Psychologisch orientierte Kategorisierung in der kognitiven Robotik mit dem Hierarchischen Dirichlet Prozess

Thomas Glassen

Vollständiger Abdruck der von der Fakultät für Luft- und Raumfahrttechnik der Universität der Bundeswehr München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Gutachter / Gutachterin:

1. Univ.-Prof. Dr. Verena Nitsch
2. Univ.-Prof. Dr. Timo von Oertzen
3. Univ.-Prof. Dr. Stefan Pickl

Die Dissertation wurde am 30.05.2017 bei der Universität der Bundeswehr München eingereicht und durch die Fakultät für Luft- und Raumfahrttechnik am 20.11.2017 angenommen. Die mündliche Prüfung fand am 15.12.2017 statt.

Danksagung

Mein Dank geht zunächst an meine Betreuerin Prof. Dr. Verena Nitsch, welche mir über die gesamte Dauer der Promotion mit Rat und Tat zur Seite stand und mich kontinuierlich unterstützte und ermutigte meine Zwischenergebnisse zu publizieren. Auch Prof. Dr. Timo von Oertzen, welcher mir trotz neuer Arbeitsstelle genügend Zeit gewährte diese Arbeit abzuschließen, möchte ich für seine Unterstützung danken. Insbesondere die vielen lehrreichen Diskussionen über Dirichlet Prozesse haben mir in hohem Maße geholfen den mathematischen Teil der Arbeit in der bestehenden Form auszuarbeiten. Abschließend aber nicht zuletzt möchte ich Annika Johnsen danken, für die wertvolle motivationale Unterstützung und die große Entlastung während arbeitsintensiver Zeiten.

Abstract (English)

Hierarchical Dirichlet Process mixture models (HDPMM) have recently been introduced not only in cognitive psychology but also in cognitive robotics. It was postulated in the context of categorization research that HDPMMs not only combine the strengths of all previous rational categorization procedures, but also unify the two prominent theories of categorization, the exemplar and prototype view, in a common categorization model. Consequently, findings were successfully modeled, for which the striking properties of an HDPMM, the dynamic complexity adaptation of the model to existing data, and the ability of cluster sharing appeared to be important key mechanisms. Researchers from cognitive robotics interpreted these successes as evidence for a good model of human categorization. In fact, HDPMMs are, however, a group of rational models that have so far hardly been explored. Hence, in this thesis the HDPMM introduced in cognitive psychology by Griffiths, Canini, Sanborn and Navarro (2007) is extensively evaluated. The study investigates the predictive power of the model regarding seven classical findings on human categorization, which are serious challenges for many of the current categorization models. It is demonstrated that two modifications to the model of Griffiths et al. (2007) are sufficient to successfully predict the majority of the findings. Furthermore, the HDPMM is compared with a prominent and well-established categorization model, the Supervised and Unsupervised Stratified Adaptive Incremental Network (SUSTAIN) with regard to data fitting and model flexibility, in which the modified model of Griffiths and colleagues can assert itself in the majority of the experiments. Finally, further possibilities for improvement, especially with regard to the application in the robotics context, are discussed.

Abstract (German)

Hierarchische Dirichlet Prozess-Mischmodelle (HDPMM) haben in neuerer Zeit nicht nur in der Kognitionspsychologie sondern auch in der kognitiven Robotik Einzug gehalten. Dabei wurde im Kontext der Kategorisierungsforschung postuliert, dass HDPMMs nicht nur die Stärken aller bisherigen rationalen Kategorisierungsverfahren in sich vereinen, sondern auch ein Zusammenführung der Exemplar- und Prototypensichtweise in einem gemeinsamen Kategorisierungsmodell darstellen. Konsequenterweise wurden Befunde erfolgreich modelliert, für welche die markanten Eigenschaften eines HDPMM, die dynamische Komplexitätsanpassung des Modells an vorliegende Daten und die Fähigkeit des Clustersharings, als wichtiger Schlüsselmechanismen hervortraten. Forscher aus der kognitiven Robotik nahmen das zum Anlass das Verfahren als ein gutes Modell menschlicher Kategorisierung zu bewerten. Tatsächlich handelt es sich bei HDPMMs jedoch um eine bisher kaum untersuchte Gruppe von rationalen Modellen. In der vorliegenden Arbeit wird deshalb das in der Kognitionspsychologie eingeführte HDPMM von Griffiths, Canini, Sanborn und Navarro (2007) ausgiebig evaluiert. Gegenstand der Untersuchung ist die Vorhersagekraft des Modells bezüglich sieben klassischer Befunde zur menschlichen Kategorisierungsleistung, welche für viele der derzeitigen Modelle aus der Kategorisierungsforschung ernsthafte Herausforderungen darstellen. Dabei wird demonstriert, dass zwei Modifikationen am Modell von Griffiths et al. (2007) ausreichen, um die Mehrzahl der Befunde erfolgreich vorherzusagen. Weiterhin wird das HDPMM mit einem prominenten und etablierten Modell, dem Supervised and Unsupervised Stratified Adaptive Incremental Network (SUSTAIN), hinsichtlich des Datenfittings und der Modellflexibilität verglichen, wobei sich das modifizierte Modell von Griffiths und Kollegen in der Mehrzahl der Experimente behaupten kann. Es werden abschließend weitere Verbesserungsmöglichkeiten insbesondere auch hinsichtlich der Anwendung im Robotikkontext diskutiert.

Inhaltsverzeichnis

1 Einführung....6
1.1 Das Forschungsfeld Robotik: Ein kurzer Überblick....6
1.2 Robotik und Kognitionswissenschaften....8
1.3 Klassische Informationsverarbeitungstheorien aus Sicht der Embodied-Cognitive-Science....9
1.4 Robotik und Bayesian-Cognitive-Science....10
1.5 Das Forschungsfeld der Kategorisierung: Ein Überblick....12
1.6 Forschungsvorhaben....15
2 Methodik....18
2.1 Der Satz von Bayes....18
2.2 Bernoulli-, Beta-, und Dirichletverteilungen....19
2.3 Der Dirichlet Prozess....21
2.4 Der Hierarchische Dirichlet Prozess....26
2.5 Das URM....28
2.6 Sampling mit dem URM....30
2.7 Durchführung einer Simulation mit dem URM....33
2.8 Bestimmung des repräsentativen Clusterings....36
2.9 SUSTAIN....39
2.10 Gegenüberstellung von URM, SUSTAIN und konkurrierenden Kategorisierungsverfahren 44
2.11 Die Parameteroptimierung....49
2.12 Zugrundeliegende Maße der Modellvergleiche....50
2.13 Statistische Auswertungen....51
3 Experimente....55
3.1 Genereller Ablauf....55
3.2 Experiment 1: Lernschwierigkeit von sechs Kategorisierungsregeln....57
3.2.1 Modellierungsergebnisse für Experiment 1....59
3.3 Experiment 2: Identifikations vs. Klassifikationsschwierigkeit bei komplexen Stimuli....67
3.3.1 Modellierungsergebnisse für Experiment 2....69
3.4 Experimente 3 und 4: Inferenz- vs. Klassifikationsschwierigkeit bei linear separablen und nicht-separablen Kategorien....72
3.4.1 Modellierungsergebnisse für die Experimente 3 und 4....76
3.5 Experimente 5 und 6: Inferenzakkuratheit bei niedrig und hoch interkorrelierenden Stimulusattributen....78
3.5.1 Modellierungsergebnisse für die Experimente 5 und 6....83
3.6 Experiment 7: Unsupervidiertes Sortieren von Stimuli in zwei Gruppen....85
3.6.1 Modellierungsergebnisse für Experiment 7....87
4 Diskussion....89
4.1 Bewertung der Modellierungsergebnisse....90
4.2 Theoretische Aspekte des Verfahrens....92
4.3 Technische Aspekte des Verfahrens....94
4.4 Konklusion....95
Literaturverzeichnis....96

Abkürzungsverzeichnis

ALCOVE	Attention Learning Covering Map
CRF	Chinese Restaurant Franchise
CRP	Chinese Restaurant Prozess
DP	Dirichlet Prozess
DPMM	Dirichlet Prozess-Mischmodell
HDP	Hierarchischer Dirichlet Prozess
HDPMM	Hierarchisches Dirichlet Prozess-Mischmodell
LSAP	Linear-Sum-Assignment-Problem
MHDP	Multimodaler Hierarchischer Dirichlet Prozess
MRMC	More Rational Model of Categorization
PPD	Posterior-Predictive-Distribution
PSDA	Posterior-Sampling-by-Direct-Assignment
RMC	Rational Model of Categorization
SSE	Summe der Fehlerquadrate
SUSTAIN	Supervised and Unsupervised Stratified Adaptive Incremental Network
URM	Unifying Rational Model
VP	Versuchsperson/en

1 Einführung

Mit der vorliegenden Arbeit soll ein Betrag zu den interdisziplinären Forschungsfeldern der kognitiven Robotik und der kognitiven Psychologie und hier im Speziellen zur Kategorisierungsforschung geleistet werden, indem ein neueres, noch kaum untersuchtes Kategorisierungsmodell evaluiert und verbessert wird. Da die Verbindung zwischen (humanoider) Robotik und Psychologie nicht auf den ersten Blick ersichtlich ist, soll zunächst ein kurzer Überblick über Robotik im Allgemeinen und anschließend schrittweise die interdisziplinären Zusammenhänge aufgezeigt werden, welche das untersuchte Modell in einem größeren Rahmen einbetten. Es folgt ein sehr ausführlicher Methodikteil, in welchem dem Leser das doch recht anspruchsvolle Modell hinsichtlich seiner Funktionsweise näher gebracht wird. Nach dem Hauptteil, in welchem das Modell an sieben Experimenten evaluiert wird, folgt eine abschließende Beurteilung der Modellierungsgüte sowie des Verfahrens aus theoretischer und technischer Sicht mit dem Ziel, den Wert des Verfahrens für die angesprochenen interdisziplinären Forschungsfelder aufzuzeigen.

