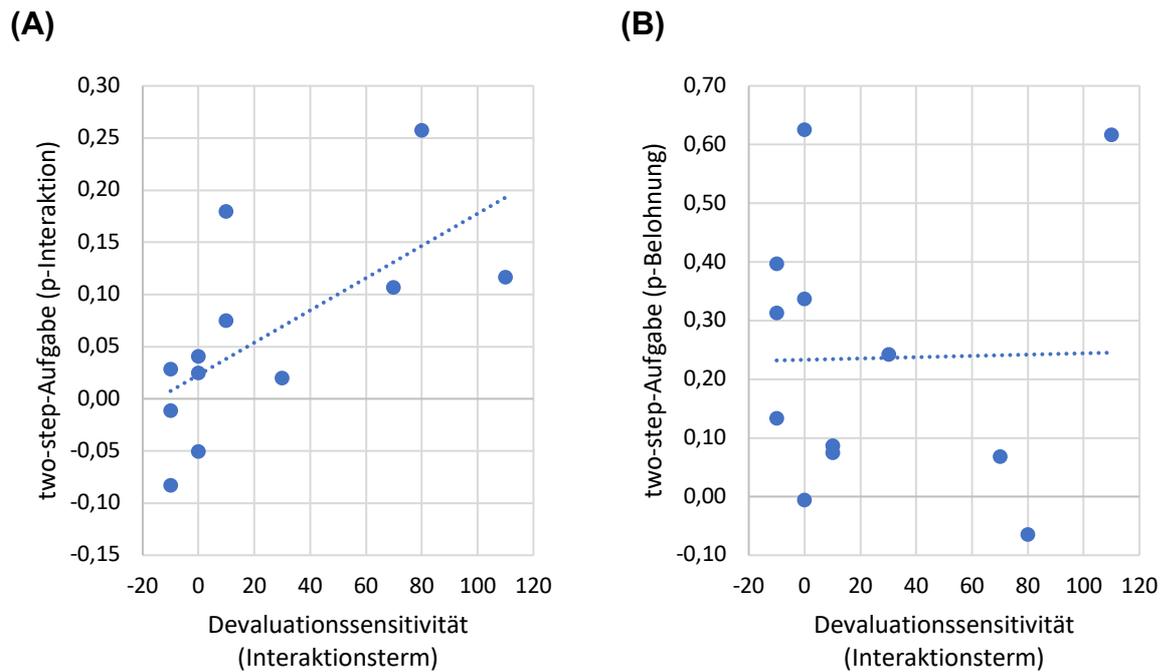


Abbildung 10: Korrelation der Parameter beider Paradigmen (N = 12)

(A) Es zeigte sich eine signifikante Korrelation zwischen der Devaluations sensitivität in der Devaluationsaufgabe und dem Interaktionsterm p-Interaktion der *two-step*-Aufgabe (in Anlehnung an Friedel et al., 2014). (B) Solch ein Zusammenhang konnte zwischen Devaluations sensitivität und dem Haupteffekt der Belohnung (p-Belohnung) in der *two-step*-Aufgabe nicht gefunden werden.

3.6. Zusammenhang zwischen Entscheidungsverhalten und kognitiven Fähigkeiten

Mit Blick auf die neuropsychologische Testung ergab sich ein recht homogene Probandenstichprobe. Eine Auswertung der Ergebnisse für die neuropsychologische Testung zeigt Tabelle 1.

	Mittelwert	Standardabweichung	Spannweite
Wörterliste	8.6	1.38	6 – 10
TMT-A	21.9s	6.53s	12 – 36s
TMT-B	53.7s	16.14s	33 – 91s
ZST	78.5	14.94	48 – 101
ZN	7.6	2.45	4 – 12
MWT-B	30.2	3.18	22 – 35

Tabelle 1: Ergebnis der neuropsychologischen Testung (N = 18)

Wörterliste: entnommen aus dem CERAD; TMT-A und TMT-B: Pfadfindertest in der Version A (nur Zahlen) und B (Zahlen und Buchstaben); ZST: Zahlen-Symbol-Test; ZN: Zahlen nachsprechen, in diesem Fall nur rückwärts; MWT-B: Mehrfachwahl-Wortschatz-Intelligenztest in der Version B

Wie im Methodenteil erwähnt, konnte in Untersuchungen bereits dargelegt werden, dass die individuelle Ausprägung von modellbasiertem Verhalten abhängig von der kognitiven Verarbeitungsgeschwindigkeit zu sein scheint, welche insbesondere durch die Leistungen im Zahlen-Symbol-Test erfasst wird (Schad et al., 2014; Sebold et al., 2014; Sjoerds et al., 2016). Diese Assoziation stellte sich bei unserer Probandenstichprobe jedoch nicht heraus. Interessanterweise ergab sich zunächst eine negative Korrelation zwischen Zahlen-Symbol-Test und dem Interaktionsterm p-Interaktion der *two-step-Aufgabe* (Spearman Rangkorrelation $r = -0.526$, $p = 0.036$), wobei dieser Effekt nach Bonferroni-Korrektur nicht mehr signifikant war ($p > 0.008$). Darüber hinaus zeigten sich keine weitere Korrelation mit anderweitigen Ergebnissen aus der neuropsychologischen Untersuchung, insbesondere nicht mit dem individuellen Grad der Devaluationssensitivität (Tabelle 2)

	p-Interaktion (Interaktionsterm <i>two-step</i> , N = 17)		Devaluationssensitivität (Interaktionsterm Devaluation, N = 14)	
	r	p	r	p
Wörterliste	-0.391	0.121	0.111	0.707
TMT-A	0.022	0.933	-0.438	0.118
TMT-B	0.222	0.408	0.423	0.150
ZST	-0.526	0.036*	-0.265	0.382
ZN	-0.090	0.731	0.016	0.958
MWT-B	0.348	0.171	0.232	0.424

Tabelle 2: Korrelationsanalysen zwischen Ergebnissen der neuropsychologischen Testung und den aus den Paradigmen ermittelten Interaktionstermen

* $p < 0,05$, r: Spearman Rangkorrelationskoeffizient, p = Irrtumswahrscheinlichkeit

4. Diskussion

4.1. Interpretation der Ergebnisse mit Hinblick auf die Konstruktvalidität

Ziel dieser Arbeit war es, die Konstruktvalidität zweier etablierter Paradigmen zu überprüfen, die sich konzeptionell aus verschiedenen Bereichen der experimentellen Psychologie entwickelt haben. Diese Paradigmen haben zum Ziel, die verschiedenen Strategien der Verhaltenskontrolle (habituelles vs. zielgerichtetes und modellfreies vs. modellbasiertes Verhalten) abzubilden und so messbar zu machen. Um der Frage der Konstruktvalidität nachzugehen, absolvierten in der Hauptstudie achtzehn gesunde Männer jeweils beide Paradigmen. Zur Erfassung von zielgerichtetem vs. habituellem Verhalten wurde eine selektive Devaluationsaufgabe eingesetzt. Mittels der in dieser Arbeit verwendeten sequentiellen Entscheidungsaufgabe (*two-step*-Aufgabe) sollten hingegen die Anteile modellbasierten und modellfreien Verhaltens erfasst werden.

Zunächst konnte Hypothese (1) dieser Arbeit bestätigt werden, denn in der Tat zeigten die Probanden in der Hauptstudie zielgerichtetes Verhalten in der Extinktionsphase der verwendeten Devaluationsaufgabe. Dies bildete sich darin ab, dass nach der Devaluation ein signifikant reduziertes Antwortverhalten für den hoch probabilistischen Stimulus in der devaluierten Bedingung nachweisbar war – in den anderen beiden Bedingungen, insbesondere der nicht devaluierten, wählten die Probanden den hoch probabilistischen Stimulus hingegen annähernd genauso häufig aus wie vor der Devaluation. Wie lässt sich nun solch eine Änderung des Verhaltens in Hinblick auf die Vermeidung eines spezifischen Stimulus erklären? Durch das Devaluationsprozedere, also die selektive Übersättigung einer vormaligen Belohnung bzw. eines Verstärkers, sollte es auch zu einer Verminderung des Anreizwertes für denjenigen Stimulus kommen, der mit diesem Verstärker assoziiert war. Das dieses Prozedere gelang, spiegelte sich einerseits in den subjektiven Ratings der Anreizwerte für die dargebotenen appetitiven Verstärker wider: Je nachdem, mit welchem Lebensmittel die Probanden übersättigt worden waren, gaben sie nach der Devaluation für den entsprechenden Verstärker (in Form eines Getränkes) einen verminderten Anreizwert an. Auch das Hungergefühl wurde nach der Devaluation als signifikant weniger angegeben. Doch die Probanden stuften das übersättigte Lebensmittel nicht nur als weniger angenehm ein, sondern sie passten ihr Auswahlverhalten in der Extinktionsphase (nach der Devaluation) anscheinend

auch dem nun veränderten Anreizwert des Verstärkers an: So vermieden die Probanden, auf Grundlage der im instrumentellen Training erlernten Assoziation zwischen einem spezifischen Stimulus und dem Antwortverhalten (Wahl des hoch probabilistischen Stimulus in jeder Bedingung), in der devaluierten Bedingung spezifisch denjenigen Stimulus, der mit dem nun weniger angenehmen Getränk bzw. Verstärker assoziiert war. Sie waren demnach sensitiv gegenüber dem Devaluationsprozedere – wenn auch in einem unterschiedlichen Ausmaß, was sich in der Höhe des Interaktionsterms für die Devaluationsaufgabe widerspiegelte. Diese Ergebnisse standen mit denen aus früheren Studien im Einklang (Schwabe and Wolf, 2009; Valentin et al., 2007)

In der *two-step*-Aufgabe zeigten die Probanden eine Mischung aus modellfreiem und modellbasiertem Verhalten, was in der signifikanten Interaktion zwischen Belohnung und Transitionsbedingung zum Ausdruck kam. Somit konnte auch Hypothese (2) bestätigt werden. Die Teilnehmer berücksichtigten demnach neben der Belohnung auch die Transitionsstruktur als Grundlage des Entscheidungsfindungsprozesses, um ihren Gewinn in dieser Aufgabe zu maximieren. Somit wurden ähnliche Ergebnisse erzielt, wie sie bereits in vorherigen Studien publiziert wurden (Daw et al., 2011; Smittenaar et al., 2013; Wunderlich et al., 2012).

Per Definition ist zielgerichtetes Verhalten in seiner Natur auch immer modellbasiert, da es sich an den möglichen Konsequenzen des Antwortverhaltens und deren Anreizwerten orientiert (Friedel et al., 2014; Voon et al., 2017). In der vorliegenden Arbeit konnte dargelegt werden, dass ein Zusammenhang zwischen diesen beiden Strategien der Verhaltenskontrolle besteht. Entsprechend der a-priori-Hypothese (3) zeigte sich eine positive Korrelation zwischen dem Interaktionsterm aus der Devaluationsaufgabe (als Gradmesser für die Devaluationssensitivität) und dem Anteil modellbasierten Verhaltens, der mittels der sequentiellen Entscheidungsaufgabe bestimmt wurde (Interaktionsterm p-Interaktion der *two-step*-Aufgabe). Daraus lässt sich ableiten: Je höher die individuelle Devaluationssensitivität – und somit das Ausmaß zielgerichteten Verhaltens in der Devaluationsaufgabe – war, desto mehr nutzten die Probanden in der *two-step*-Aufgabe eine modellbasierte Strategie um ihren Gewinn zu maximieren. Diese Befunde unterstützen die Annahme, dass die beiden erläuterten theoretischen

Konzepte tatsächlich auf einem gemeinsamen Mechanismus beruhen, so wie es bereits in den letzten Jahren zwar postuliert, aber bis dato nicht empirisch an einer Stichprobe überprüft worden war (Dolan and Dayan, 2013). Das Vorhandensein eines Zusammenhangs zwischen modellbasierter und zielgerichteter Verhaltenskontrolle konnte auch in späteren Studien repliziert werden. So zeigten Gillan et al. (2015) an zwei Stichproben mit insgesamt 185 gesunden Probanden, dass das Ausmaß modellbasierten Verhaltens als Indikator für den Grad der Sensitivität gegenüber einer selektiven Devaluation herangezogen werden kann. Auch in einer weiteren Studie (Sjoerds et al., 2016), bei der die Konstruktvalidität zwischen der *two-step*-Aufgabe und einer Variante des Devaluationsparadigmas, adaptiert nach de Wit et al. (2012), überprüft wurde, zeigte sich eine Relation zwischen modellbasiertem und zielgerichtetem Verhalten.

Betrachtet man die bisherigen Befunde von fMRT-Studien zu Verhaltens- und Lernmechanismen, so gibt es auch hier evidente Hinweise darauf, dass zielgerichtetes Verhaltensmuster und modellbasierte Strategien der Verhaltenskontrolle auf ähnlichen (neuronalen) Mechanismen zu beruhen scheinen, (Daw and O'Doherty, 2014). Es gibt es in den Neurowissenschaften zunehmend Evidenz dafür, dass sowohl der ventromediale präfrontale Kortex als auch der orbitofrontale Kortex beim Menschen eine wichtige Struktur bei der Entstehung zielgerichteter Verhaltensweisen darstellt (Balleine and O'Doherty, 2010; de Wit et al., 2012, 2009; Valentin et al., 2007). Auch in Bezug auf modellbasiertes Verhalten konnten jene Strukturen als zugrundeliegende neurobiologische Korrelate identifiziert werden (Gläscher et al., 2010; Smittenaar et al., 2013; Voon et al., 2015).

Um abzubilden, in welchen Ausmaß Probanden modellfreie Strategien in der *two-step-Aufgabe* nutzten, wurde zusätzlich der Parameter p-Belohnung (Haupteffekt der Belohnung) für jeden Probanden ermittelt. Ein individuell niedriger Interaktionsterm in der Devaluationsaufgabe sprach für eine geringe Devaluationssensitivität und folglich für einen verminderten Grad zielgerichteter sowie ein Überwiegen habitueller Verhaltensstrategien. Bei gegebener Konstruktvalidität zwischen den untersuchten Paradigmen wäre zu erwarten gewesen, dass Probanden, die eine verminderte Sensitivität gegenüber einer selektiven Devaluation und damit eher habituelles Verhaltensmuster zeigen, entsprechend größeres Ausmaß modellfreier Verhaltensstrategien nutzten. Allerdings war kein Zusammenhang zwischen dem

Haupteffekt der Belohnung (p-Belohnung) und der Devaluationssensitivität im Sinne einer negativen Korrelation nachweisbar, sodass Hypothese 4 dieser Arbeit letztlich nicht belegt werden konnte. Auch in den beiden bereits erwähnten Studien zeigte sich kein Zusammenhang zwischen modellfreiem Verhalten, welches über den Haupteffekt der Belohnung abgebildet wird, und der Devaluationssensitivität (Gillan et al., 2015; Sjoerds et al., 2016). Einen möglichen Erklärungsansatz für diesen Befund lieferten bereits Valentin et al. (2007) in einer früheren fMRT-Untersuchung. Sie konnten zeigen, dass während einer selektiven Devaluationsaufgabe auf neuronaler Ebene keine Aktivierung ebenjener Strukturen stattfindet, die eigentlich mit habituellen Verhaltensstrategien assoziiert sind. Die Autoren schlussfolgerten daraus, dass die verwendete Devaluationsaufgabe womöglich nicht dazu geeignet ist, habituelles Verhalten im Sinne einer verminderten Devaluationssensitivität zu evozieren und zu erfassen. Devaluationsparadigmen scheinen eher zielgerichtetes Verhalten abzubilden, wohingegen habituelles Verhalten besser durch Lernaufgaben mit extensivem Training induziert werden kann (Tricomi et al., 2009).

Demnach ist kritisch zu hinterfragen, inwieweit die in dieser Arbeit verwendete Devaluationsaufgabe (bzw. die Abbildung der Devaluationssensitivität über den Interaktionsterm) überhaupt dazu geeignet ist, Aussagen über den Anteil habitueller Verhaltensstrategien zu treffen. Letztlich zeigte sich ja auch in der Auswertung der Hauptstudie, dass die Probanden überwiegend zielgerichtetes Verhalten im Sinne einer hohen Devaluationssensitivität aufwiesen – wenn auch in einem unterschiedlichen Ausmaß. Daraus folgt, dass für zukünftige Studien zur Überprüfung der Konstruktvalidität beider Konzepte auch solche Paradigmen verwendet werden sollten, die habituelle Verhaltensweisen induzieren und erfassen. Eine Möglichkeit wäre es, ein Paradigma mehrstufig zu konzipieren, sodass zunächst habituelles Verhalten induziert (z.B. durch extensives Training) und zum Vorteil bei der Bewältigung einer Aufgabe wird. Im nächsten Schritt würde dann nach einer Devaluationsphase, welche zu einer selektiven Abwertung einer vormaligen Belohnung führen soll, zielgerichtetes Verhalten die profitablere Verhaltensstrategie darstellen. In der zweiten Hälfte der Untersuchung könnte dann also untersucht werden, ob die Probanden in der Lage sind, ihre Verhaltensstrategie anzupassen.

Ein weiterer Aspekt, der als mögliche Erklärung für den fehlenden Zusammenhang zwischen der Devaluationssensitivität und dem Ausmaß für modellfreies Lernen berücksichtigt werden sollte, ist auch die Struktur der verwendeten sequentiellen

Entscheidungsaufgabe. In der *two-step*-Aufgabe ändern sich die Belohnungswahrscheinlichkeiten auf der zweiten Ebene im Verlauf der Aufgabe dynamisch und langsam (*random walk*). Auch modellfreies Verhalten bleibt hierbei sensitiv gegenüber Änderungen in der Umwelt, ohne dabei jedoch die Aufgabenstruktur zu berücksichtigen. Im Gegensatz dazu beruht habituelles Lernen auf der Assoziation zwischen Stimulus und Antwortverhalten und wird durch häufige Wiederholungen etabliert, bleibt dabei aber relativ stabil gegenüber plötzlichen Veränderungen (Colwill and Rescorla, 1988). Durch die langsame Veränderung der Belohnungswahrscheinlichkeiten (*random walk*) sollte ein Überlernen infolge von repetitiven, unveränderten Abläufen bei der *two-step*-Aufgabe verhindert werden. Wie zuvor erläutert, ist dieser Prozess des extensiven Lernens jedoch eine wichtige Komponente bei der Induktion habituellen Verhaltens (Tricomi et al., 2009). Folglich sind Aufgaben zur sequentiellen Entscheidungsfindung möglicherweise nicht gut geeignet, um jene Prozesse zu induzieren oder identifizieren, die zu habituellem Verhalten im Sinne einer verminderten Sensitivität gegenüber einer Devaluation bzw. gegenüber einer Änderung des Anreizwertes der Belohnung führen. Insofern lässt sich konstatieren, dass habituelles Lernen durch die zwei verwendeten Paradigmen vermutlich nur unzureichend erfasst werden konnte.

Darüber hinaus ist in Erwägung zu ziehen, dass die Paradigmen womöglich nur unterschiedliche Aspekte der von Dolan und Dayan (2013) postulierten reflexiven – also auf zurückliegenden Lernerfahrungen beruhenden – Verhaltensstrategie abbilden. Folglich stellen habituelle und modellfreie Strategien vielleicht nur Teilaspekte eines umfassenderen Systems der reflexiven Verhaltenskontrolle dar, die dementsprechend auch unterschiedlich operationalisiert werden – ähnlich beispielsweise dem Konstrukt der Intelligenz, bei dem die Teilbereiche Sprachverständnis, Verarbeitungsgeschwindigkeit, logisches Denken und Arbeitsgedächtnis jeweils über unterschiedliche Tests erfasst werden (Wechsler, 1981).

Dass entgegen der der a-priori-Hypothese keine negative Korrelation zwischen dem Ausmaß modellfreien Verhaltens und der Devaluationssensitivität gefunden wurde, bedeutet im Umkehrschluss demnach nicht, dass nicht doch ein Zusammenhang zwischen diesen Parametern bestehen könnte. Gillan et al. (2015) führten in ihrer Studie daher eine weitere Analyse durch (Median Split), bei der die Devaluationssensitivität als abhängige Variable miteinfluss. Dabei ergab sich, dass

Probanden, die beim Devaluationsparadigma in der Extinktionsphase überwiegend habituelle Verhaltensstrategien nutzten, in der sequentiellen Entscheidungsaufgabe keine modellbasierte Strategie anwandten. Obgleich also das Ausmaß modellfreien Verhaltens keinen Prädiktor für die Devaluationssensitivität bzw. für das Ausmaß habituellen Verhaltens darstellt, so geht ein Überwiegen habitueller Verhaltensstrategien anscheinend dennoch mit einer Einschränkung in der Nutzung modellbasierter Strategien einher.

Trotzdem stellt sich weiterhin die Frage, inwieweit modellfreie und habituelle Verhaltensmechanismen überhaupt gleichzusetzen sind. Die nicht gefundene Korrelation könnte auch ein Hinweis darauf sein, dass diese Strategien der Verhaltenskontrolle durch unterschiedliche neuronale Prozesse vermittelt werden. Bezieht man wiederum fMRT-Studien in die Argumentation mit ein, lassen sich auch hier Belege für diese Annahme finden: So konnten für habituelle und modellfreie Verhaltensmuster bisher nur wenige Übereinstimmungen in den zugrundeliegenden neuronalen Netzwerken gefunden werden (Daw and O'Doherty, 2014). Als homologes Areal für das dorsolaterale Striatum bei Ratten gilt das Putamen (als Teil des dorsalen Striatums) beim Menschen, weshalb diese Region insbesondere dem habituellen System zugeordnet wird (Daw and O'Doherty, 2014; de Wit et al., 2012; Tricomi et al., 2009). Im Gegensatz dazu gelten der Belohnungsvorhersagefehler (*reward prediction error*) und seine neuronale Repräsentation im ventralen Striatum als zugrundeliegender Mechanismus für modellfreie Verhaltensstrategien beim Menschen (Gläscher et al., 2010; O'Doherty et al., 2003). Diese Befunde untermauern einmal mehr die Hypothese, dass modellbasierte und zielgerichtete Strategien auf neurobiologischer Ebene auf ähnlichen Strukturen beruhen, wohingegen dies für ihre Konterparts weniger zuzutreffen scheint.

Interessanterweise zeigte sich in den durchgeführten Analysen im Gegensatz zu anderen Untersuchungen (Schad et al., 2014; Sebold et al., 2014; Sjoerds et al., 2016) zunächst ein negativer (nicht positiver) Zusammenhang zwischen dem Ergebnis des Zahlen-Symbol-Testes und dem Interaktionsterm der *two-step*-Aufgabe. Allerdings war dieser nach Korrektur für multiples Testen nicht mehr signifikant. Wie bereits zu Beginn dargelegt, beanspruchen modellbasierte Strategien ein höheres Maß an kognitiver Prozessleistung, was Komponenten wie Aufmerksamkeit, Arbeitsgedächtnis und kognitive Verarbeitungsgeschwindigkeit

miteinschließt. Solche Parameter können mithilfe der neuropsychologischen Testung abgebildet werden. Es wäre zu erwarten gewesen, dass eine bessere Leistung im Zahlen-Symbol-Test mit einem höheren Interaktionsterm p -Interaktion in der *two-step*-Aufgabe, also einem Überwiegen modellbasierter Strategien, einhergehen würde. Ein solches Ergebnis konnte jedoch mittels der vorliegenden Arbeit nicht reproduziert werden. Als mögliche Ursache ist hier der Deckeneffekt zu diskutieren: Alle Studienteilnehmer zeigten gute Ergebnisse in der neuropsychologischen Testung, was ein Hinweis auf eine insgesamt hohe kognitive Leistungsfähigkeit der Probanden ist. Daraus resultiert aber auch eine geringe Varianz innerhalb der Stichprobe, welche möglicherweise so gering war, dass bestehende Unterschiede zwischen den Probanden durch die entsprechenden Untersuchungen nicht ausreichend differenziert werden konnte. Dies könnte auch eine Erklärung dafür sein, warum in Bezug auf die weiterführenden Korrelationsanalysen keine Zusammenhänge mit anderweitigen Ergebnissen aus der neuropsychologischen Untersuchung gefunden werden konnten.

Der genannte Deckeneffekt könnte aber auch einen Einfluss auf die Korrelation zwischen den Parametern der beiden Paradigmen gehabt haben: Möglicherweise zeigten die Probanden aufgrund ihrer guten kognitiven Performance überwiegend zielgerichtete bzw. modellbasierte Verhaltensmechanismen, sodass Unterschiede hier gut erfasst und ein signifikanter Zusammenhang zwischen der Devaluationssensitivität und dem Ausmaß modellbasierten Verhaltens nachgewiesen werden konnte. Dies würde aber auch im Umkehrschluss bedeuten, dass habituelle bzw. modellfreie Strategien kaum oder gar nicht bei den Probanden vorhanden waren und dadurch auch kein Zusammenhang zwischen dem Haupteffekt für Belohnung als Maß für modellfreies Lernen und der Devaluationssensitivität gefunden werden konnte – selbst, wenn dieser eigentlich vorhanden gewesen wäre. Für zukünftige Untersuchungen wäre also eine größere Varianz zwischen den Probanden auch in Hinblick auf ihre kognitiven Fähigkeiten wünschenswert, um jenen Deckeneffekt zu reduzieren bzw. zu vermeiden.