1.1 Das Forschungsfeld Robotik: Ein kurzer Überblick

Auch wenn das vom tschechischen Schriftsteller Karel Capek eingeführte Wort „Robot“ ursprünglich die künstlich erschaffenen menschenähnlichen Zwangsarbeiter in seinem Theaterstück R.U.R. bezeichnete (Capek & Shenef, 2016), sind unter dem heutigen Wort „Roboter“ nicht nur anthropomorphe autonome Konstruktionen zu verstehen. So gibt es heutzutage neben bipedalen humanoiden Robotern (Kaneko, Harada, & Kanehiro, 2008) auch radbetriebene (Lafaye, Gouaillier, & Wieber, 2014), schwimmende (Suzumori, Endo, Kato, & Suzuki, 2007), fliegende (Gurdan et al., 2007) und statische Varianten (Brogårdh, 2007). Roboter unterscheiden sich aber nicht nur in ihrer Fortbewegungsart. Tatsächlich ist die Bandbreite der existierenden Systeme, welche unter dem Begriff „Roboter“ firmieren, derart groß, dass eine generelle Klassifikation nach Typen ein schwieriges Unterfangen ist (Angeles, 2014). Poole (1989) konnte Roboter beispielsweise noch nach Anzahl der Freiheitsgrade, Art der Bewegung, Plattform, Energiequelle, Intelligenz und Anwendungsgebiet unterscheiden. Dabei handelte es sich jedoch beim Großteil der damals existierenden Roboter um industrielle (statische) Varianten. Fast 25 Jahre später ist die Diversität nach Angeles (2014) bestenfalls nach drei Gesichtspunkten zu unterteilen: Die Größe, die Art der

Funktion und das Anwendungsgebiet. So existieren beispielsweise die Forschungsgebiete Nano- und Microrobotics, welche sich mit der Entwicklung und Erforschung von Robotern im Milli-, Micro- und Nanometerbereich beschäftigen (Abbott, Nagy, Beyeler, & Nelson, 2007; Dong & Nelson, 2007; Hill, Amodeo, Joseph, & Patel, 2008), die Fachbereiche Laboratory- und Underwaterobotics, welche sich auf Roboter für die Automatisierung von Labortätigkeiten wie z.B. chemische Analysen bzw. Roboter für Unterwasserarbeiten spezialisiert haben (Little, 1993; Yuh, 2000), oder Cloud- und BEAM-Robotics, welche eine zentralisierte Informationsverarbeitung von Robotern mittels Cloud-basierter Technologie bzw. die Steuerung von Robotern ausschließlich über analoge Schaltungen forcieren (Lorencik & Sincak, 2013; Pransky, 2014). Auch wenn sich die vielen Forschungsnischen in ihren Interessengebieten deutlich unterscheiden, lassen sich jedoch sieben Funktionsbereiche bei Robotern identifizieren, welche übergeordnete Forschungsschwerpunkte innerhalb der Robotik darstellen: *Manipulation*, *Locomotion*, *Navigation*, *Planning*, *Sensing*, *Learning* und *Interaction*.

Manipulation beinhaltet die Erforschung von Methoden der aktiven Veränderung der Umwelt, beispielsweise durch Greifer oder künstliche Hände in unterschiedlichen Kontexten (Mason, 2001; Murray, Li, & Sastry, 1994; Shimoga, 1996). Zu aktuellen Herausforderungen gehören etwa die Koordination multipler Manipulatoren bei der Bewegung eines einzelnen Objekts (Mattar, 2013; Okamura, Smaby, & Cutkosky, 2000) oder der Umgang mit unstrukturierten und dynamischen anstelle von statischen und kontrollierten Umgebungen (Katz, Kenney, & Brock, 2008; Kemp, Edsinger, & Torres-Jara, 2007).

Im Forschungschwerpunkt Locomotion, werden Fortbewegungsformen von Robotern studiert und entwickelt, beispielsweise pedale, radgestützte, serpentine oder schwimmende (Apostolopoulos, 2001; Bhatti, Plummer, Iravani, & Ding, 2015; Transeth, Pettersen, & Liljebäck, 2009; Zhang, Hu, Shen, & Xie, 2006). Ziel ist es unter Anderem effizientere Bewegungsabläufe von Robotern zu ermöglichen, sei es z.B. durch die Reduzierung des Kraftaufwands bei der Lokomotion durch die Ausnutzung physikalischer Eigenschaften einer bestimmten Konstruktion (McGeer, 1990) oder durch die Identifikation von leichter berechenbaren Bewegungsmodellen (Holmes, Full, Koditschek, & Guckenheimer, 2006). Wie auch im Forschungsschwerpunkt Manipulation spielen dabei Herangehensweisen aus der Kinematik, Dynamik und Kontrolltheorie eine wesentliche Rolle (Featherstone & Orin, 2000; Lenarcic & Husty, 2012; Spong & Fujita, 2011).

Der Schwerpunkt Navigation umfasst neben der Bewegungsplanung für eine anschließende Raumdurchquerung, das heißt die Separierung eines größeren Handlungsstrangs in notwendige Teilbewegungen (Hwang & Ahuja, 1992; Masehian & Sedighizadeh, 2007), auch die Kartierung der Umwelt, das sogenannte Mapping (Aulinas, Petillot, Salvi, & Lladó, 2008; Thrun, 2002). Letzteres

gilt insbesondere dann als problematisch, wenn die Position des Roboters unbekannt ist. Die zugehörige Problemstellung ist auch geläufig unter dem Namen *Simultaneous Location and Mapping* und gilt als eines der wichtigsten Probleme auf dem Weg zum komplett autonomen Roboter (Thrun & Leonard, 2008).

Als Planning ist der Forschungsschwerpunkt benannt, welcher sich mit Planungsalgorithmen und Prozessen beschäftigt, welche Handlungsabfolgen von Robotern zur Erreichung eines bestimmten Ziels organisieren, von Fehlern bereinigen und optimieren sollen (Dean & Kambhampati, 1996; Ghallab, Nau, & Traverso, 2016; McDermott, 1992). Ob es sich dabei tatsächlich um ein distinktes Aufgabengebiet handelt und sich nicht beispielsweise mit Learning überschneidet, gilt jedoch als kontrovers (McDermott, 1992).

Sensing umfasst die Wahrnehmung der Umwelt mittels Sensoren, beispielsweise visuelle, taktile oder akustische, und die Verarbeitung der zugehörigen Signale über entsprechender Algorithmen (Chen, Li, & Kwok, 2011; Desouza & Kak, 2002; Leonard & Durrant-Whyte, 1992; Nicholls & Lee, 1989). Anhaltende Bestrebungen stellen beispielsweise die Verbesserung der Reliabilität und Sensitivität von Sensoren (Yousef, Boukallel, & Althoefer, 2011), die Entwicklung effizienter Sensor-Array Systeme (Dahiya, Metta, Valle, & Sandini, 2010) und die Integration multipler Sensordaten in ein Gesamtbild dar (Khaleghi, Khamis, Karray, & Razavi, 2013).

Unter Learning wird im Allgemeinen das autonome Erlernen von notwendigen Fertigkeiten zur Bewältigung einer Aufgabe durch den Roboter verstanden (Argall, Chernova, Veloso, & Browning, 2009; Connell & Mahadevan, 1993; Jabin, 2010; Kaelbling, Littman, & Moore, 1996). Die gesuchten Verfahren sollen dabei die Handlungsfähigkeit von Robotern in unbekannten dynamischen Umgebungen und lebenslanges Lernen ermöglichen (Connell & Mahadevan, 1993; Thrun & Mitchell, 1995).

Im Forschungsschwerpunkt Interaction werden die Faktoren erforscht, welche die Qualität der Interaktion zwischen Roboter und Mensch beeinflussen (Breazeal, 2004; Goodrich & Schultz, 2008). Hierzu gehören beispielsweise neben optischen Merkmalen des robotischen Erscheinungsbildes und dessen situationsspezifischen Verhaltens auch die Einstellung der sozial interagierenden Person gegenüber Technologie (Fink, 2012; Nitsch & Glassen, 2015).