Letztlich sollte bei der Diskussion der divergierenden Untersuchungsergebnisse ein – zwar genereller, aber nicht unwesentlicher – Aspekt berücksichtigt werden: Die beiden vorgestellten Konstrukte und deren Operationalisierungen sind im Grunde genommen nur eine vereinfachte Darstellung der tatsächlich zugrundeliegenden

Mechanismen (Dolan and Dayan, 2013). Daraus ergibt sich ein weiterer möglicher Erklärungsansatz dafür, dass zwar eine positive signifikante Korrelation zwischen Devaluationssensitivität und dem Ausmaß modellbasierten Verhaltens (p-Interaktion), aber kein negativer Zusammenhang zwischen Devaluationssensitivität und dem Haupteffekt der Belohnung als Gradmesser für modellfreies Verhalten gefunden werden konnte: Die Konzepte von zielgerichteten vs. habituellem sowie modellfreiem vs. modellbasiertem Verhalten sind in ihrem Kern möglicherweise unterschiedlich konfiguriert und daher nur bedingt vergleichbar. Bezüglich des Konzepts modellfreier und modellbasierter Strategien der Verhaltenskontrolle gibt es auf der Grundlage neurobiologischer Untersuchungen Hinweise darauf, dass dieses Konstrukt auf einer Balance zwischen diesen Systemen basiert (Dolan and Dayan, 2013; Gillan et al., 2015). Bisher besteht allerdings nur wenig Evidenz darüber, ob die Interaktion zwischen den verschiedenen Systemen auf einem „ausgewogenen“ Verhältnis beruht, welches durch einen übergeordneten Vermittler moderiert wird, oder doch eher einen kompetitiven Charakter im Sinne einer wechselseitigen Inhibition besitzt (Daw and O’Doherty, 2014). So gehen Daw et al. (2015) davon aus, dass es durchaus eine Konkurrenzsituation zwischen dem modellfreien und modellbasierten System gibt. Dabei soll es abhängig von der Situation sowie unter Berücksichtigung einer Kosten-Nutzen-Abwägung (modellbasierte Strategien erfordern immerhin deutlich mehr kognitive Ressourcen als modellfreie) zum Überwiegen des einen oder anderen Systems kommen (Niv et al., 2007). Dabei scheinen nicht nur die einzelnen Komponenten (modellbasiert vs. modellfrei) und deren Ausprägung bei der Verhaltenskontrolle von Bedeutung zu sein, sondern letztlich auch deren Zusammenspiel bzw. das Gleichgewicht zwischen diesen beiden Systemen. Ungeachtet dessen, ob nun ein Gleichgewicht oder eine Konkurrenzsituation zwischen diesen beiden Strategien der Verhaltenskontrolle vorherrscht, ist noch nicht vollständig verstanden, wie ihr Zusammenspiel letztlich dirigiert wird. Abermals lieferten fMRT-Studien diesbezüglich wichtige Informationen für ein besseres Verständnis der zugrundeliegenden Mechanismen: Es konnte gezeigt werden, dass es im Verlauf der *two-step*-Aufgabe auf neuronaler Ebene zu einer Überlappung beider Formen des Vorhersagefehlers (also sowohl RPE als auch SPE) kommt. Dabei fungiert das ventrale Striatum möglicherweise als jene Struktur, die die zugrundeliegenden neuronalen Systeme modellfreier und modellbasierter Strategien der Verhaltenskontrolle in den Entscheidungsprozess integriert (Daw et al., 2011).

Lee et al. (2014) hingegen konnten Hinweise dafür finden, dass es innerhalb eines neuronalen Netzwerkes eine unabhängige Kontrollinstanz zu geben scheint, die zwischen den modellfreien und modellbasierten Systemen der Verhaltenskontrolle vermittelt. Die Aufgabe des Vermittlers liegt darin, die beiden Systeme in Hinblick auf ihre Effizienz hin zu vergleichen und gegeneinander abzuwägen. Unter Berücksichtigung der Kompetenzen der verschiedenen Systeme und den jeweiligen Anforderungen, die in einer bestimmten Situation an den Agenten gestellt werden, moderiert der Vermittler den individuellen Ausprägungsgrad der Systeme in dieser Situation und initiiert letztlich die entsprechenden Handlungsoptionen.

Während also bei modellfreien und modellbasierten Strategien die neuronalen Systeme auf die eine oder andere Weise miteinander zu interagieren scheinen, zeichnet sich für habituelle und zielgerichtete Verhaltensmechanismen ein etwas anderes Bild: Läsionsstudien bei Ratten konnten zeigen (Balleine and Dickinson, 1998; Yin et al., 2005; Yin and Knowlton, 2006), dass es bei der Inaktivierung jener Strukturen, die dem habituellen System der Verhaltenskontrolle zugesprochen werden, nicht zu einer Veränderung oder gar Einschränkung des zielgerichteten Systems kommt. Dies führt zu der Annahme, dass es sich beim Konzept von zielgerichteter und habitueller Verhaltenskontrolle um ein dichotomes Konstrukt handelt, bei dem beide Systeme separat und unabhängig voneinander arbeiten. Solch eine Konstellation wird auch als „doppelte Dissoziation“ bezeichnet (Daw and O’Doherty, 2014). Allerdings ergaben sich in den letzten Jahren zunehmend Hinweise darauf, dass zielgerichtete und habituelle Systeme nicht vollständig autonom sind. Schwabe und Wolf (2009) diskutierten diesbezüglich bereits 2007 im Rahmen einer Devaluationsstudie, dass die Amygdala aufgrund ihres modulierenden Einflusses auf verschiedene neuronale Systeme bei der Koordination von habitueller und zielgerichteter Verhaltenskontrolle mitwirkt.

Interessanterweise fanden Sebold et al. (2014) unter Anwendung der *two-step*-Aufgabe hingegen heraus, dass alkoholabhängige Patienten zwar Defizite im modellbasierten Verhalten zeigen, der Anteil der modellfreien Strategien sich dabei aber im Vergleich zu Gesunden nicht unterscheidet. Dieser Befund lässt eine andere Annahme über das Zusammenspiel der verschiedenen Strategien der Verhaltenskontrolle zu: Möglicherweise stellen das habituelle und modellfreie System sehr robuste, bereits von Beginn an bestehende Verhaltensmuster dar, die

gewissermaßen das Fundament für alle sich im Verlauf des Lebens entwickelnden Strategien bilden (Wood and Rüniger, 2016). Die Varianz von Verhalten wird dann durch den unterschiedlichen Einfluss einer übergeordneten Struktur, nämlich den modellbasierten bzw. zielgerichteten System, realisiert. Folglich arbeiten die divergierenden Systeme nicht separat oder parallel, sondern ein System wird auf hierarchische Weise von einem anderen dominiert und modifiziert. Diese Allgegenwärtigkeit einer modulierenden Instanz in Form des modellbasierten bzw. zielgerichteten Systems wurde bereits früher postuliert (Dezfouli and Balleine, 2013, 2012; Doll et al., 2012). Dass unterschiedliche und vor allem dysfunktionale Lern- und Verhaltensmechanismen also gegebenenfalls auf einer Varianz bzw. Abnahme des modellbasierten bzw. zielgerichteten Verhaltens beruhen und nicht auf einer Zunahme des Gegenparts, ist auch eine weitere mögliche Erklärung dafür, dass sich in der vorliegenden Studie zwar ein Zusammenhang zwischen modellbasiertem Verhalten (p -Interaktion) und individueller Devaluationssensitivität feststellen ließ, dieser jedoch nicht für modellfreies Lernen (Haupteffekt Belohnung) und habituelles Verhalten (in Form einer verminderten Devaluationssensitivität) nachgewiesen werden konnte. Denn durch beide Paradigmen werden primär die Varianzen der modulierenden, übergeordneten Instanzen erfasst, was ein Vergleich der beiden Paradigmen in Hinblick auf modellbasiertes bzw. zielgerichtetes Verhalten möglich macht und dementsprechend zu einer positiven Korrelation führt. Die Fundamente, also habituelle und modellfreie Strategien der Verhaltenskontrolle, können jedoch – wie bereits dargelegt – mittels der hier verwendeten Paradigmen möglicherweise nur unzureichend operationalisiert werden.

Zusammenfassend lassen sich somit folgende Erklärungsansätze für die divergierenden Ergebnisse dieser Studie und die scheinbar nur teilweise vorliegende Konstruktvalidität der beiden verwendeten Paradigmen finden:

- Zielgerichtete Strategien stellen eine Facette modellbasierten Verhaltens dar, sodass diesbezüglich ein Zusammenhang nachgewiesen werden konnte. Indes sind habituelle und modellfreie Strategien möglicherweise Teilaspekte eines umfassenderen Systems der reflexiven Verhaltenskontrolle und werden dementsprechend unabhängig voneinander ausgebildet und operationalisiert.

- Zielgerichtete und modellbasierten Verhaltensstrategien liegen ähnliche neuronale Systeme zugrunde, für habituelle und modellfreie Systeme ließen sich solche Übereinstimmungen nicht finden.
- Habituelle und modellfreie Strategien werden hierarchisch vom modellbasierten bzw. zielgerichteten System kontrolliert. Die beiden Paradigmen erfassen aber letztlich nur die modulierenden, übergeordneten Instanzen und können daher nur eingeschränkt Rückschlüsse auf das individuelle Ausmaß von habituellem und modellfreiem Verhalten geben. Insbesondere habituelle Verhaltensstrategien werden durch das verwendete Devaluationsparadigma nur unzureichend erfasst bzw. induziert.
- Die beiden vorgestellten Konstrukte sind in sich unterschiedlich konfiguriert: Während ein Zusammenspiel zwischen modellbasiert und modellfreien System im Sinne einer Balance oder auch gegenseitigen Inhibition vermutet wird, nimmt man für das zielgerichtete und habituelle System an, dass diese eher unabhängig voneinander arbeiten (Dichotomie, doppelte Dissonanz).
- Deckeneffekt: Alle Probanden zeigten eine insgesamt gute kognitive Performance und ein Überwiegen zielgerichteter bzw. modellbasiertes Verhaltensmechanismen. Dadurch bestand keine große Varianz innerhalb der Stichprobe. Bestehende Unterschiede in der Ausprägung von habituellen und modellfreien Strategien konnten in der Folge womöglich durch die verwendeten Paradigmen nicht gut induziert bzw. nicht ausreichend differenziert werden.

4.2. Limitationen der Studie

Bereits während der Pilotierungsphase des Devaluationsparadigmas, aber insbesondere nach Beendigung der Studie und der Auswertung der Ergebnisse ergaben sich einige Faktoren, die die Aussagekraft des Studienergebnisses möglicherweise schmälern. Diese Limitationen sollen im Folgenden näher beleuchtet werden und müssen bei der Interpretation der Befunde berücksichtigt werden.

Als einer der größten Limitationsfaktoren ist sicherlich die geringe Fallzahl anzumerken (13 bzw. nach Ausschluss eines Ausreißers 12 Probanden in den Korrelationsanalysen). Hinzu kommt, dass es sich bei den Ergebnissen um Korrelationen handelt, folglich sind Aussagen über die kausalen Zusammenhänge nicht möglich. Die Befunde dieser Studie konnten jedoch in anderen Studien an

deutlich größeren Stichproben (N = 90) repliziert werden (Gillan et al., 2015; Sjoerds et al., 2016).

Weitere limitierende Faktoren ergeben sich insbesondere mit Hinblick auf das experimentelle Design des Devaluationsparadigmas. Bereits in der Pilotierungsphase wurde deutlich, dass bei der Auswahl der Getränke und verabreichten Speisen diverse Faktoren zu beachten sind. Im Gegensatz zur Hauptstudie wurden in der Pilotierungsphase zunächst Schokoladenmilch und Orangensaft als appetitive Verstärker verabreicht. Es zeigte sich allerdings nach Analyse der Verhaltensdaten keine selektive Devaluation, d.h. das Auswahlverhalten für den hoch probabilistischen Stimulus nahm sowohl in der devaluierten als auch in der nicht devaluierten Bedingung ab – unabhängig davon, welches Lebensmittel übersättigt worden war. Als Ursache ist die mögliche sensorische Ähnlichkeit zwischen Schokoladenmilch und Orangensaft zu diskutieren. Beides sind süße Getränke, sodass durch die Sättigung möglicherweise keine selektive Devaluation für die spezifischen Reize, sondern lediglich eine Sättigung bezüglich „süß“ stattfand. Zudem berichteten einige Probanden, dass die Aufeinanderfolge von Schokoladenmilch und Orangensaft als unangenehm empfunden wurde. In zukünftigen Studien sollte folglich darauf geachtet werden, dass die dargebotenen Lebensmittel und Getränke in ihrer sensorischen Spezifität gut voneinander zu unterscheiden sind. Daher entschieden wir uns schließlich für die Verwendung von Schokoladenmilch und Tomatensaft in der Hauptstudie.

Zudem sollten die verwendeten Nahrungsmittel von den Teilnehmern als ausreichend angenehm empfunden bzw. wirklich gemocht werden. Dieses Kriterium erwies sich insbesondere bei der Rekrutierung alkoholabhängiger Probanden als problematisch: Es ließen sich nicht genug alkoholabhängige Probanden finden, die Schokoladenmilch als ausreichend angenehm einstufen. Dies ist möglicherweise auf den Einfluss des Suchtmittels Alkohol zurückzuführen, welches in Form von Wein, Bier und Schnaps eher eine herbe Geschmacksrichtung darstellt. Außerdem schienen jene alkoholabhängigen Patienten, die an der Pilotphase teilnahmen, im Gegensatz zu den gesunden Probanden deutlich mehr Schwierigkeiten mit dem experimentellen Aufbau und dem Verständnis des Devaluationsparadigmas zu haben. Aus diesem Grund musste die ursprüngliche Idee, auch alkoholabhängige Patienten in die Studie einzuschließen, letztlich verworfen werden. Daher wurden letztlich nur gesunde Testpersonen untersucht. Aber auch unter den gesunden

Probanden erwies sich die Rekrutierung als schwierig, da solche Personen gefunden werden mussten, die sowohl Tomatensaft als auch Schokoladenmilch als angenehm empfanden.

Bisweilen stellte sich heraus, dass manche Probanden bei der Rekrutierung zwar für die jeweiligen Lebensmittel angegeben hatten, diese gerne zu sich zu nehmen, im Augenblick der Testung allerdings trotz Nahrungskarenz keinen Appetit mehr darauf hatten. Ob die Probanden jedoch wirklich die Nahrungskarenz einhielten, wurde nicht objektiv durch bestimmte Messmethoden wie z.B. Blutzuckermessung überprüft. Es erscheint daher umso wichtiger, die Einschätzung von Hunger und dem augenblicklichen Appetit auf die dargebotenen Produkte auch kurz vor der eigentlichen Testung zu erfassen um mögliche Einflussfaktoren zu berücksichtigen und ggf. zu diesem Zeitpunkt bereits die Untersuchung abzubrechen.

Darüber hinaus sollten überwiegend männliche Probanden in Folgestudien eingeschlossen werden. Zum einen, weil Frauen insgesamt häufiger eine restriktives Essverhalten zeigen (Wardle et al., 2004), auch wenn sie angeben, derzeit keine diätetische Lebensweise zu führen. Andererseits sollte mit Blick auf mögliche weitere Untersuchungen an alkoholabhängigen Patienten berücksichtigt werden, dass der Anteil männlichen Patienten in dieser klinischen Population höher ist (Wilsnack et al., 2009).

Außerdem berichteten einige Probanden, dass sie den Geschmack der Getränke aufgrund der Verabreichung über die Schläuche mitunter als verändert oder unangenehm empfanden. Ebenso sorgte die kurzzeitige Abfolge der Getränke zuweilen für Geschmacksirritationen. Es zeigte sich zudem, dass die Präferenz dafür, was nach der nächtlichen bzw. morgendlichen Nahrungskarenz gern gegessen wird, vereinzelt die Durchführung und den Effekt der Devaluation erschwerte (z.B. lieber ein Brötchen anstatt Tomatensuppe essen). Kritisch ist insbesondere anzumerken, dass sich die Verwendung von Schokoladenpudding und Tomatensuppe nur als eingeschränkt dienlich erwies, da sie sich deutlich in ihrem kalorischen Gehalt unterschieden und folglich zu unterschiedlichen Sättigungseffekten führten. Vereinzelt erwies es sich sogar als schwierig, die Testpersonen zum Essen zu animieren, wenn sie bereits nach wenigen Löffeln die Devaluation abrechnen wollten, obwohl sie noch nicht ausreichend gesättigt waren. Im Gegensatz dazu berichteten einige Teilnehmer, dass sie kaum genug von Schokolade bekommen konnten, sodass in diesem Fall, trotz dem Verzehr großer Mengen Pudding, Schokolade als

nicht weniger angenehm eingestuft wurde. Schließlich mussten aufgrund dieser fehlenden Sättigungseffekte zwei Probanden ausgeschlossen werden. Die Höhe der Konsummenge während der Devaluation scheint folglich ein kritisches, vor allem aber auch ein sehr individuelles Maß zu sein. Letzteres wird durch die Befunde von Gillan et al. (2015) untermauert: Sie ließen bei ihren Analysen den Konsum als Kovariable in die lineare Regression mit einfließen. Es zeigte sich hierbei, dass der Konsum weder die Devaluationssensitivität voraussagte, noch mit dem Parameter für modellbasiertes Verhalten interagierte.

Alles in allem scheint der experimentelle Aufbau mithilfe von Lebensmitteln für die Probanden einen kritischen Punkt beim Devaluationsparadigma darzustellen. Positiv ist hierbei allerdings anzumerken, dass in dieser Studie eine selektive Devaluation induziert werden konnte. Dies spiegelte sich in einem reduzierten Antwortverhalten für denjenigen Stimulus wider, der nach der Devaluation mit dem übersättigten und folglich weniger angenehmen Getränk assoziiert war. Eine alternative und möglicherweise praktikablere Umsetzung der Devaluationsaufgabe bei Verwendung von Lebensmitteln als Verstärker verfolgten Horstmann et al. (2015): Hier wurde den Probanden während der instrumentellen Lernaufgabe die jeweilige Belohnung lediglich in Form eines Bildes angezeigt, wobei ihnen im Voraus mitgeteilt wurde, dass sie am Ende des ersten Durchgangs die gewonnenen Snacks ausgehändigt bekommen würden. Auf diese Weise könnten mögliche Geschmacksirritationen vermieden werden. Auch der komplexe apparative Aufbau mit Pumpen und Schläuchen würde überflüssig werden, was wiederum zweckmäßiger für Testung mit funktioneller Bildgebung wäre. Ein Nachteil hierbei wäre allerdings, dass kein sensorisches Signal beim fMRT erfasst werden würde. Des Weiteren wurde bei der erwähnten Studie eine „implizite“ Devaluation realisiert. Das bedeutet, dass den Probanden nicht direkt gesagt wurde, dass sie nun so viel von den Lebensmitteln zu sich nehmen sollten, bis sie gesättigt wurden. Stattdessen erhielten die Teilnehmer nach der Lernaufgabe freien Zugang zu demjenigen Snack, der devaluiert werden sollte. Dabei lud man sie unter dem Deckmantel einer Titelgeschichte dazu ein, sich so oft wie möglich an den Snacks zu bedienen, während sie 30 Minuten lang (unbeobachtet) einen Naturfilm schauten, zu dessen Inhalt sie im Nachhinein in Form einer vermeintlichen Befragung Auskunft geben sollten. Diese Form der Devaluation umgeht eventuell die Schwierigkeit, Probanden zum Essen animieren zu müssen, da sie einerseits ausreichend Zeit bekommen und sich zudem unbeobachtet fühlen, was

die Menge der aufgenommenen Snacks angeht. So rückt das Prozedere der Devaluation aus dem Fokus der Teilnehmer und wirft möglicherweise weniger Fragen bei ihnen auf. In ihren so genannten Hawthorne-Experimenten stellten Roethlisberger et al. (1939) nämlich fest, dass Teilnehmer eines Experimentes sich anders verhalten, wenn sie Kenntnis darüber haben, Teil einer Untersuchung zu sein. Dies wiederum kann zu einer Verfälschung der Ergebnisse führen, da die Probanden sich möglicherweise nicht „natürlich“ verhalten. So kann es passieren, dass sie sich sozial erwünscht verhalten oder auch Aufgaben auf solch eine Weise erledigen, wie sie – vermeintlich – von ihnen verlangt wird. Dies wiederum hat Folgen für die interne Validität der Untersuchungsergebnisse. In Bezug auf das Devaluationsparadigma würde dies bedeuten, dass die Probanden weniger Snacks essen, als sie eigentlich bis zur Erreichung eines ausreichenden Sättigungsgefühls benötigen würden.

Da die Besonderheit bei dem hier verwendeten Devaluationsparadigma darin lag, dass Nahrung sowohl als instrumenteller Verstärker als auch bei der Devaluation verwendet wurde (Grund hierfür ist, dass Nahrung einen primären Verstärker darstellt), sollte auch das Gewicht als mutmaßlicher Einflussfaktor diskutiert werden. Womöglich kann ein erhöhter BMI ein Marker für veränderte Essgewohnheiten sein (z.B. Höhe Kalorienzufuhr, Präferenz für bestimmte Lebensmittel, Grad des Sättigungseffektes), selbst wenn bislang keine Essstörung, wie sie im Eating Attitudes Tests erfasst wird, nachzuweisen ist. Zudem konnte gezeigt werden, dass adipöse Probanden ihr Verhalten weniger gut adaptieren können, wenn es zu einer Änderung des Anreizwertes des Lebensmittels kommt (Horstmann et al., 2015). Aus behavioraler Sicht zeigten sich bei der erwähnten Studie eine geringe Sensitivität gegenüber einer Devaluation bei höherem BMI sowie vermehrt habituelle Verhaltensmuster. Es muss jedoch hinzugefügt werden, dass keiner der von der hiesigen Untersuchung ausgeschlossenen Probanden einen deutlich erhöhten BMI aufwies (durchschnittlich 28.8 kg/m^2 , $SD = 5,72$). Allerdings zeigte sich in den Analysen auch kein Zusammenhang zwischen Devaluationssensitivität und dem BMI (Spearman Rangkorrelation $r = 0.471$, $p = 0.104$).

Aufgrund der Tatsache, dass zwei der Probanden in der Devaluationsaufgabe nicht das Lernkriterium erfüllten, also in den letzten zehn Durchgängen der instrumentellen Aufgabe nicht häufig genug den hoch probabilistischen Stimulus eines jeden Paares ausgewählt hatten, ist zu diskutieren, ob die verwendeten abstrakten Bilder zu

ähnlich waren und so das Lernen der entsprechenden Zuordnung zu den Belohnungen erschwert wurde. So berichteten einige Teilnehmer, dass sie manche Stimuli gar verwechselten und demnach nicht die eigentlich von ihnen gewünschten Belohnung erhielten. Nichtsdestotrotz ist hierbei zu bedenken, dass eine zu einfache Gestaltung der instrumentellen Lernaufgabe wiederum andere limitierende Faktoren mit sich bringt. Einfach gehaltene Stimuli können zu einer geringeren Aufmerksamkeit bzw. einem geringeren Arousal bei den Probanden führen, was zur Folge hätte, dass es nur zu einer unzureichenden Aktivierung des eigentlich vorhandenen Leistungsvermögens kommt (Yerkes and Dodson, 1908). Erhalten die Probanden zudem mehr Zeit für die instrumentelle Lernaufgabe, werden also mehr Durchgänge zugelassen, wird dadurch möglicherweise ein Überlernen induziert, was wiederum habituelle Verhaltensmuster fördern würde (Tricomi et al., 2009).

Ein interessanter Aspekt wäre außerdem, ob die Zuordnung zwischen Stimulus und Belohnung eher explizit oder implizit gelernt wurde. Dabei würde Letzteres bedeuten, dass die Probanden die Zuordnung gelernt haben, ohne im Nachhinein tatsächlich sagen zu können, welcher Stimulus zu welchem Getränk führte. Über diese postexperimentelle Abfrage könnte gegebenenfalls auch die Frage geklärt werden, ob das Lernen der Zuordnung für Nicht-Lerner tatsächlich zu schwierig war oder ob womöglich andere, bereits genannte Einflussfaktoren eine Rolle spielten. Darüber hinaus stellte für manche Probanden auch die Auftretenswahrscheinlichkeit der Belohnungen eine Schwierigkeit dar. Diese lag für den hoch probabilistischen Stimulus bei $p = 0.5$ für Schokoladenmilch oder Tomatensaft sowie bei $p = 0.25$ für Früchtetee. Folglich war die Wahrscheinlichkeit, bei der Auswahl dieses Stimulus eine Belohnung zu erhalten bei 75%, bei dem anderen, niedrig probabilistischen Stimulus hingegen war dies nur zu 25% der Fall. Da sich diese Wahrscheinlichkeit jedoch nicht auf einzelne Blöcke (z.B. von zehn Durchgängen) bezog, sondern über die gesamte instrumentelle Lernaufgabe randomisiert wurde, kam es mitunter dazu, dass einige Teilnehmer am Anfang zufällig häufiger keine Belohnung erhielten. Zwar handelte es sich hierbei nicht um einen systematischen Fehler, weil alle Testpersonen unter den gleichen Bedingungen die Aufgabe absolvierten. Allerdings führten mehr „Nullrunden“ zu Beginn der instrumentellen Lernaufgabe vielleicht auch zu mehr Verunsicherung und so unter Umständen zu erschwerten Lernbedingungen.