1.2 Robotik und Kognitionswissenschaften

Das in der naturgemäß interdisziplinären Robotik umgesetzte Wissen stammt nicht nur aus den drei Kerndisziplinen Computer-Science, Mechanical-Engineering und Electrical-Engineering, sondern

auch zunehmend aus den Bio-, Sozial-, und Kognitionswissenschaften (Birk, 2011; McKee, 2006). Letztere liefern wichtige Erkenntnisse, welche nicht nur in die ältere Forschungsrichtung Biorobotics bzw. Bio-inspired-Robotics (Beer, Quinn, Chiel, & Ritzmann, 1997; Pfeifer, Lungarella, & Iida, 2007, 2012), sondern auch in die seit Anfang des Jahrhunderts neu entstandenen Forschungsfelder Developmental- und Cognitive-Robotics einfließen (Asada, MacDorman, Ishiguro, & Kuniyoshi, 2001; Christaller, 1999; Lungarella, Metta, Pfeifer, & Sandini, 2003). Als Auslöser für diese Entwicklung lässt sich neben dem generellen Trend, Verhaltensmuster von Robotern erlernen zu lassen statt diese selbst zu programmieren (Bakker & Kuniyoshi, 1996; Connell & Mahadevan, 1993; Dorigo & Colombetti, 1994; Schaal, 1999), auch ein Paradigmenwechsel in den Kognitionswissenschaften vermuten (A. Clark & Grush, 1999). So äußerten sich zunehmend mehr Forscher kritisch über eine eingeschränkte Sichtweise auf das klassische Konstrukt Kognition, welche bisher als ein von Perzeption und Handlung unabhängiger Prozess in einem modularen System aufgefasst wurde (M. Anderson, 2003; Beer, 2000; Andy Clark, 1999). Als Gegenbewegung entwickelte sich daher eine als *Embodied-Cognition* oder *Grounded-Cognition* bezeichnete Forschungsrichtung, welche Kognition, Perzeption und Handlung und mittelbar den eigenen Körper und die nähere Umgebung in einem holistischen Rahmen betrachtet (Barsalou, 2007; Glenberg, Witt, & Metcalfe, 2013; L. Smith & Gasser, 2005). Konsequenterweise wurden dabei Roboter als ideale Plattform vorgeschlagen, um kognitionswissenschaftliche Theorien in der Praxis zu überprüfen (D'Mello & Franklin, 2011; Morse, Herrera, Clowes, Montebelli, & Ziemke, 2011).

1.3 Klassische Informationsverarbeitungstheorien aus Sicht der Embodied-Cognitive-Science

Heutzutage wird kaum mehr bezweifelt, dass Embodiment einen wesentlichen Einfluss auf Kognition hat. Der Disput hat sich vielmehr auf die Frage verlagert, ob ausnahmslos jeder Bereich der Kognition einen handlungsbezogenen bzw. perzeptuellen Ursprung hat. Mahon und Caramazza (2008) halten beispielsweise die bestehenden empirischen Belege für eine gegroundete Konzeptverarbeitung für ebenso plausibel für die klassische Disembodied-Cognition. Auch Dove (2010) führt für die menschliche Sprache theoretische und empirische Befunde auf, welche für eine Form der Disembodied-Cognition sprechen. Chatterjee (2010) sieht zusätzlich in der Fähigkeit zur Abstraktion eine wesentliche Herausforderung für die Grounding-Theorie. Wie ein Jahrzehnt zuvor sind deshalb noch heute Forscher der Überzeugung, dass die Fähigkeit zur von Perzeption und

Handlung unabhängigen Kognition eine spezies-typische Besonderheit darstellt (Mahon, 2015; M. Wilson, 2002).

Trotz anhaltender Kontroversen werden klassische, unter dem Paradigma der amodalen symbolischen Informationsverarbeitung postulierte Theorien der Kognitionswissenschaften auch von Vertretern der Embodied-Cognition als weiterhin wertvoll erachtet. So glaubt beispielsweise Barsalou (2007, 2010), dass auf höherer theoretischer Ebene, die ursprünglichen Theorien im Wesentlichen erhalten bleiben und Wilson und Clark (2009), dass sich der klassische Computationalismus mühelos zu einem gegroundeden Computationalismus erweitern lässt. Hierzu ergänzend sind Pezzulo, Barsalou, Cangelosi, Fischer, McRae und Spivey (2013) der Überzeugung, dass viele bereits existierende Typen von computationalen Modellen wie konnektionistische, dynamisch-systemische und bayes'sche, gut geeignet sind, um gegroundete Phänomene zu simulieren.

1.4 Robotik und Bayesian-Cognitive-Science

Eine in den letzten Jahren in den Kognitionswissenschaften entstandene und mittlerweile einflussreiche Forschungsrichtung, welche sich ausschließlich bayes‘scher Modellierung bedient, ist die sogenannte Bayesian-Cognitive-Science (Chater & Oaksford, 2008; Oaksford & Chater, 1998). Postulat dieser Disziplin ist, dass approximierte Bayes‘sche Inferenz der zentrale zugrundeliegende Mechanismus aller im Gehirn stattfindenden kognitiven Prozesse ist (Chater, Tenenbaum, & Yuille, 2006; Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010; Tenenbaum, Kemp, Griffiths, & Goodman, 2011). Basis jeglicher Modellentwicklung stellt dabei die *Rational-Analysis* nach Anderson dar (J. Anderson, 1990; J. Anderson & Milson, 1989). Hierbei handelt es sich um eine sechstufige Methode zur Theorieentwicklung, nach welcher ein Modell zur potentiellen Funktionsweise eines kognitiven Prozesses über angenommene Ziele des Prozesses, Eigenschaften der Umwelt und computationalen Limitierungen abzuleiten ist (Chater & Oaksford, 1999). Die zugrundeliegende Annahme ist, dass biologische Systeme durch fortwährende Evolutionsprozesse an ihre Umwelt adaptiert sind und dadurch optimale Lösungen für bestehende Probleme entwickelt haben. Die Rational-Analysis beschäftigt sich daher mit der Frage, wie ein gegebenes Problem unter umwelt- und systembedingten Voraussetzungen in optimaler Weise gelöst werden kann. Dabei spielt Bayes‘sche Statistik als mathematisch optimale Herangehensweise an unsicherheitsbedingte Entscheidungen eine zentrale Rolle.

Nach bestem Wissen des Autors wurden, bis auf eine Ausnahme, keine Ansätze aus der Bayesian-Cognitive-Science in Cognitive-Developmental-Robotics verfolgt. Dies mag sicherlich darin begründet liegen, dass bestehende Theorien der Bayes‘schen Kognition häufig nur in konzeptioneller Form vorliegen und noch weit von einer praktischen Umsetzbarkeit entfernt sind (siehe auch Glassen & Nitsch, 2016; Holyoak & Cheng, 2010). Die angesprochene Ausnahme stellt hierbei der *Multimodale Hierarchische Dirichlet Prozess* (MHDP) von Nakamura, Nagai und Iwahashi (2011) dar. Dabei handelt es sich um ein Clustermodell, welches multimodale Messdaten von Objekten einer Trainingssitzung, beispielsweise auditive, visuelle und haptische Informationen, in Klassen einteilt und es dadurch ermöglicht fehlende Modalitätsinformationen eines neuen Objektes in einer anschließenden Erkennungsphase zu inferieren. Der MHDP basiert auf dem Hierarchischen *Dirichlet Prozess* (HDP) von Teh, Jordan, Beal und Blei (2006), ein Prior für statistische Modelle, sogenannte Topic Models, welche gemeinsame inhaltliche Themen einer definierten Anzahl an Dokumenten anhand der darin vorkommenden Wörter identifizieren sollen (Blei, 2012). Griffiths et al. (2007) erkannten bereits 2007, dass sich über diesen Prior ein Modell formulieren lässt, welches eine Vereinheitlichung aller existierenden rationalen Kategorisierungsmodelle innerhalb der Bayesian-Cognitive-Science darstellt. Darüber hinaus hat es Eigenschaften, welche es in beschränktem Maße erlauben, Vorwissen in die Kategorisierung mit einzubeziehen. Letztere wurden in einem Artikel von Canini, Shashkov und Griffiths (2010) aufgegriffen, um erfolgreiches Transferlernen mit dem HDP bei verschiedenen Kategorisierungsaufgaben zu demonstrieren. Dieser Befund wurde schließlich von Nakamura et al. (2011) als Evidenz für ein gutes Modell der menschlichen Kategorisierung und damit als empirische Untermauerung ihres Kategorisierungsansatzes bei Robotern angesehen. Tatsächlich handelt es sich jedoch beim Modell von Griffiths et al. (2007), im Weiteren als *Unifying Rational Model* (URM) bezeichnet, um ein Modell, welches bisher kaum evaluiert wurde. So finden sich in der Literatur bisher nur Modellierungsergebnisse zu den Kategorisierungsexperimenten von Smith und Minda (1998) in Griffiths et al. (2007) und Nosofsky, Gluck, Palmeri, McKinley und Glauthier (1994) in Glassen und Nitsch (2015). Ob nun dieses Modell gerechtfertigt als gutes Modell menschlicher Kategorisierung bezeichnet werden kann, ist die zentrale Frage dieser Arbeit. Bevor jedoch näher auf die Motivation dieser Frage eingegangen wird, soll zunächst das Forschungsfeld der Kategorisierung vorgestellt werden, zu welchem diese Arbeit einen ebenso wesentlichen Beitrag darstellen soll.

1.5 Das Forschungsfeld der Kategorisierung: Ein Überblick

Die Kategorisierung als Forschungsrichtung hat eine lange Tradition innerhalb der interdisziplinären Kognitionswissenschaften (z.B. Hull, 1920), mehrfach Paradigmenwechsel durchlaufen und neben einer unzähligen Anzahl an Befunden zur menschlichen Kategorisierung eine Reihe an einflussreichen formalen Modellen hervorgebracht. Während man in den Anfängen des Fachgebiets Kategorien noch als Klassen von Objekten verstand, welche sich anhand von für die Kategorie hinreichenden und notwendigen Merkmalen als gleichartige Mitglieder einer Kategorie auszeichnen, musste man diese als klassische Sicht bezeichnete Theorie bald verwerfen (Komatsu, 1992; Medin, 1989). So stellte man fest, dass es nicht nur Kategorien gibt, welche dieser Definition nach übereinstimmender Meinung gerecht werden (Sternberg & Sternberg, 2012), sondern auch Kategorien, bei welchen sich Menschen nicht immer einig sind, welche Objekte dazugehören bzw. wo die Grenzen der Kategorie verlaufen (Waldmann, 2017). Zudem zeigte sich bei Versuchspersonen, dass oft keine übereinstimmende Ansicht gegeben ist, welche Merkmale von Objekten hinreichend bzw. notwendig für eine Kategorie sind und dass manche Mitglieder einer Kategorie als typischer für diese Kategorie wahrgenommen werden als andere (Waldmann, 2017).

Hieraus entwickelte sich die Theorie der Prototypen, welche besagt, dass keine notwendigen oder hinreichenden Merkmale die Zugehörigkeit zu einer Kategorie definieren, sondern die Ähnlichkeit eines Objekts zum Prototyp einer Kategorie (Medin, 1989). Prototypen stellen dabei eine Art mittleres Objekt der Kategorie bzw. die mittlere Tendenz der Merkmale aller Mitglieder der Kategorie dar und verkörpern die einzigen Informationen welche über eine Kategorie gespeichert werden (Waldmann, 2017). Die Prototypensicht wird deshalb auch als eine Abstraktionstheorie bezeichnet (J. Anderson, 2015).

In der Prototypentheorie geht man davon aus, dass Typikalität, also die Frage danach, wie typisch ein Mitglied für seine Kategorie ist, von der Ähnlichkeit des Objekts zum Prototyp abhängt. Je unähnlicher sich beide sind, desto untypischer gilt das Objekt für die Kategorie (Waldmann, 2017). Gleichzeitig lässt sich über diese Theorie erklären warum wir Objekte als Mitglieder derselben Kategorie erkennen können, welche keine ihrer Merkmale teilen. Wittgenstein nennt hierfür das Spiel als Beispiel (Welding, 2013). So gibt es viele unterschiedliche Arten von Spielen, etwa Spiele die Spaß machen, Spiele welche man alleine oder mit anderen spielen kann, Spiele bei denen man würfeln muss oder nur Karten hat, Spiele wo es um Geld geht oder Spiele bei denen niemand gewinnen kann (Sternberg & Sternberg, 2012). Obwohl nun alle Mitglieder der Kategorie Spiel kein Merkmal teilen, wissen wir dennoch wann etwas als Spiel zu bewerten ist. Das hierbei angewandte

oft nicht bewusste Bewertungsmaß, ist die sogenannte Familienähnlichkeit. Sie liegt dann vor, wenn sich Objekte viele Merkmale teilen, dennoch aber keines gemeinsam haben (Medin, Wattenmaker, & Hampson, 1987). Zunächst ging man davon aus, dass natürliche Kategorien, also keine künstlich im Labor vorgegebenen, generell eine Familienähnlichkeitsstruktur aufweisen (Lassaline & Murphy, 1996). Tatsächlich aber hat sich gezeigt, dass es domänenabhängige Unterschiede gibt. So zeigt sich zwar in sozialen Kategorien, beispielsweise Personenkategorien wie Wissenschaftler, eine Familienähnlichkeitsstruktur, während hingegen bei Artefakten, also Kategorien welche von Menschen hergestellte Objekte umfassen, häufiger nicht linear separable Kategorien vorzufinden sind (Waldmann, 2017). Mit letztgenannter Art von Kategorien, sind solche Kategorien gemeint, für deren Mitglieder keine $w_1,\ldots,w_n$ Gewichte und Threshold θ existieren, mit welchen für jedes Mitglied i einer Kategorie A mit den Merkmalen $[a_{i1}, \ldots,a_{in}]$ und jedes Mitglied j der Kategorie B mit den Merkmalen $[b_{j1}, \ldots,b_{jn}]$ folgende Ungleichung gilt (Bobrowski & Łukaszuk, 2009):

$$\sum_{k=1}^{k=n} w_k \cdot a_{ik} < \theta < \sum_{k=1}^{k=n} w_k \cdot b_{jk} \quad (1)$$

Anders ausgedrückt sind zwei Kategorien dann linear separabel, wenn ihre in einem n-dimensionalen Merkmalsraum positionierten Mitglieder durch eine Hyperebene getrennt werden können (Blair & Homa, 2001).

Die Tatsache, dass auch nicht linear separable Kategorien existieren und die Prototypentheorie diese Art der Kategorien nicht erklären kann, hat unter anderem dazu geführt, dass die Prototypentheorie zunehmend angezweifelt und einer neuen aufkommenden Theorie, der sogenannten Exemplarsicht, gegenübergestellt wurde (Ashby & Maddox, 2004; Murphy & Medin, 1985; J. Smith & Minda, 1998).

Diese alternative Theorie distanziert sich von der Idee der Abstraktion und postuliert stattdessen eine getrennte Speicherung von Kategorienmitgliedern (Medin, 1989; Ross & Makin, 1999). Hierdurch lassen sich nicht nur die Existenz von nicht linear separablen Kategorienstrukturen nachvollziehen, sondern auch andere, für die Prototypentheorie kritische Befunde erklären, beispielsweise warum Versuchspersonen neben der zentralen Tendenz von Kategorien, auch die Variabilität ihrer Mitglieder sowie die relative Größe von Kategorien berücksichtigen (Waldmann, 2017).

Wenn auch das bisher einflussreichste und gegenüber Kontrahenten der Prototypensicht am meisten bewährte formale Kategorisierungsmodell, das *Generalized Context Model* (Nosofsky, 1986), ein Exemplarmodell darstellt (Vanpaemel & Lee, 2012; Wills & Pothos, 2012), bestanden dennoch wichtige Kritikpunkte an der Theorie. So ist zum einen die Speicherungshypothese generell unrealistisch, da wir schlicht und einfach nicht die Kapazität zur Verfügung haben um jedes jemals

beobachtete Objekt individuell zu speichern (Collier, 2005) und zum anderen kann die Exemplartheorie nicht erklären auf welcher Basis Objekte zu Kategorien zusammengefasst werden (Waldmann, 2017). Letztere Kritik trifft beispielsweise auf die Prototypen- bzw. die klassische Sicht nicht zu, da bei diesen Familienähnlichkeit bzw. definierende Merkmale als Basis der Kategorisierung gelten und somit etwa eine natürliche Kategorie mit Bäumen und Kaffeetassen als Mitglieder, anders als in der Exemplarsicht, unplausibel wäre.

Nicht nur dieser Punkt, sondern auch die zuvor erwähnten Tatsachen, dass bestimmte Domänen Kategorien mit Prototypenstruktur aufweisen und manche Kategorien auch über definierende Merkmale ihrer Mitglieder bestimmt sind, boten den Nährboden für einem Multisystemansatz. Tatsächlich wurde hierfür in neuerer Zeit neurowissenschaftliche Evidenz präsentiert (Ashby, Alfonso-Reese, Turken, & Waldron, 1998), wenn auch weiterhin kontrovers diskutiert wird, ob viele der propagierten Belege für Multisysteme nicht auch durch Einzelsysteme erklärt werden können (Nosofsky & Johansen, 2000; Zaki, Nosofsky, Jessup, & Unverzagt, 2003).

Unabhängig von dieser neueren Entwicklung im Forschungsfeld der Kategorisierung zeigten die in der Vergangenheit postulierten Theorien zur menschlichen Kategorisierung, die klassische, die Prototypen- und die Exemplarsicht, eine gemeinsame Schwäche: Sie basieren auf dem Konzept der Ähnlichkeit, welche bisher als kontextinvariant und symmetrisch gesehen wurde und anhand eines stabilen Merkmalsraums bestimmt sein soll (Waldmann, 2017). Dieses Verständnis bildet die Realität jedoch nicht adäquat ab. So konnte beispielsweise gezeigt werden, dass Grau im Kontext der Bewertung von Haar, als weiß eingestuft wird aber bei Wolken als Schwarz, ein Granatapfel einem Apfel ähnlicher ist als ein Apfel einem Granatapfel und Versuchspersonen im Verlauf eines Lernprozesses beginnen, zuvor ungenutzte Merkmale zur Klassifikation heranziehen (Waldmann, 2017).