Letztlich sollte auch ein möglicher Reihenfolgeeffekt in dieser Studie diskutiert werden. Alle Probanden absolvierten das gesamte Experiment immer nach dem gleichen Ablauf: Devaluationsparadigma, *two-step*-Aufgabe und neuropsychologische Testung. Es erfolgte also kein Ausbalancieren der Reihenfolge. Dabei ist zu bedenken, dass das Leistungsvermögen eines Probanden im Verlauf eines Experimentes abnimmt. Je länger eine Untersuchung andauert, desto mehr ermüdet der Proband, was wiederum eine Einbuße seiner Performance bedeutet. In der Tat gaben einige der Probanden in der mündlichen Nachbefragung an, dass sie insbesondere nach der *two-step*-Aufgabe erschöpft und müde waren. Zum einen lag dies an der ständigen Fokussierung auf den Computerbildschirm, zum anderen aber auch an der anspruchsvollen Aufgabenstellung. Dies wiederum führte möglicherweise dazu, dass die Probanden in der abschließenden neuropsychologischen Testung ihr volles Leistungspotential nicht abrufen konnten und es zu einer Verzerrung der Untersuchungsergebnisse kam. Dies könnte auch zum fehlenden Zusammenhang zwischen den Ergebnissen der neuropsychologischen Untersuchungen und dem gezeigten modellbasierten Verhalten beigetragen haben. Es sollte daher zukünftig auf ausreichend Pausen sowie ein Ausbalancieren der einzelnen Teilabschnitte der Untersuchung geachtet werden. Gegebenenfalls sollte auch erwogen, die verschiedenen Paradigmen an unterschiedlichen Untersuchungstagen durchzuführen. Denn mit einer Gesamtdauer von ca. drei Stunden je Proband handelte es sich bei der vorliegenden Studie doch um eine recht lange Untersuchungszeit, welche von einigen Probanden, trotz vorheriger Aufklärung über den Untersuchungsablauf, nicht gut toleriert wurde.

4.3. Bedeutung der Ergebnisse für die Untersuchung veränderter Lern- und Verhaltensmechanismen bei Alkoholabhängigkeit

Die Ergebnisse sowohl dieser Arbeit als auch anderer, genannter Studien lassen darauf schließen, dass ein Zusammenhang zwischen zielgerichteten und modellbasierten Strategien der Verhaltenskontrolle besteht – nicht nur auf behavioraler, sondern auch auf neuronaler Ebene. Letztlich lassen sich zielgerichtete Verhaltensstrategie sogar als eine Facette des modellbasierten Systems beschreiben (Friedel et al., 2014; Voon et al., 2017). Habituelle und modellfreie Verhaltensstrategien scheinen hingegen unterschiedliche Aspekte des eingangs erwähnten „retrospektiven“ Systems der Verhaltenskontrolle abzubilden. Was

bedeutet dies nun für die Untersuchung veränderter Strategien der Verhaltenskontrolle bei Alkoholabhängigkeit? Bei der Konzeptualisierung von Alkoholabhängigkeit als Folge dysfunktionaler Lern- und Verhaltensstrategien wird davon ausgegangen, dass es zu einer Verschiebung von zielgerichtetem bzw. modellbasiertem Verhalten hinzu habituellen bzw. modellfreien Strategien kommt. Zur Untersuchung der zugrundeliegenden, veränderten Mechanismen wird dabei auf die verschiedenen Operationalisierungen der theoretischen Konstrukte (habituell vs. zielgerichtet bzw. modellfrei vs. modellbasiert) zurückgegriffen, die bereits an gesunden Probanden untersucht und etabliert worden sind. Womöglich sind die Ergebnisse dieser Untersuchungen und Schlussfolgerungen aber aufgrund der lediglich teilweise vorhandenen Konstruktvalidität der unterschiedlichen Paradigmen nur bedingt vergleichbar, da ja habituelle und modellfreie Strategien unterschiedliche Prozesse darzustellen scheinen – also genau jene Verhaltensmuster, die bei der Alkoholabhängigkeit von besonderem Interesse sind. Folglich sollten zur Untersuchung klinischer Populationen jeweils solche Paradigmen verwendet und entwickelt werden, die diese Verhaltensstrategien gut erfassen und abbilden können. So konnte in der Tat durch eine Abwandlung der Devaluationsaufgabe, adaptiert nach de Wit et al. (2009), bei alkoholabhängigen Patienten ein Überwiegen habituellen Verhaltens im Sinne einer auf der Assoziation zwischen Stimulus und Antwortverhalten basierenden Strategie (S-R-Assoziation) nachgewiesen werden (Sjoerds et al., 2013).

Zudem ist in Hinblick auf Studien mit Patienten zu beachten, dass diese Population mitunter über limitierte kognitive Fähigkeiten und Ressourcen verfügt. Dies spielt in zweierlei Hinsicht eine wichtige Rolle: Zum einen müssen die Paradigmen derart adaptiert und konzipiert werden, dass sie von Patienten verstanden und angemessen bewältigt werden können. Andererseits stellen eingeschränkte kognitive Kapazitäten (Schad et al., 2014; Smittenaar et al., 2013) und Stresserleben (Otto et al., 2013; Schwabe and Wolf, 2009) modulierende Faktoren bei der individuellen Balance von Strategien der Verhaltenskontrolle dar. Folglich sollten bei Untersuchungen an klinischen Populationen diese Parameter ebenso berücksichtigt und bei der Entstehung sowie Aufrechterhaltung von Alkoholabhängigkeit diskutiert werden (siehe hierzu auch Hägele et al., 2014).

Darüber hinaus offenbarte sich in der vorliegenden Arbeit aber ein weiteres Problem: Selbst wenn es möglich ist, die verschiedenen Strategien der Verhaltenskontrolle

mithilfe entsprechender Paradigmen messbar zu machen, so geben sie keinen Aufschluss darüber, welche Ursachen den veränderten Verhaltensmechanismen letztlich zugrunde liegen. Umso wichtiger erscheint die Weiterentwicklung der Paradigmen um einen noch differenzierten Einblick in die zugrundeliegenden Mechanismen zu gewinnen. Hier könnte sich abermals die funktionelle Bildgebung als hilfreiches Instrument erweisen.

Die Antwort auf die Frage danach, wie Lern- und Verhaltensmechanismen bei Alkoholabhängigkeit verändert sind und inwiefern Defizite oder ein Übermaß an Verhaltensstrategien bestehen, bestimmen letztlich auch die Entwicklung der therapeutischen Ansätze zur Behandlung ebenjener Erkrankungen: Muss der therapeutische Schwerpunkt darauf gelegt werden, die „moderierende Instanz“ und folglich alternative Denkansätze und Verhaltensmechanismen zu stärken? Oder sollte es eher Ziel sein, habituelle oder modellfreie Verhaltensmuster abzuschwächen? Ist möglicherweise eine Kombination aus beiden Strategien am erfolgversprechendsten? Wenn es in Zukunft möglich sein wird, durch weitere Untersuchungen solche Variablen zu erfassen, die z.B. modellbasiertes Verhalten stärken und eine Ausprägung habitueller Verhaltensmuster verhindern (Gillan et al., 2015), könnten entsprechende Fertigkeiten auch trainiert werden. Diese könnte einen weiteren Baustein in der Therapie der Alkoholabhängigkeit darstellen. Die bisherigen Befunde aus den Forschungsbestrebungen haben immerhin schon zu einigen neuen Ansätzen geführt. So konnte in einer ersten Studie eine signifikante Reduktion der Rückfallrate bei alkoholabhängigen Patienten durch das Überwinden automatisierter Tendenzen mithilfe eines entsprechenden Verhaltenstrainings gezeigt werden (Eberl et al., 2013). Bei diesem so genannten Approach Avoidance Task (AAT) werden alkoholabhängige Patienten darin trainiert, alkoholassoziierte Reize mehr von sich wegzuschieben um dadurch automatisierte, habituelle Annäherungstendenzen zu überwinden.

4.4. Ausblick

Das Hauptresultat der Studie liegt darin, dass mittels Korrelationsanalysen ein positiver Zusammenhang zwischen modellbasiertem Antwortverhalten während einer sequentiellen Entscheidungsaufgabe und dem Ausmaß für zielgerichteten Verhaltens im Sinne einer hohen Devaluationssensitivität im Rahmen eines Devaluationsparadigmas gezeigt werden konnte. Diese Ergebnisse liefern erste

evidente Hinweise darauf, dass beiden Operationalisierungen letztlich auf einem gemeinsamen Mechanismus zu beruhen scheinen. Folglich spräche dies für die Konstruktvalidität der beiden verwendeten Paradigmen. Einschränkend ist allerdings zu sagen, dass dieser Zusammenhang nicht für habituelles Verhalten, als Ausdruck einer verminderten Devaluationssensitivität, und modellfreie Strategien nachgewiesen werden konnte. Mögliche Erklärungsansätze wurden im Rahmen dieser Arbeit ausführlich diskutiert. Auch wenn die Ergebnisse letztlich auf dem Hintergrund der aufgeführten Limitationen kritisch zu betrachten sind, so sollte auch die Relevanz der Ergebnisse nicht unterschätzt werden. Denn im Rahmen dieser Studie wurde erstmals die in der Wissenschaft seit vielen Jahren zwar vermutete, bisher aber nicht gesicherte Konstruktvalidität der beiden Paradigmen untersucht. Die Überlegungen zu möglichen Ursachen der vorliegenden Ergebnisse, aber auch die kritische Diskussion von Negativergebnissen bieten indes die Möglichkeit, weitere Forschungsbestrebungen voranzutreiben. Erste Folgestudien konnten die Ergebnisse zwischenzeitlich bereits replizieren und haben so zu einigen Weiterentwicklungen der Operationalisierungen geführt (Gillan et al., 2015; Sjoerds et al., 2013).

Dennoch ist in Anbetracht der dargelegten divergierenden Befunde nur folgerichtig zu betonen, dass für zukünftige Studien Paradigmen entwickelt werden sollten, die in der Lage sind, die verschiedenen Strategien der Verhaltenskontrolle zu erfassen und voneinander zu trennen – vor allem in Hinblick auf habituelle und modellfreie Verhaltensstrategien. Insbesondere beim Devaluationsparadigma wäre eine Anpassung in solch einer Weise wünschenswert, dass es sich als praktikabel erweist, auch in Hinblick auf die Umsetzung bei der Verwendung der funktionellen Bildgebung. Einige Möglichkeiten der Adaptation wurden in der vorliegenden Arbeit unter Berücksichtigung von anderen Studiendesigns diskutiert (Gillan et al., 2015; Horstmann et al., 2015; Sjoerds et al., 2016, 2013).

Ein weiteres Augenmerk sollte zudem daraufgelegt werden, das experimentelle Design so zu konzipieren, dass es möglich wird, eine potentiell bestehende übergeordnete Instanz, also einen Vermittler zwischen den verschiedenen Systemen der Verhaltenskontrolle zu erfassen. Denn selbst wenn man zukünftig dazu im Stande sein wird, die verschiedenen Verhaltensstrategien abzubilden und zu analysieren, so liefert dies noch keine Erklärung dafür, ob und wie die zugrundeliegenden neuronalen Systeme mit- oder gegeneinander arbeiten.

Letzten Endes sind die Rückschlüsse in Hinblick auf die spezifischen, zugrundeliegenden Mechanismen der Verhaltenskontrolle, die allein auf der Analyse und Interpretation von Verhaltensdaten beruhen, stets in ihrer Aussagekraft begrenzt. Ein wichtiges Instrument stellt deshalb die funktionelle Bildgebung (fMRT) dar, die es ermöglicht, die beteiligten neuronalen Strukturen zu identifizieren sowie mögliche Überschneidungen und Interaktionsprozesse zu analysieren. Dabei ist vor allem das Verständnis der Integration der verschiedenen Informationen und Strategien in das große neuronale Netzwerk von Lern- und Entscheidungsprozessen, welche letztlich das individuelle Verhalten prägen und ausbilden, von besonderem Interesse.

Von besonderer Relevanz wäre auch die Testung an klinischen Populationen, um auf diese Weise ein besseres Verständnis über dysfunktionale Lern- und Verhaltensmechanismen als Komponente bei der Entstehung und Aufrechterhaltung psychischer Erkrankungen zu erlangen. Solche Untersuchungen sollten auch unter Berücksichtigung der eingeschränkten kognitiven Fähigkeiten bei dieser Population erfolgen, bieten aber letzten Endes die Chance der Entwicklung neuer therapeutischer und pharmakologischer Ansätze bei der Behandlung jener Erkrankungen.