Solche Befunde hatten zur Folge, dass in zunehmendem Maße anstelle einer ähnlichkeitsbasierten eine theoriebasierte Kategorisierung diskutiert wurde (Medin, 1989). Dennoch blieben ähnlichkeitsbasierte Ansätze weiterhin attraktiv, da mit Einzug theorienbasierter Theorien nun auch in der Kategorisierungsforschung die Debatte zwischen den Nativisten, welche bereichspezifisches angeborenes Wissen vermuten, und den Konstruktivisten, die jegliche Art von Wissen als erlernt ansehen, vermehrt Relevanz hatte (Gentner & Medina, 1998; Silvén, 2002; Vosniadou, 1994). Dabei wird in Betracht gezogen, dass zumindest zu Beginn eines Lernprozesses, ähnlichkeitsbasierte Mechanismen eine Rolle spielen und im weiteren Verlauf durch theorienbasierte Mechanismen abgelöst werden (Goldstone, 1994). Welche genauen ähnlichkeitsbasierten Mechanismen hierbei eine Rolle spielen, ist dabei weiterhin Gegenstand aktueller Forschung, wobei sich mittlerweile unzählige Vorschläge in Form von formalen Modellen der ähnlichkeitsbasierten Kategorisierung

wiederfinden. Eine Fülle an Vorschlägen reicht jedoch nicht aus um den gesuchten Mechanismen auf die Spur zu kommen, bestehende Modelle müssen auch ausgiebig evaluiert werden. So bemängeln Wills und Pothos (2012), dass genau dieses Defizit den derzeitigen Fortschritt in der Kategorisierungsforschung behindert.

1.6 Forschungsvorhaben

In dieser Arbeit soll dieses Defizit aufgegriffen, eine umfassendere Evaluation des URM durchgeführt, sowie Verbesserungsmöglichkeiten des Modells vorgestellt werden. Die generelle Motivation des Vorgehens besteht dabei nicht nur in der Erarbeitung von Grundlagen zur Plausibilisierung des Modells als psychologisch fundiertes Modell im Robotikkontext, sondern auch in der allgemeinen Entwicklungsunterstützung von zukünftigen psychologisch orientierten Klassifikationsverfahren auf Basis des HDP in der Robotik, sowie in der Unterstützung der Kategorisierungsforschung durch systematische Analyse eines neueren bisher kaum getesteten Verfahrens.
Den beiden erstgenannten Punkten wird dabei neben der Evaluation der bisher von Griffiths et al. (2007) vorgeschlagenen Modellvariante an ausgewählten Kategorisierungsaufgaben und der Analyse und theoretischen Überprüfung von implizit getroffenen Annahmen im Modell bezüglich der Natur der menschlichen Kategorisierung auch durch die in dieser Arbeit abschließende Beurteilung des Verfahrens hinsichtlich der Anwendbarkeit in dynamischen Kontexten auf Basis der akkumulierten (technischen) Erkenntnisse entsprochen. Hierzu zählen neben der Benennung wichtiger Eckdaten des Verfahrens wie Trainings- und Inferenzaufwand und einer Auflistung bestehender Optimierungsmöglichkeiten zur Erlangung einer höheren Ausführungsgeschwindigkeit auch die Bewertung der Nützlichkeit und Zuverlässigkeit von präsentierten Modellverbesserungen, sowie eine Beurteilung des Aufwands von Verfahrensanpassungen an spezifische Kontexte.
Zusätzlich wird diese Arbeit auch als ein Beitrag zum kognitionswissenschaftlichen Forschungsfeld der Kategorisierung verstanden, in welchem ein genereller Mangel an umfassenden Evaluationen der bisher vorgeschlagenen und konkurrierenden Kategorisierungsmodellen vorliegt (Wills & Pothos, 2012). Dieser Umstand trägt dazu bei, dass die Identifizierung eines einheitlichen Modells, welches möglichst alle bekannten charakteristischen Eigenschaften menschlicher Kategorisierung vorhersagen kann, erschwert wird. Letzteres wiederum führt dazu, dass der wesentliche Kategorisierungsmechanismus beim Menschen weiterhin ungeklärt bleibt oder zumindest keine ernstzunehmenden Kandidaten vorliegen.

Ein Fortschritt in diesem Bereich kommt außerdem unmittelbar der psychologischen Robotikforschung zu Gute. Beispielsweise ist durch das *No Free Lunch Theorem* bekannt, dass es kein Klassifikationsverfahren geben kann, welches in der mittleren Generalisierungsleistung über alle denkbaren Klassifikationsprobleme einem alternativen Verfahren überlegen ist (Lattimore & Hutter, 2013; Schaffer, 1994). Hieraus ergibt sich zwangsläufig, dass ein gutes Klassifikationsverfahren für Roboter ein Verfahren sein muss, welches eine Optimierung für Klassifikationsprobleme aufweist, welche im Operationsumfeld des Roboters verstärkt auftreten. Da nun viele Roboter in einem Umfeld agieren, für welches menschliche Gehirne bereits ein offensichtlicherweise hoch angepasstes und optimiertes Klassifikationsverfahren aufweisen, liegt der Schluss nahe, dass eine Identifizierung des menschlichen Kategorisierungsmechanismus eine starke Verbesserung der bisher in Robotern verwendeten Verfahren ermöglicht.

Mit Bezug auf die Kategorisierungsforschung soll deshalb die vorliegenden Arbeit dazu beitragen die Akkuratheit der Prognosen des URM bei ausgewählten, bisher mit dem Verfahren nicht modellierten Aufgabenstellungen von Kategorisierungsexperimenten zu bestimmen, welche zu charakteristischen Befunden der menschlichen Kategorisierungsleistung führten. Hierbei ist es zudem möglich erste Aussagen bezüglich der Prognosestärken und -schwächen des Verfahrens zu treffen und gegebenenfalls eine Eingrenzung des vorhersagbaren Leistungsspektrums vorzunehmen. Nicht zuletzt lässt sich durch die Evaluation des rationalen Kategorisierungsmodells URM bewerten, inwiefern die durchaus starke Annahme, dass der menschliche Kategorisierungsmechanismus eine Clusterstruktur nach dem Rich-Gets-Richer-Prinzip bestimmt, gerechtfertigt ist. Details hierzu folgen im Methodikteil.

Das URM soll im Folgenden an sieben in der Literatur bekannten und schwer zu modellierenden Befunden zur menschlichen Kategorisierungsleistung (= Kategorisierungsphänomene) aus Love, Medin und Gureckis (2004) evaluiert und gegebenenfalls Modellverbesserungen vorgeschlagen und getestet werden. Jedes dieser Experimente stellt dabei eine besondere Herausforderung an das zu testende Kategorisierungsmodell dar, welche üblicherweise, wenn überhaupt, nur von vereinzelten Modellen in der Literatur gemeistert werden konnten. Die Testreihe startet in Experiment 1 mit dem führenden Benchmark für formale Modelle der Kategorisierung, den charakteristischen Lernkurven von Shepard, Hovland und Jenkins (1961), bei welchen man annimmt, dass nur Modelle mit einem selektiven Aufmerksamkeitsmechanismus in der Lage sind die Befunde vorherzusagen (Kurtz, Levering, Stanton, Romero, & Morris, 2013; Nosofsky et al., 1994). Anschließend folgt in Experiment 2 eine Evaluation des Modells an der bisher einflussreichsten Kategorienstruktur in der Literatur, den sogenannten 5-4 Kategorien, welche als besonders schwierig für Prototypenmodelle gelten (Blair & Homa, 2003; J. Smith & Minda, 2000). Wird diese Kategorienstruktur bei Stimuli

mit idiosynkratischen Informationen verwendet, zeigt sich zudem ein Lernvorteil in der Identifikations- gegenüber der Kategorisierungsbedingung, welcher sonst in Abwesenheit solcher Informationen in der Literatur als umgekehrt bzw. höchstens gleichartig berichtet wurde (Blair & Homa, 2003; Medin, Dewey, & Murphy, 1983; Reed, 1978; Shepard et al., 1961).

Experiment 3 und 4 schließen an mit Befunden von Yamauchi und Markman (1998) und Yamauchi, Love und Markman (2002), welche die lange Zeit bestehende Annahme widerlegen konnten, dass Kategorienrepräsentationen invariante Strukturen aufweisen, welche nur durch die in der Umwelt vorliegenden objektiven Merkmalszusammenhänge determiniert sind (Waldmann, 2017). Stattdessen hat nach Yamauchi und Kollegen ebenfalls die Kategoriennutzung, zum Beispiel zur Inferenz von Merkmalen oder nur zur Klassifikation von Stimuli, einen Einfluss auf die Repräsentation. Da es sich hierbei um neuere Erkenntnisse in der Kategorisierungsforschung handelt, dürften viele etablierten Modelle Schwierigkeiten haben diese Befunde vorherzusagen.

Experiment 5 und 6 sind zusätzlich problematisch für klassische Assoziationsmodelle, welche von einer Kompetition der Merkmale bei der Vorhersage der Kategorienzugehörigkeit ausgehen, sodass die Kantengewichte für alternative prädiktive Merkmale nicht angepasst werden (Waldmann, 2017). Die hier zu evaluierenden Experimente von Billman und Knutson (1996) ergaben nämlich, dass die Leistung bei VP nach unsupervidiertem Erlernen von Kategorien durch hoch interkorrelierende Merkmale zunahm und somit eine Unterstützung statt Kompetition zwischen alternativen Prädiktoren vorlag (Goldstone & Kersten, 2003).

Abschließend wird in Experiment 7 ein weiterer konträrer Befund zu einer gängigen Annahme in der Kategorisierungsforschung aufgegriffen, und zwar der Annahme dass natürliche Kategorien generell nach Familienähnlichkeit organisiert sind (Lassaline & Murphy, 1996). Tatsächlich gingen auch Medin, Wattenmaker und Hampson (1987) zunächst von dieser Annahme aus, stellten jedoch schnell fest, dass wenn VP zehn Tiere in zwei gleichgroße Gruppen einteilen sollten, diese anstatt eine Sortierung nach der tatsächlich vorliegenden Familienähnlichkeit eine Sortierung entlang nur einer Dimension bevorzugten. Viele Kategorisierungsmodelle, welche auf dieser klassischen Annahme beruhen, dürften somit Schwierigkeiten haben den Befund zu prognostizieren.

Zusätzlich zur Evaluation des URM wird das Verfahren mit einem bereits bekannten Modell, dem *Supervised and Unsupervised Stratified Adaptive Incremental Network* (SUSTAIN; Love & Medin, 1998) hinsichtlich seiner Modellierungsgüte und -flexibilität in den sieben angesprochenen Experimenten verglichen. Nach bestem Wissen des Autors, soll SUSTAIN bisher das einzige Modell sein, welches alle sieben Phänomene nachbilden könne (Love et al., 2004). Bis auf den Modellierungscode des ersten Experiments existieren jedoch nach Todd Gureckis alle übrigen Programmcodes nicht mehr (persönliche Kommunikation, 14. April 2016). Dieser verbleibende

Modellierungscode des ersten Experiments enthält jedoch hinsichtlich der Experimentalprozedur leichte Abweichungen vom Originalexperiment. Es wurde deshalb eine Remodellierung aller Experimente auch mit SUSTAIN angestrebt um zu prüfen, ob die Befunde repliziert werden können.

2 Methodik

Im Folgenden sollen die zwei Kategorisierungsmodelle, das URM und SUSTAIN, sowie die für die Evaluation verwendeten Berechnungs- und Auswertungsmethoden vorgestellt werden. Beginnend mit dem Satz von Bayes, als basalstes Prinzip der Modellberechnungen rationaler Modelle, wird zunächst schrittweise über Bernoulli-, Beta- und Dirichletverteilungen an den recht komplexen DP herangeführt, anschließend übergeleitet zum HDP, um schließlich die konkrete Funktionsweise des URM zu besprechen.

2.1 Der Satz von Bayes

Der Satz von Bayes (Bayes & Price, 1763) ist ein grundlegender Satz aus der Wahrscheinlichkeitstheorie. Er besagt, dass die Wahrscheinlichkeit eines Ereignisses *A* unter der Bedingung, dass Ereignis *B* eingetreten ist, genauso groß ist wie die Wahrscheinlichkeit des Ereignisses *B*, unter der Bedingung, dass Ereignis *A* bereits vorliegt, multipliziert mit dem Verhältnis der Wahrscheinlichkeiten von *A* und *B*:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \tag{2}$$

Diese Gleichung lässt sich gering umformen, indem man die Wahrscheinlichkeit *P*(*B*) durch das Äquivalent ersetzt, welches sich durch den Satz der totalen Wahrscheinlichkeit ergibt:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|\bar{A}) \cdot P(\bar{A})} \tag{3}$$

Wir stellen uns nun vor, wir hätten in (2) nicht das Ereignis *A* betrachtet, sondern sein Komplement $\bar{A}$, dann gilt analog zur Umformulierung in (3) die Gleichung:

$$P(\bar{A}|B) = \frac{P(B|\bar{A}) \cdot P(\bar{A})}{P(B|A) \cdot P(A) + P(B|\bar{A}) \cdot P(\bar{A})} \tag{4}$$

An diesem einfachen Beispiel von zwei disjunkten Teilmengen A und $\bar{A}$ einer Grundmenge von Elementarereignissen Ω lässt sich erkennen, dass sich der Satz von Bayes auf eine beliebige Anzahl an disjunkten Teilmengen der Grundmenge Ω erweitern lässt. Generell gilt also:

$$P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{\sum_j P(B|A_j) \cdot P(A_j)} \tag{5}$$

Wir können dabei die A_i als die möglichen Ziehungen einer diskreten Zufallsvariable A auffassen wobei die $P(A_i)$ die Wahrscheinlichkeiten eine solche Ziehung vor Eintreten des Ereignisses B beschreiben und $P(A_i \mid B)$ die Wahrscheinlichkeiten nach Vorliegen dieses Ereignisses. Da nun (5) für alle Ziehungen aus A gilt, lässt sich somit (2) auch als Beziehung zwischen der Verteilung einer diskreten Zufallsvariable A vor und nach dem Vorliegen des Ereignisses B interpretieren. Dieser Zusammenhang lässt sich ebenfalls auf kontinuierliche Zufallsvariablen übertragen (Gelman et al., 2014). Dabei gilt für Realisierungen a und b aus den kontinuierlichen Zufallsvariablen A und B

$$f(a|b) = \frac{g(b|a) \cdot f(a)}{\int g(b|a) \cdot f(a) da} \tag{6}$$

Den Faktor $f(a)$ nennt man in diesem Zusammenhang die Priorverteilung und $f(a \mid b)$ konsequenterweise die Posteriorverteilung. $g(b \mid a)$ wiederum heißt Likelihood und der Nenner der rechten Gleichungsseite von (6) die Evidenz.

Der Satz von Bayes ermöglicht es somit eine Wahrscheinlichkeits- oder auch Wahrscheinlichkeitsdichteverteilung nach Vorliegen eines Ereignisses zu aktualisieren und damit in eine neue Verteilung der Posterior zu überführen. Das aufgetretene Ereignis können beispielsweise Beobachtungsdaten sein und der Prior die Glaubensverteilung für einen spezifischen Modellparameter. Wenn wir dabei den aktualisierten Prior, die Posteriorverteilung, als neuen Prior für zukünftige Beobachtungen verwenden, können wir nach und nach eine realistischere Verteilung des Modellparameters anhand von Erfahrungen erhalten. Dieses Bayes‘sche Lernen ist die Basis aller rationalen Kategorisierungsverfahren. Wie an (6) zu sehen ist, spielen Verteilungen dabei eine wesentliche Rolle. Im Folgenden sollen deshalb die für das URM relevanten Verteilungen, beginnend mit der Bernoulliverteilung, näher erläutert werden.

2.2 Bernoulli-, Beta-, und Dirichletverteilungen

Bei der Bernoulliverteilung handelt es sich um die einfachste der im Verlauf dieser Arbeit benötigten Verteilungen. Sie beschreibt die Wahrscheinlichkeiten zweier möglicher Realisierungen

einer binären Zufallsvariable, beispielsweise Kopf und Zahl als Ausgang eines Münzwurfs. Ihre Wahrscheinlichkeitsfunktion ist formuliert als (Kruschke, 2010):

$$P(v;p) = p^{v} \cdot (1-p)^{1-v} \quad \textit{für } v = 0{,}1 \tag{7}$$

p ist der einzige Parameter dieser Funktion. Er definiert die Wahrscheinlichkeit des Ereignisses v=1. Da es sich um eine Wahrscheinlichkeitsverteilung handelt und die Summe der Wahrscheinlichkeiten aller Ausgänge dieser binären Zufallsvariable somit immer 1 ergeben müssen, ist die Wahrscheinlichkeit des Ereignisses v=0 konsequenterweise $1 - p$.

Nun können wir uns folgende Situation vorstellen: Wir haben eine Münze und möchten angeben für wie wahrscheinlich wir es halten, dass die Münze in einem bestimmten Bereich möglicher Wurftendenzen liegt. Es könnte sich beispielsweise um eine Trickmünze handeln, bei welcher eine Seite häufiger auftritt, oder aber eine faire Münze, welche gleichhäufig auf beide Seiten fällt. Haben wir nun beispielsweise mit dieser Münze zehn Würfe durchgeführt und dabei achtmal Kopf und zweimal Zahl erhalten, so würden wir glauben, dass eine Bernoulliverteilung mit einem p zwischen 0.6 und 0.8 die wahre Tendenz der Münze wahrscheinlicher enthält als ein p zwischen 0.4 und 0.6.