5. Literaturverzeichnis

- Adams, C.D., Dickinson, A., 1981. Instrumental responding following reinforcer devaluation. *Q. J. Exp. Psychol. Sect. B* 33, 109–121.
<https://doi.org/10.1080/14640748108400816>
- Addis, D.R., Schacter, D.L., 2011. The hippocampus and imagining the future: where do we stand? *Front. Hum. Neurosci.* 5, 173.
<https://doi.org/10.3389/fnhum.2011.00173>
- Alvares, G.A., Balleine, B.W., Guastella, A.J., 2014. Impairments in goal-directed actions predict treatment response to cognitive-behavioral therapy in social anxiety disorder. *PloS One* 9, e94778.
<https://doi.org/10.1371/journal.pone.0094778>
- Balleine, B.W., 2005. Neural bases of food-seeking: affect, arousal and reward in corticostriatolimbic circuits. *Physiol. Behav.* 86, 717–730.
<https://doi.org/10.1016/j.physbeh.2005.08.061>
- Balleine, B.W., Dickinson, A., 1998. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37, 407–419.
- Balleine, B.W., O'Doherty, J.P., 2010. Human and rodent homologues in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacol. Off. Publ. Am. Coll. Neuropsychopharmacol.* 35, 48–69. <https://doi.org/10.1038/npp.2009.131>
- Belin, D., Jonkman, S., Dickinson, A., Robbins, T.W., Everitt, B.J., 2009. Parallel and interactive learning processes within the basal ganglia: relevance for the understanding of addiction. *Behav. Brain Res.* 199, 89–102.
<https://doi.org/10.1016/j.bbr.2008.09.027>
- Bellman, R., 1957. *Dynamic Programming*. Princeton University Press, Princeton, NJ.
- Berridge, K.C., Robinson, T.E., 2016. Liking, wanting, and the incentive-sensitization theory of addiction. *Am. Psychol.* 71, 670–679.
<https://doi.org/10.1037/amp0000059>
- Bouton, M.E., 2002. Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. *Biol. Psychiatry* 52, 976–986.
- Carter, B.L., Tiffany, S.T., 1999. Meta-analysis of cue-reactivity in addiction research. *Addict. Abingdon Engl.* 94, 327–340.

- Childress, A.R., Ehrman, R.N., Wang, Z., Li, Y., Sciortino, N., Hakun, J., Jens, W., Suh, J., Listerud, J., Marquez, K., Franklin, T., Langleben, D., Detre, J., O'Brien, C.P., 2008. Prelude to passion: limbic activation by “unseen” drug and sexual cues. *PloS One* 3, e1506.
<https://doi.org/10.1371/journal.pone.0001506>
- Childress, A.R., Mozley, P.D., McElgin, W., Fitzgerald, J., Reivich, M., O'Brien, C.P., 1999. Limbic activation during cue-induced cocaine craving. *Am. J. Psychiatry* 156, 11–18. <https://doi.org/10.1176/ajp.156.1.11>
- Cohen, N.J., Eichenbaum, H., 1995. *Memory, Amnesia, and the Hippocampal System*. A Bradford Book.
- Collins, A.G.E., Frank, M.J., 2012. How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *Eur. J. Neurosci.* 35, 1024–1035.
<https://doi.org/10.1111/j.1460-9568.2011.07980.x>
- Colwill, R.M., Rescorla, R.A., 1988. The role of response-reinforcer associations increases throughout extended instrumental training. *Anim. Learn. Behav.* 16, 105–111. <https://doi.org/10.3758/BF03209051>
- Cools, R., 2011. Dopaminergic control of the striatum for high-level cognition. *Curr. Opin. Neurobiol.* 21, 402–407. <https://doi.org/10.1016/j.conb.2011.04.002>
- Cooper, R.P., Ruh, N., Mareschal, D., 2014. The goal circuit model: a hierarchical multi-route model of the acquisition and control of routine sequential action in humans. *Cogn. Sci.* 38, 244–274. <https://doi.org/10.1111/cogs.12067>
- Corbit, L.H., Balleine, B.W., 2005. Double dissociation of basolateral and central amygdala lesions on the general and outcome-specific forms of pavlovian-instrumental transfer. *J. Neurosci. Off. J. Soc. Neurosci.* 25, 962–970.
<https://doi.org/10.1523/JNEUROSCI.4507-04.2005>
- Dalley, J.W., Everitt, B.J., Robbins, T.W., 2011. Impulsivity, compulsivity, and top-down cognitive control. *Neuron* 69, 680–694.
<https://doi.org/10.1016/j.neuron.2011.01.020>
- Danner, U.N., Aarts, H., de Vries, N.K., 2007. Habit formation and multiple means to goal attainment: repeated retrieval of target means causes inhibited access to competitors. *Pers. Soc. Psychol. Bull.* 33, 1367–1379.
<https://doi.org/10.1177/0146167207303948>

- Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P., Dolan, R.J., 2011. Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69, 1204–1215. <https://doi.org/10.1016/j.neuron.2011.02.027>
- Daw, N.D., Niv, Y., Dayan, P., 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8, 1704–1711. <https://doi.org/10.1038/nn1560>
- Daw, N.D., O'Doherty, J.P., 2014. Chapter 21 - Multiple Systems for Value Learning, in: Glimcher, P.W., Fehr, E. (Eds.), *Neuroeconomics (Second Edition)*. Academic Press, San Diego, pp. 393–410. <https://doi.org/10.1016/B978-0-12-416008-8.00021-8>
- de Wit, S., Corlett, P.R., Aitken, M.R., Dickinson, A., Fletcher, P.C., 2009. Differential engagement of the ventromedial prefrontal cortex by goal-directed and habitual behavior toward food pictures in humans. *J. Neurosci. Off. J. Soc. Neurosci.* 29, 11330–11338. <https://doi.org/10.1523/JNEUROSCI.1639-09.2009>
- de Wit, S., van de Vijver, I., Ridderinkhof, K.R., 2014. Impaired acquisition of goal-directed action in healthy aging. *Cogn. Affect. Behav. Neurosci.* 14, 647–658. <https://doi.org/10.3758/s13415-014-0288-5>
- de Wit, S., Watson, P., Harsay, H.A., Cohen, M.X., van de Vijver, I., Ridderinkhof, K.R., 2012. Corticostriatal connectivity underlies individual differences in the balance between habitual and goal-directed action control. *J. Neurosci. Off. J. Soc. Neurosci.* 32, 12066–12075. <https://doi.org/10.1523/JNEUROSCI.1088-12.2012>
- Deserno, Lorenz, Huys, Q.J.M., Boehme, R., Buchert, R., Heinze, H.-J., Grace, A.A., Dolan, R.J., Heinz, A., Schlagenhauf, F., 2015. Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. *Proc. Natl. Acad. Sci. U. S. A.* 112, 1595–1600. <https://doi.org/10.1073/pnas.1417219112>
- Deserno, L., Wilbertz, T., Reiter, A., Horstmann, A., Neumann, J., Villringer, A., Heinze, H.-J., Schlagenhauf, F., 2015. Lateral prefrontal model-based signatures are reduced in healthy individuals with high trait impulsivity. *Transl. Psychiatry* 5, e659. <https://doi.org/10.1038/tp.2015.139>

- Dezfouli, A., Balleine, B.W., 2013. Actions, action sequences and habits: evidence that goal-directed and habitual action control are hierarchically organized. *PLoS Comput. Biol.* 9, e1003364. <https://doi.org/10.1371/journal.pcbi.1003364>
- Dezfouli, A., Balleine, B.W., 2012. Habits, action sequences and reinforcement learning. *Eur. J. Neurosci.* 35, 1036–1051. <https://doi.org/10.1111/j.1460-9568.2012.08050.x>
- Dickinson, A., 1985. Animal intelligence - Actions and habits: the development of behavioural autonomy. *Phil Trans R Soc Lond B* 308, 67–78. <https://doi.org/10.1098/rstb.1985.0010>
- Dickinson, A., Balleine, B., 1994. Motivational control of goal-directed action. *Anim. Learn. Behav.* 22, 1–18. <https://doi.org/10.3758/BF03199951>
- Dolan, R.J., Dayan, P., 2013. Goals and habits in the brain. *Neuron* 80, 312–325. <https://doi.org/10.1016/j.neuron.2013.09.007>
- Doll, B.B., Simon, D.A., Daw, N.D., 2012. The ubiquity of model-based reinforcement learning. *Curr. Opin. Neurobiol.* 22, 1075–1081. <https://doi.org/10.1016/j.conb.2012.08.003>
- Eberl, C., Wiers, R.W., Pawelczack, S., Rinck, M., Becker, E.S., Lindenmeyer, J., 2013. Approach bias modification in alcohol dependence: do clinical effects replicate and for whom does it work best? *Dev. Cogn. Neurosci.* 4, 38–51. <https://doi.org/10.1016/j.dcn.2012.11.002>
- Effertz, T., 2020. Die volkswirtschaftlichen Kosten von Alkohol- und Tabakkonsum in Deutschland, in: Deutsche Hauptstelle für Suchtfragen (Ed.), *DHS Jahrbuch Sucht 2020*. Pabst Science Publishers, Lengerich.
- Everitt, B.J., Robbins, T.W., 2005. Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat. Neurosci.* 8, 1481–1489. <https://doi.org/10.1038/nn1579>
- Faure, A., Haberland, U., Condé, F., El Massioui, N., 2005. Lesion to the nigrostriatal dopamine system disrupts stimulus-response habit formation. *J. Neurosci. Off. J. Soc. Neurosci.* 25, 2771–2780. <https://doi.org/10.1523/JNEUROSCI.3894-04.2005>
- Flagel, S.B., Akil, H., Robinson, T.E., 2009. Individual differences in the attribution of incentive salience to reward-related cues: Implications for addiction. *Neuropharmacology* 56 Suppl 1, 139–148. <https://doi.org/10.1016/j.neuropharm.2008.06.027>

- Friedel, E., Koch, S.P., Wendt, J., Heinz, A., Deserno, L., Schlagenhauf, F., 2014. Devaluation and sequential decisions: linking goal-directed and model-based behavior. *Front. Hum. Neurosci.* 8, 587. <https://doi.org/10.3389/fnhum.2014.00587>
- Garbusow, M., Friedel, E., Sebold, M., Beck, A., Heinz, A., Smolka, M.N., 2013. Wege in die Abhängigkeit. *SUCHT* 59, 187–199. <https://doi.org/10.1024/0939-5911.a000254>
- Garbusow, M., Schad, D.J., Sebold, M., Friedel, E., Bernhardt, N., Koch, S.P., Steinacher, B., Kathmann, N., Geurts, D.E.M., Sommer, C., Müller, D.K., Nebe, S., Paul, S., Wittchen, H.-U., Zimmermann, U.S., Walter, H., Smolka, M.N., Sterzer, P., Rapp, M.A., Huys, Q.J.M., Schlagenhauf, F., Heinz, A., 2016. Pavlovian-to-instrumental transfer effects in the nucleus accumbens relate to relapse in alcohol dependence. *Addict. Biol.* 21, 719–731. <https://doi.org/10.1111/adb.12243>
- Garner, D.M., Olmsted, M.P., Bohr, Y., Garfinkel, P.E., 1982. The eating attitudes test: psychometric features and clinical correlates. *Psychol. Med.* 12, 871–878.
- Gillan, C.M., Kosinski, M., Whelan, R., Phelps, E.A., Daw, N.D., 2016. Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *eLife* 5. <https://doi.org/10.7554/eLife.11305>
- Gillan, C.M., Morein-Zamir, S., Kaser, M., Fineberg, N.A., Sule, A., Sahakian, B.J., Cardinal, R.N., Robbins, T.W., 2014. Counterfactual processing of economic action-outcome alternatives in obsessive-compulsive disorder: further evidence of impaired goal-directed behavior. *Biol. Psychiatry* 75, 639–646. <https://doi.org/10.1016/j.biopsych.2013.01.018>
- Gillan, C.M., Otto, A.R., Phelps, E.A., Daw, N.D., 2015. Model-based learning protects against forming habits. *Cogn. Affect. Behav. Neurosci.* 15, 523–536. <https://doi.org/10.3758/s13415-015-0347-6>
- Gillan, C.M., Robbins, T.W., 2014. Goal-directed learning and obsessive-compulsive disorder. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 369. <https://doi.org/10.1098/rstb.2013.0475>
- Gläscher, J., Daw, N., Dayan, P., O'Doherty, J.P., 2010. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* 66, 585–595. <https://doi.org/10.1016/j.neuron.2010.04.016>