Eine Funktion, deren Flächeninhalt unter den jeweiligen Kurvensabschnitten (z.B. 0.4 – 0.6) genau dieser Wahrscheinlichkeit des zugehörigen Tendenzbereichs entspricht, ist folgende:

$$f(p;\alpha,\beta) = \frac{p^{\alpha} \cdot (1-p)^{\beta} \cdot \binom{\alpha+\beta}{\alpha}}{\int_0^1 u^{\alpha} \cdot (1-u)^{\beta} \cdot \binom{\alpha+\beta}{\alpha} du} = \frac{p^{\alpha} \cdot (1-p)^{\beta}}{\int_0^1 u^{\alpha} \cdot (1-u)^{\beta} du} \tag{8}$$

Wie man sehen kann, beschreibt die Funktion die multiplizierten Wahrscheinlichkeiten von α Ausgängen v=1 und β Ausgängen v=0 einer Bernoulliverteilung für jeden Parameter p normalisiert über eine von α und β abhängigen Konstante, sodass die Größe der Fläche unter der Kurve auf dem Intervall [0, 1] 1 ergibt. Dekrementiert man nun in dieser Formel die Anzahl der Ereignisbeobachtungen α und β um jeweils 1 erhalten wir die Betaverteilung (Gupta & Nadarajah, 2004):

$$f(p;\alpha,\beta) = \frac{p^{\alpha-1} \cdot (1-p)^{\beta-1}}{\int_0^1 u^{\alpha-1} \cdot (1-u)^{\beta-1} du} \tag{9}$$

Die Betaverteilung kann also zur Beschreibung der Wahrscheinlichkeitsdichte aller Häufigkeitsproportionen auf dem Intervall (0, 1) zweier Ausgänge einer binären Zufallsvariable herangezogen werden. Die Wahrscheinlichkeitsdichte bezeichnet dabei die Höhe der Verteilung sowie die Steigung der Wahrscheinlichkeit an einer Stelle, während die Wahrscheinlichkeit für einen Ereignisbereich der Größe der Fläche unter dem zugehörigen Verteilungsabschnitt entspricht.

Wir stellen uns nun vor, wir hätten keine binäre Zufallsvariable wie ein Münzwurf, sondern eine diskrete Zufallsvariable mit mehr als zwei möglichen Realisierungen, beispielsweise ein Würfelwurf. In diesem Fall eignet sich die Betaverteilung nicht mehr für eine Beschreibung der Dichte der Häufigkeitsproportionen, da sie nur zwei mögliche Ausgänge in Betracht zieht. Wir benötigen stattdessen eine multivariate Erweiterung der Betaverteilung, die sogenannte Dirichletverteilung. Ihre Dichtefunktion ist wie folgt (Frigyik, Kapila, & Gupta, 2010):

$$f(p_{1,\dots}, p_V; \alpha_{1,\dots}, \alpha_V) = \frac{\Gamma(\sum_{i=1}^{V} \alpha_i)}{\prod_{i=1}^{V} \Gamma(\alpha_i)} \cdot \prod_{i=1}^{V} p_i^{a_i - 1} \tag{10}$$

Γ bezeichnet dabei die Gammafunktion, eine erweiterte Fakultätsfunktion mit der Eigenschaft Γ(n) = (n – 1)!, welche im Gegensatz zur Fakultätsfunktion auch für reelle und komplexe Zahlen als Argumente definiert ist. Analog zu α und β der Betaverteilung lassen sich die Parameter α_i bei der Dirchletverteilung als Anzahl der Beobachtungen von V disjunkten Ereignissen interpretieren.

2.3 Der Dirichlet Prozess

Bisher wurden nur Verteilungen der parametrischen Bayes-Statistik vorgestellt. Sie charakterisieren sich dadurch, dass Ziehungen aus ihnen immer durch eine endliche Anzahl an Informationen beschrieben werden können. Sie unterscheiden sich deshalb von Verteilungen der non-parametrischen Bayes-Statistik, bei welchen eine endliche Informationsmenge für die Beschreibung einer Ziehung nicht mehr ausreicht (Teh, 2010). Zu Letzteren gehört der Dirichlet Prozess (DP), welcher folgendermaßen definiert ist (Müller & Quintana, 2004):

$$\begin{aligned} &\textit{Wenn } G \sim \mathrm{DP}(H, \alpha), \\ &\textit{dann gilt } (G(C_1), \dots, G(C_K)) \sim \mathrm{Dir}(\alpha H(C_1), \dots, \alpha H(C_K)) \end{aligned} \tag{11}$$

„~“ ist als „verteilt als“ und „Dir“ als Dirichletverteilung zu lesen. Die C_i stellen eine endliche Partitionierung eines messbaren Raumes S dar. Dabei entspricht eine Ziehung aus einem Dirichlet Prozess einer diskreten Verteilung über dem Raum S, bei welcher die Wahrscheinlichkeiten der Raumcluster C_k dirichletverteilt sind. Das Dirichlet Prozess verteilte G kann man sich am besten folgendermaßen vorstellen: Es sei ein Stab der Länge 1 gegeben. Wir brechen zunächst ein Stück des Stabes ab und weisen einer zufälligen Ziehung aus einer Basisverteilung H die Länge des abgebrochenen Stabstücks als neue Wahrscheinlichkeit zu. Dann wiederholen wir diesen Brechvorgang unendlich oft auf das jeweils übrig gebliebene Stabstück. α beeinflusst dabei wie

groß die abgebrochenen Stabstücke sind. Ein kleines α begünstigt große abgebrochene Stücke, ein großes α führt hingegen zu vielen kleinen Stabstücken. Formal ist diese sogenannte *Stick-Breaking Construction* über die Addition von gewichteten Punktmassefunktionen δ definiert, welche ungewichtet eine Wahrscheinlichkeit von 1 an einer festgelegten Stelle θ_k repräsentieren (Teh, 2010):

$$\begin{aligned} b_k &\sim Beta(1, \alpha) \\ \theta_k &\sim H \\ \pi_k &= b_k \cdot \prod_{i=1}^{k-1} (1 - b_i) \\ G &= \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k} \end{aligned} \tag{12}$$

Die hieraus resultierende Verteilung G, welche demnach als eine Diskretisierung der Verteilung H mit umverteilten Wahrscheinlichkeiten gesehen werden kann, entspricht dann einer Ziehung aus einem DP mit Basisverteilung H und Konzentrationsparameter α. Abbildung 1 demonstriert das Verfahren anschaulich.

In der non-parametrischen Bayes‘schen Statistik wird der DP üblicherweise für Clusterverfahren, sogenannte Mischmodelle, verwendet. Dabei dient der DP als Prior über mögliche Clusterparametersets, welcher, aktualisiert über bereits beobachtete Clusterparametersets, eine

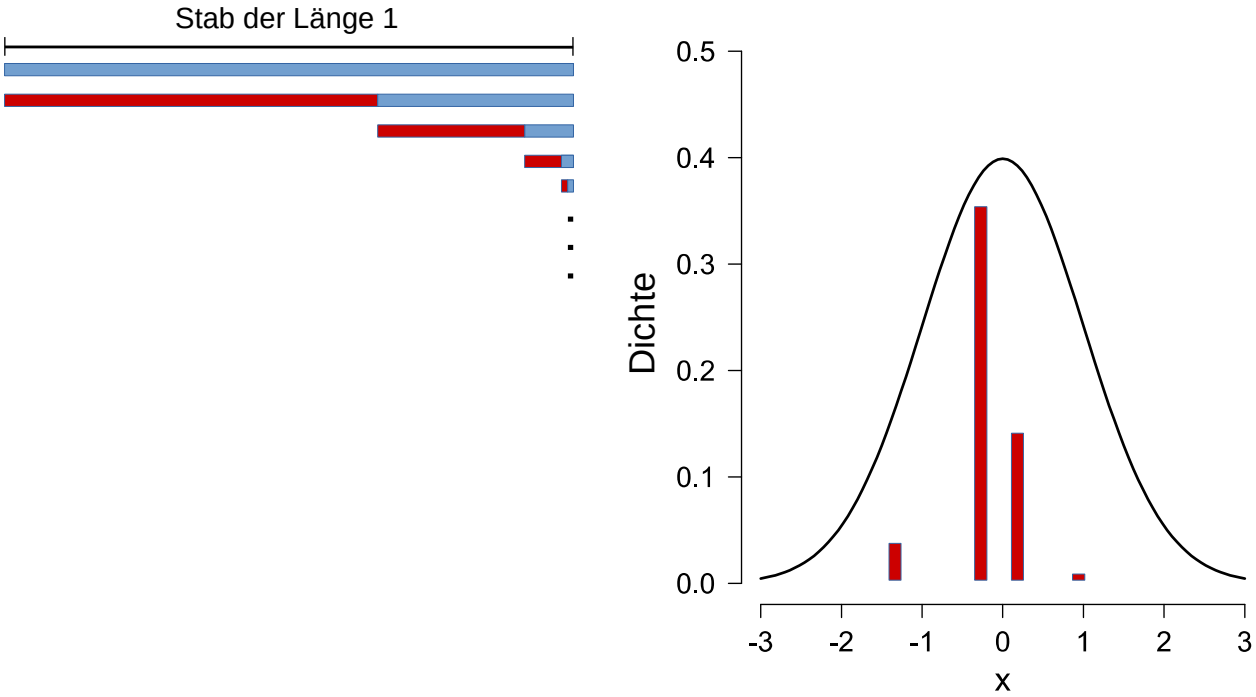


Abbildung 1. Schematische Darstellung der Stick-Breaking Construction mit der Standardnormalverteilung als Basisverteilung H und Konzentrationsparameter $\alpha = 1$.