- Gremel, C.M., Costa, R.M., 2013. Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. *Nat. Commun.* 4, 2264. <https://doi.org/10.1038/ncomms3264>
- Hägele, C., Friedel, E., Kienast, T., Kiefer, F., 2014. How do we “learn” addiction? Risk factors and mechanisms getting addicted to alcohol. *Neuropsychobiology* 70, 67–76. <https://doi.org/10.1159/000364825>
- Heinz, A., 2017. *New Understanding of Mental Disorders*. MIT University Press Group Ltd, Cambridge, Massachusetts.
- Heinz, A., Batra, A., Scherbaum, N., Gouzoulis-Mayfrank, E., Lutz, U., Mörsen, C., Reuter, J., Müller, C., Charlet, K., 2012. *Neurobiologie der Abhängigkeit: Grundlagen und Konsequenzen für Diagnose und Therapie von Suchterkrankungen*, 1. Aufl. ed. Kohlhammer W., GmbH, Stuttgart.
- Heinz, A., Friedel, E., Beck, A., 2013. Alkoholabhängigkeit als gelerntes Verhaltensmuster. *SUCHT* 59, 183–184. <https://doi.org/10.1024/0939-5911.a000252>
- Horga, G., Maia, T.V., Marsh, R., Hao, X., Xu, D., Duan, Y., Tau, G.Z., Gianiello, B., Wang, Z., Kangarlu, A., Martinez, D., Packard, M.G., Peterson, B.S., 2015. Changes in corticostriatal connectivity during reinforcement learning in humans. *Hum. Brain Mapp.* 36, 793–803. <https://doi.org/10.1002/hbm.22665>
- Horstmann, A., Dietrich, A., Mathar, D., Pössel, M., Villringer, A., Neumann, J., 2015. Slave to habit? Obesity is associated with decreased behavioural sensitivity to reward devaluation. *Appetite* 87, 175–183. <https://doi.org/10.1016/j.appet.2014.12.212>
- Hull, C.L., 1943. *Principles of Behavior: An Introduction to Behavior Theory*. Appleton-Century-Crofts.
- Huys, Q.J.M., Eshel, N., O’Nions, E., Sheridan, L., Dayan, P., Roiser, J.P., 2012. Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Comput. Biol.* 8, e1002410. <https://doi.org/10.1371/journal.pcbi.1002410>
- Kierkegaard, S., 1923. *Die Tagebücher 1834 - 1855*. In zwei Bänden ausgewählt und übersetzt von Theodor Haecker. Brenner-Verlag, Innsbruck.
- Killcross, S., Coutureau, E., 2003. Coordination of actions and habits in the medial prefrontal cortex of rats. *Cereb. Cortex N. Y. N* 1991 13, 400–408.

- Lee, D., 2013. Decision Making: from Neuroscience to Psychiatry. *Neuron* 78, 233–248. <https://doi.org/10.1016/j.neuron.2013.04.008>
- Lee, S.W., Shimojo, S., O'Doherty, J.P., 2014. Neural computations underlying arbitration between model-based and model-free learning. *Neuron* 81, 687–699. <https://doi.org/10.1016/j.neuron.2013.11.028>
- Liljeholm, M., Tricomi, E., O'Doherty, J.P., Balleine, B.W., 2011. Neural correlates of instrumental contingency learning: differential effects of action-reward conjunction and disjunction. *J. Neurosci. Off. J. Soc. Neurosci.* 31, 2474–2480. <https://doi.org/10.1523/JNEUROSCI.3354-10.2011>
- Montague, P.R., Dayan, P., Sejnowski, T.J., 1996. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci. Off. J. Soc. Neurosci.* 16, 1936–1947.
- Morris, J.C., Heyman, A., Mohs, R.C., Hughes, J.P., van Belle, G., Fillenbaum, G., Mellits, E.D., Clark, C., 1989. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* 39, 1159–1165.
- Morris, R.W., Quail, S., Griffiths, K.R., Green, M.J., Balleine, B.W., 2015. Corticostriatal control of goal-directed action is impaired in schizophrenia. *Biol. Psychiatry* 77, 187–195. <https://doi.org/10.1016/j.biopsych.2014.06.005>
- Niv, Y., Daw, N.D., Joel, D., Dayan, P., 2007. Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology (Berl.)* 191, 507–520. <https://doi.org/10.1007/s00213-006-0502-4>
- O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H., Dolan, R.J., 2003. Temporal difference models and reward-related learning in the human brain. *Neuron* 38, 329–337.
- O'Keefe, J., Nadel, L., 1978. *The Hippocampus as a Cognitive Map*. Oxford University Press, Oxford : New York.
- Otto, A.R., Raio, C.M., Chiang, A., Phelps, E.A., Daw, N.D., 2013. Working-memory capacity protects model-based learning from stress. *Proc. Natl. Acad. Sci. U. S. A.* 110, 20941–20946. <https://doi.org/10.1073/pnas.1312011110>
- Pawlow, I.P., 1927. *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex*. Oxford University Press: Humphrey Milford.

- Reiter, A.M.F., Deserno, L., Kallert, T., Heinze, H.-J., Heinz, A., Schlagenhauf, F., 2016. Behavioral and Neural Signatures of Reduced Updating of Alternative Options in Alcohol-Dependent Patients during Flexible Decision-Making. *J. Neurosci. Off. J. Soc. Neurosci.* 36, 10935–10948. <https://doi.org/10.1523/JNEUROSCI.4322-15.2016>
- Reiter, A.M.F., Heinze, H.-J., Schlagenhauf, F., Deserno, L., 2017. Impaired Flexible Reward-Based Decision-Making in Binge Eating Disorder: Evidence from Computational Modeling and Functional Neuroimaging. *Neuropsychopharmacol. Off. Publ. Am. Coll. Neuropsychopharmacol.* 42, 628–637. <https://doi.org/10.1038/npp.2016.95>
- Roethlisberger, F.J., Dickson, W.J., Wright, H.A., Company, W.E., 1939. *Management and the Worker: An Account of a Research Program Conducted by the Western Electric Company, Hawthorne Works, Chicago.* Harvard University Press.
- Schad, D.J., Jünger, E., Sebold, M., Garbusow, M., Bernhardt, N., Javadi, A.-H., Zimmermann, U.S., Smolka, M.N., Heinz, A., Rapp, M.A., Huys, Q.J.M., 2014. Processing speed enhances model-based over model-free reinforcement learning in the presence of high working memory functioning. *Front. Psychol.* 5, 1450. <https://doi.org/10.3389/fpsyg.2014.01450>
- Schmidt, K.-H., Metzler, P., 1992. *Wortschatztest: WST.* Beltz, Weinheim.
- Schoenmakers, T.M., de Bruin, M., Lux, I.F.M., Goertz, A.G., Van Kerkhof, D.H.A.T., Wiers, R.W., 2010. Clinical effectiveness of attentional bias modification training in abstinent alcoholic patients. *Drug Alcohol Depend.* 109, 30–36. <https://doi.org/10.1016/j.drugalcdep.2009.11.022>
- Schultz, W., Dayan, P., Montague, P.R., 1997. A neural substrate of prediction and reward. *Science* 275, 1593–1599.
- Schwabe, L., Wolf, O.T., 2009. Stress prompts habit behavior in humans. *J. Neurosci. Off. J. Soc. Neurosci.* 29, 7191–7198. <https://doi.org/10.1523/JNEUROSCI.0979-09.2009>
- Sebold, M., Deserno, L., Nebe, Stephan, Nebe, Stefan, Schad, D.J., Garbusow, M., Hägele, C., Keller, J., Jünger, E., Kathmann, N., Smolka, M.N., Smolka, M., Rapp, M.A., Schlagenhauf, F., Heinz, A., Huys, Q.J.M., 2014. Model-based and model-free decisions in alcohol dependence. *Neuropsychobiology* 70, 122–131. <https://doi.org/10.1159/000362840>

- Sebold, M., Nebe, S., Garbusow, M., Guggenmos, M., Schad, D.J., Beck, A., Kuitunen-Paul, S., Sommer, C., Frank, R., Neu, P., Zimmermann, U.S., Rapp, M.A., Smolka, M.N., Huys, Q.J.M., Schlagenhauf, F., Heinz, A., 2017. When Habits Are Dangerous: Alcohol Expectancies and Habitual Decision Making Predict Relapse in Alcohol Dependence. *Biol. Psychiatry* 82, 847–856. <https://doi.org/10.1016/j.biopsych.2017.04.019>
- Sjoerds, Z., de Wit, S., van den Brink, W., Robbins, T.W., Beekman, A.T.F., Penninx, B.W.J.H., Veltman, D.J., 2013. Behavioral and neuroimaging evidence for overreliance on habit learning in alcohol-dependent patients. *Transl. Psychiatry* 3, e337. <https://doi.org/10.1038/tp.2013.107>
- Sjoerds, Z., Dietrich, A., Deserno, L., de Wit, S., Villringer, A., Heinze, H.-J., Schlagenhauf, F., Horstmann, A., 2016. Slips of Action and Sequential Decisions: A Cross-Validation Study of Tasks Assessing Habitual and Goal-Directed Action Control. *Front. Behav. Neurosci.* 10, 234. <https://doi.org/10.3389/fnbeh.2016.00234>
- Smittenaar, P., FitzGerald, T.H.B., Romei, V., Wright, N.D., Dolan, R.J., 2013. Disruption of Dorsolateral Prefrontal Cortex Decreases Model-Based in Favor of Model-free Control in Humans. *Neuron* 80, 914–919. <https://doi.org/10.1016/j.neuron.2013.08.009>
- Sutton, R.S., 1988. Learning to predict by the methods of temporal differences. *Mach. Learn.* 3, 9–44. <https://doi.org/10.1007/BF00115009>
- Sutton, R.S., Barto, A.G., 1998. *Reinforcement Learning: An Introduction*. The MIT Press, Cambridge, Mass.
- Swanson, L.W., 2000. Cerebral hemisphere regulation of motivated behavior. *Brain Res.* 886, 113–164.
- Talmi, D., Seymour, B., Dayan, P., Dolan, R.J., 2008. Human pavlovian-instrumental transfer. *J. Neurosci. Off. J. Soc. Neurosci.* 28, 360–368. <https://doi.org/10.1523/JNEUROSCI.4028-07.2008>
- Thorn, C.A., Atallah, H., Howe, M., Graybiel, A.M., 2010. Differential dynamics of activity changes in dorsolateral and dorsomedial striatal loops during learning. *Neuron* 66, 781–795. <https://doi.org/10.1016/j.neuron.2010.04.036>
- Thorndike, E.L., 1927. The Law of Effect. *Am. J. Psychol.* 39, 212–222. <https://doi.org/10.2307/1415413>

- Thorndike, E.L., 1898. *Animal Intelligence: An Experimental Study of the Associative Processes in Animals*. Columbia University.
- Tolman, E.C., 1948. Cognitive maps in rats and men. *Psychol. Rev.* 55, 189–208.
- Tombaugh, T.N., 2004. Trail Making Test A and B: normative data stratified by age and education. *Arch. Clin. Neuropsychol. Off. J. Natl. Acad. Neuropsychol.* 19, 203–214. [https://doi.org/10.1016/S0887-6177\(03\)00039-8](https://doi.org/10.1016/S0887-6177(03)00039-8)
- Tricomi, E., Balleine, B.W., O'Doherty, J.P., 2009. A specific role for posterior dorsolateral striatum in human habit learning. *Eur. J. Neurosci.* 29, 2225–2232. <https://doi.org/10.1111/j.1460-9568.2009.06796.x>
- Valentin, V.V., Dickinson, A., O'Doherty, J.P., 2007. Determining the neural substrates of goal-directed learning in the human brain. *J. Neurosci. Off. J. Soc. Neurosci.* 27, 4019–4026. <https://doi.org/10.1523/JNEUROSCI.0564-07.2007>
- van der Meer, M., Kurth-Nelson, Z., Redish, A.D., 2012. Information processing in decision-making systems. *Neurosci. Rev. J. Bringing Neurobiol. Neurol. Psychiatry* 18, 342–359. <https://doi.org/10.1177/1073858411435128>
- Voon, V., Derbyshire, K., Rück, C., Irvine, M.A., Worbe, Y., Enander, J., Schreiber, L.R.N., Gillan, C., Fineberg, N.A., Sahakian, B.J., Robbins, T.W., Harrison, N.A., Wood, J., Daw, N.D., Dayan, P., Grant, J.E., Bullmore, E.T., 2015. Disorders of compulsivity: a common bias towards learning habits. *Mol. Psychiatry* 20, 345–352. <https://doi.org/10.1038/mp.2014.44>
- Voon, V., Reiter, A., Sebold, M., Groman, S., 2017. Model-Based Control in Dimensional Psychiatry. *Biol. Psychiatry* 82, 391–400. <https://doi.org/10.1016/j.biopsych.2017.04.006>
- Wardle, J., Haase, A.M., Steptoe, A., Nillapun, M., Jonwutiwes, K., Bellisle, F., 2004. Gender differences in food choice: the contribution of health beliefs and dieting. *Ann. Behav. Med. Publ. Soc. Behav. Med.* 27, 107–116.
- Watkins, E.R., Nolen-Hoeksema, S., 2014. A habit-goal framework of depressive rumination. *J. Abnorm. Psychol.* 123, 24–34. <https://doi.org/10.1037/a0035540>
- Wechsler, D., 1981. *WAIS-R manual: Wechsler adult intelligence scale-revised*. Psychological Corporation, New York.

- Wiers, C.E., Stelzel, C., Park, S.Q., Gawron, C.K., Ludwig, V.U., Gutwinski, S., Heinz, A., Lindenmeyer, J., Wiers, R.W., Walter, H., Bermpohl, F., 2014. Neural correlates of alcohol-approach bias in alcohol addiction: the spirit is willing but the flesh is weak for spirits. *Neuropsychopharmacol. Off. Publ. Am. Coll. Neuropsychopharmacol.* 39, 688–697. <https://doi.org/10.1038/npp.2013.252>
- Wilsnack, R.W., Wilsnack, S.C., Kristjanson, A.F., Vogeltanz-Holm, N.D., Gmel, G., 2009. Gender and alcohol consumption: patterns from the multinational GENACIS project. *Addict. Abingdon Engl.* 104, 1487–1500. <https://doi.org/10.1111/j.1360-0443.2009.02696.x>
- Wittchen, H.-U., Fydrich, T., Zaudig, M., 1997. Strukturiertes Klinisches Interview für DSM- IV. SKID-I und SKID-II. Achse I: Psychische Störungen / Achse II: Persönlichkeitsstörungen. Hogrefe.
- Wittchen, H.-U., Hoyer, J., 2011. Was ist Klinische Psychologie? Definitionen, Konzepte und Modelle, in: *Klinische Psychologie & Psychotherapie*, Springer-Lehrbuch. Springer, Berlin, Heidelberg, pp. 11–25. https://doi.org/10.1007/978-3-642-13018-2_1
- Wood, W., Quinn, J.M., Kashy, D.A., 2002. Habits in everyday life: thought, emotion, and action. *J. Pers. Soc. Psychol.* 83, 1281–1297.
- Wood, W., Rünger, D., 2016. Psychology of Habit. *Annu. Rev. Psychol.* 67, 289–314. <https://doi.org/10.1146/annurev-psych-122414-033417>
- Wrase, J., Schlagenhauf, F., Kienast, T., Wüstenberg, T., Bermpohl, F., Kahnt, T., Beck, A., Ströhle, A., Juckel, G., Knutson, B., Heinz, A., 2007. Dysfunction of reward processing correlates with alcohol craving in detoxified alcoholics. *NeuroImage* 35, 787–794. <https://doi.org/10.1016/j.neuroimage.2006.11.043>
- Wunderlich, K., Smittenaar, P., Dolan, R.J., 2012. Dopamine Enhances Model-Based over Model-Free Choice Behavior. *Neuron* 75, 418–424. <https://doi.org/10.1016/j.neuron.2012.03.042>
- Yerkes, R.M., Dodson, J.D., 1908. The relation of strength of stimulus to rapidity of habit-formation. *J. Comp. Neurol. Psychol.* 18, 459–482. <https://doi.org/10.1002/cne.920180503>
- Yin, H.H., Knowlton, B.J., 2006. The role of the basal ganglia in habit formation. *Nat. Rev. Neurosci.* 7, 464–476. <https://doi.org/10.1038/nrn1919>

- Yin, H.H., Knowlton, B.J., Balleine, B.W., 2004. Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur. J. Neurosci.* 19, 181–189.
- Yin, H.H., Mulcare, S.P., Hilário, M.R.F., Clouse, E., Holloway, T., Davis, M.I., Hansson, A.C., Lovinger, D.M., Costa, R.M., 2009. Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill. *Nat. Neurosci.* 12, 333–341. <https://doi.org/10.1038/nn.2261>
- Yin, H.H., Ostlund, S.B., Knowlton, B.J., Balleine, B.W., 2005. The role of the dorsomedial striatum in instrumental conditioning. *Eur. J. Neurosci.* 22, 513–523. <https://doi.org/10.1111/j.1460-9568.2005.04218.x>

Eidesstattliche Versicherung

Ich, Jean Wendt, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: „Messung verschiedener Formen der Verhaltenskontrolle und deren Relevanz bei Alkoholabhängigkeit“ selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren/innen beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) werden von mir verantwortet.

Ich versichere ferner, dass ich die in Zusammenarbeit mit anderen Personen generierten Daten, Datenauswertungen und Schlussfolgerungen korrekt gekennzeichnet und meinen eigenen Beitrag sowie die Beiträge anderer Personen korrekt kenntlich gemacht habe (siehe Anteilserklärung). Texte oder Textteile, die gemeinsam mit anderen erstellt oder verwendet wurden, habe ich korrekt kenntlich gemacht.

Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem Erstbetreuer, angegeben sind. Für sämtliche im Rahmen der Dissertation entstandenen Publikationen wurden die Richtlinien des ICMJE (International Committee of Medical Journal Editors; www.icmje.org) zur Autorenschaft eingehalten. Ich erkläre ferner, dass ich mich zur Einhaltung der Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis verpflichte.

Weiterhin versichere ich, dass ich diese Dissertation weder in gleicher noch in ähnlicher Form bereits an einer anderen Fakultät eingereicht habe.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§§156, 161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

Datum

Unterschrift

Anteilerklärung an etwaigen erfolgten Publikationen

Jean Wendt hatte Anteil an der folgenden Publikation:

Friedel, E., Koch, S.P., Wendt, J., Heinz, A., Deserno, L., Schlagenhaut, F., 2014. Devaluation and sequential decisions: linking goal-directed and model-based behavior. *Front. Hum. Neurosci.* 8, 587. <https://doi.org/10.3389/fnhum.2014.00587>

Beitrag im Einzelnen:

- Rekrutierung und Screening (Einschluss/ Ausschluss) der Probanden
- Erarbeitung des experimentellen Designs
- Versuchsplanung und -organisation
- Durchführung und Auswertung der neuropsychologischen Testung
- Datenerhebung inklusive Aufklärungen bei allen Probanden sowohl in der Pilotstudie, als auch in der Hauptstudie
- Datenablage, Datenqualitätssicherung und Datenorganisation
- Verwaltung der Probandenkasse und Abrechnung von Probandengelder

Für die Studie wurde das *two-Step-Experiment* genutzt – ein innerhalb der Forschergruppe bereits etabliertes Paradigma, welches in Anlehnung an Daw et. al (2011) von meinen Co-Autoren erarbeitet wurden. Das Devaluationsparadigma wurde in Anlehnung an das jenes von Valentin et. al (2007) gestaltet und programmiert, wobei sich die präsentierten Stimuli und verwendeten Verstärker unterschieden. Basierend auf der Grundidee, die Konstruktvalidität der beiden Paradigmen zu überprüfen, wurde durch mich das experimentelle Design und die Versuchsanordnung geplant und umgesetzt.

Sowohl die statistischen Auswertungen, als auch die Abbildungen und der Fließtext des genannten Zeitschriftenartikels wurden durch meine Co-Autoren realisiert, basieren aber vollständig auf den von mir erhobenen Daten. Dabei flossen auch meine Ideen bezüglich der Ergebnisdiskussion mit ein, wodurch es nach der Pilotstudie auch zu Anpassungen des experimentellen Designs in der Hauptstudie kam.

Da Teile dieser Monographie auf bereits veröffentlichten Daten basieren, sind dementsprechend auch einige Abbildungen dieser Arbeit an jene der genannten Publikation angelehnt und entsprechend gekennzeichnet. Darüber hinaus wurden in der vorliegenden Arbeit bereits veröffentlichte Daten um weitere Berechnungen ergänzt und die entsprechenden Ergebnisse aufgeführt (u.a. nicht devaluierte und neutrale Bedingung beim Devaluationsparadigma, Haupteffekt Transition, Korrelation Devaluations sensitivität und p-Belohnung, Korrelation Gewinn und p-Interaktion). Trotz der bereits vorhandenen Daten und Ergebnisse im Rahmen der erwähnten Publikation, wurde die statistische Auswertung für alle erhobenen Daten im Rahmen der Monographie erneut meinerseits durchgeführt, da auch der inhaltliche Schwerpunkt (Fokus auf klinischen Kontext) ein anderer als in genannter Publikation ist. Aufgrund dieser Berechnungen, welche auch zur einheitlichen Anpassung von Nachkommastellen durchgeführt wurden, waren einige Ergebnisse geringfügig abweichend von den bereits veröffentlichten Daten. Die Pilotstudie wurde nicht veröffentlicht, die in dieser Monographie genannten Daten und Ergebnisse wurde eigens durch mich erhoben und ausgewertet.

Unterschrift, Datum und Stempel des betreuenden Hochschullehrers

Unterschrift der Doktorandin

Lebenslauf

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Publikationsliste

Friedel, E., Koch, S.P., Wendt, J., Heinz, A., Deserno, L., Schlagenhauf, F., 2014. Devaluation and sequential decisions: linking goal-directed and model-based behavior. *Front. Hum. Neurosci.* 8, 587. <https://doi.org/10.3389/fnhum.2014.00587>

Garbusow, M., Schad, D.J., Sommer, C., Jünger, E., Sebold, M., Friedel, E., Wendt, J., Kathmann, N., Schlagenhauf, F., Zimmermann, U.S., Heinz, A., Huys, Q.J.M., Rapp, M.A., 2014. Pavlovian-to-instrumental transfer in alcohol dependence: a pilot study. *Neuropsychobiology* 70, 111–121. <https://doi.org/10.1159/000363507>

Danksagung

An dieser Stelle möchte ich allen beteiligten Personen meinen großen Dank aussprechen, die mich bei der Anfertigung und Vollendung meiner Dissertation unterstützt haben:

Zunächst danke ich meinem Doktorvater Prof. Dr. med. Florian Schlagenhaut für die Möglichkeit, in seinem Arbeitskreis meine Promotionsarbeit anfertigen und einen Teil meiner Forschung an einer so renommierten Einrichtung wie dem Max-Planck-Institut absolvieren zu dürfen. Darüber hinaus danke ich ihm für die vielen produktiven Gespräche, seine Diskussionsbereitschaft, seine hilfreiche und konstruktive Kritik sowie die langjährige Betreuung.

Mein besonderer Dank gilt zudem Dr. rer. nat. Stefan Paul Koch für die warmherzige Zusammenarbeit und die enorme Unterstützung bei der Umsetzung und Durchführung der gesamten Arbeit. Mit beispiellosem Engagement und einer unermüdlichen Gelassenheit hat er es mir ermöglicht, meine Ideen in Bezug auf das Thema in die Tat umzusetzen, auf dem Themenfeld der Statistik den Überblick zu behalten und auch bei aufkommenden Problemen stets einen kühlen Kopf zu bewahren. Dies hat wesentlich zum Gelingen meiner Dissertation beigetragen.

Außerdem möchte ich mich bei Dr. med. Dr. phil. Dipl.-Psych. Eva Friedel von ganzem Herzen bedanken, die mich im Rahmen meiner Promotion nicht nur mit fachlichem Rat und wertvollen Anregungen, sondern auch mit aufmunternden Worten und Zuspruch unterstützt hat.

Nicht unerwähnt lassen will ich zudem meine langjährige Arbeitsgruppe, allen voran meine Kolleginnen Dr. rer. nat. Dipl.-Psych. Maria Garbusow und Dr. rer. nat. Dipl.-Psych. Maria Sebold. Über viele Jahre in einer solch angenehmen Arbeitsatmosphäre promovieren und neben dem Studium arbeiten zu können, war nicht nur eine besondere Erfahrung, die viele schöne Erinnerungen hinterlässt, sondern hat mir auch in schwierigen Zeiten geholfen, mein Ziel nicht aus den Augen zu verlieren.

Lea Prengel danke ich für die wertvolle Unterstützung und hilfreichen Anmerkungen bei den statistischen Auswertungen.

Für die mühevollen Arbeit des Korrekturlesens und die vielen wertvollen Hinweise möchte ich zudem Jana Loeding und Dr. phil. Juliane Fleischer einen großen Dank aussprechen.

Ganz besonders möchte ich mich bei meinen Freunden und langjährigen Wegbegleitern bedanken. Speziell bin ich für ihre Geduld und Rücksichtnahme, für ihr Zuhören zu jeder Tageszeit und ihre uneingeschränkte Hilfe, egal ob in Wort oder Tat, dankbar. Ohne diesen emotionalen Beistand wären mir die Stolpersteine auf dem Weg zur Promotion vermutlich als unüberwindbare Mauern erschienen und das Zustandekommen dieser Arbeit für mich nicht denkbar gewesen.