Clusteringeigenschaft aufweist, sodass die Wahrscheinlichkeit zukünftig beobachteter Parametersets in der erwarteten Ziehung aus dem DP proportional zur Häufigkeit bereits beobachteter Clusterparametersets ist. Unter einem Clusterparameterset ist dabei ein Satz an Parametern zu verstehen, welcher die generierenden Verteilungen aus denen die Mitglieder des Clusters stammen, konkretisiert. Detaillierter betrachtet, lässt sich die Verwendbarkeit des DP als Prior eines Mischmodells folgendermaßen herleiten:

Wir stellen uns vor, die Basisverteilung H des DP ist eine Wahrscheinlichkeitsverteilung über dem Raum Θ, welcher alle möglichen Clusterparametersets beinhaltet. Eine Ziehung aus dem DP stellt dann eine durch den Konzentrationsparameter α kontrollierte Wahrscheinlichkeitsumverteilung dar, sodass niedrigere α höhere Konzentrationen der Wahrscheinlichkeit auf wenige Parametersets begünstigen. Der Erwartungswert des DP entspricht dabei punktweise der Basisverteilung (Teh, 2010), wobei α die Stärke der Streuung um den Erwartungswert reguliert. Im Weiteren wird auf die strenge, jedoch korrektere Formulierung der nur punktweisen Übereinstimmung von Erwartungswert und Basisverteilung eines DP zugunsten eines unmittelbareren Verständnisses der Zusammenhänge verzichtet.

Wir wollen nun für N+1 Objekte $o_{1,\ldots,N+1}$ ein Clustering erhalten und nehmen dazu an, wir wüssten bereits für die ersten N Objekte $o_{1,\ldots,N}$ jeweils die zugehörigen Clusterparametersets $\theta_{1,\ldots,N}$. Ausgehend von diesem Wissen prognostizieren wir ein geeignetes Parameterset θ_{N+1} für das letzte Objekt o_{N+1}, indem wir aus der *Posterior-Predictive-Distribution* (PPD) des DP ein neues Parameterset ziehen (Teh, 2010):

$$\theta_{N+1} \mid \theta_1, \ldots, \theta_N \sim E(G \mid \theta_1, \ldots, \theta_N) \tag{13}$$

Um das zu ermöglichen, müssen wir jedoch zunächst wissen, was der Erwartungswert eines DP ist, wenn bereits N beobachtete Parametersets $\theta_{1,\ldots,N}$ vorliegen. Hierzu betrachten wir die Definition des DP unter dieser Situation:

$$(G(C_1),\ldots,G(C_K)) \mid \theta_1, \ldots, \theta_N \sim \mathrm{Dir}(\alpha H(C_1)+n_1,\ldots,\alpha H(C_K)+n_K) \tag{14}$$

Da die Parameter der Dirichletverteilung, wie oben eingeführt, als Anzahl der Ereignisbeobachtungen interpretiert werden können, ist sofort ersichtlich, dass nichts anderes zu tun ist, als die Parameter der Dirichletverteilung um die Anzahl der beobachteten Parametersets zu erhöhen, welche zum assoziierten Raumcluster gehören. Die jeweiligen Erhöhungen sind in (14) mit $n_{1,\ldots,K}$ angegeben. Bei Betrachtung der aktualisierten Dirichletverteilung und unter Rückbesinnung auf die Definition (11) wird dann ersichtlich, dass der zugehörige DP folgendermaßen aussehen muss (Teh, 2010):

$$G \mid \theta_1, \ldots, \theta_N \sim DP\left(\alpha+N, \frac{1}{\alpha+N}\left(\alpha H + \sum_{i=1}^{N} \delta_{\theta_i}\right)\right) \text{ mit } \sum_{i=1}^{N} \delta_{\theta_i}(C_k) = n_k \quad (15)$$

Da uns bekannt ist, dass der Erwartungswert eines DP dessen Basisverteilung darstellt, wissen wir nun, dass für den Erwartungswert eines aktualisierten DP gilt:

$$E(G \mid \theta_1, \ldots, \theta_N) = \frac{1}{\alpha+N}\left(\alpha H + \sum_{i=1}^{N} \delta_{\theta_i}\right) \quad (16)$$

Aufgrund von (13) ist uns deshalb ebenfalls die PPD bekannt, welche wir zum besseren Verständnis geringfügig umstellen können (Teh, 2010):

$$\theta_{N+1} \mid \theta_1, \ldots, \theta_N \sim \frac{1}{\alpha+N}\left(\alpha H + \sum_{k=1}^{K} n_k \cdot \delta_{\theta_k^*}\right) \quad (17)$$

Die θ_k^* stellen dabei die einzigartigen Clusterparametersets dar und n_k ihre Häufigkeit unter den Beobachtungen. Wie man an (17) erkennen kann, weist der DP die eingangs beschriebene Clustering Eigenschaft auf, dass das nächste Parameterset θ_{N+1} mit einer Wahrscheinlichkeit proportional zur Häufigkeit n_k das Parameterset θ_k^* ist. Ein völlig neues Clusterparameterset wird hingegen mit einer Wahrscheinlichkeit proportional zu α prognostiziert. Diese Art von Clustering folgt demnach einem Rich-Gets-Richer-Prinzip, wobei bereits große Cluster voraussichtlich schneller wachsen als kleinere. Ein solches Phänomen ist ebenfalls in komplexen Netzwerken, wie beispielsweise biologischen Zellen, sozialen Gesellschaften oder dem Internet vorzufinden (Barabási, 2009).

(17) hat noch eine weitere Eigenschaft, welche sich auf die Wahrscheinlichkeiten unterschiedlicher Reihenfolgen derselben Beobachtungen bezieht. Diese sind nämlich für die Wahrscheinlichkeit des finalen Clusterings gänzlich irrelevant, sodass die erste Beobachtung ebenso die letzte hätte sein können. Wenn so etwas für Beobachtungen zutrifft, nennt man sie *austauschbar*. Hierbei handelt es sich um eine wichtige Voraussetzung, um über ein spezielles Verfahren zur Posteriorapproximation, das sogenannte *Gibbs Sampling* (Geman & Geman, 1984) effizient Samples zu generieren.

Um nun N+1 Beobachtungen über ein *Dirichlet Prozess-Mischmodell* (DPMM) zu clustern, kombiniert man (17) mit einer Likelihoodfunktion L

$$p(\theta_{N+1} \mid o_{N+1}, \theta_1, \ldots, \theta_N) = L(o_{N+1} \mid \theta_{N+1}) \cdot p(\theta_{N+1} \mid \theta_1, \ldots, \theta_N) \quad (18)$$

wählt beliebige initiale Clusterparametersets für $\theta_{1,\ldots,N}$ und zieht für θ_{N+1} ein neues Parameterset über (18). Anschließend redefiniert man jedes θ_i als θ_{i+1} und θ_{N+2} als θ_1 und zieht danach für das neue θ_{N+1} über (18) ein neues Parameterset. Dieses Vorgehen, welches als Gibbs Sampling bekannt ist,

wiederholt man bis für jede Beobachtung o_i ein neues Clusterparameterset ermittelt wurde. Das Ergebnis ist dann ein Sample aus der Posteriorverteilung des DPMM.

Die eben beschriebene Methode unter Verwendung der Verteilung in (18) wird jedoch üblicherweise nicht angewandt, da es sich um ein sehr ineffizientes Verfahren handelt (Neal, 2000). Der Grund hierfür ist, dass jedes Parameterset einzeln angepasst werden muss und somit ein großer Cluster nur sehr langsam hinsichtlich eines für alle Mitglieder des Clusters geeigneteren Parametersets optimiert werden kann.

Es gibt jedoch eine alternative Methode, welche diese Mängel nicht aufweist (Neal, 2000). Hierzu benötigen wir einen *Chinese Restaurant Prozess* (CRP), welcher am besten über folgende Metapher verstanden werden kann (siehe auch Gershman & Blei, 2012): Wir stellen uns ein chinesisches Restaurant vor, welches unendlich Tische beherbergt. Zu Beginn ist das Restaurant leer und ein neu eintreffender Gast setzt sich an den ersten Tisch. Jeder weitere Gast, der das Restaurant betritt, wählt nun einen Tisch mit einer Wahrscheinlichkeit proportional zur Anzahl der Gäste an diesem Tisch und mit einer Wahrscheinlichkeit proportional zu α *an* einen neuen Tisch. Formal lassen sich die Wahrscheinlichkeiten der Tischwahl t_{N+1} für den neuesten Gast mit Indexnummer *N+1* wie folgt zusammenfassen:

$$P(t_{N+1} = k \mid t_{1:N}) = \begin{cases} \dfrac{n_k}{N + \alpha} & \text{für mit } n_k \text{ Gästen besetzten Tisch } k \\ \dfrac{\alpha}{N + \alpha} & \text{für unbesetzten Tisch } k \end{cases} \tag{19}$$

Wir modifizieren nun (19) indem wir die Wahrscheinlichkeit für einen unbesetzten Tisch mit der Prior-Predictive-Distribution der Basisverteilung H für ein neues zu klassifizierendes Objekt o_{N+1} und die Wahrscheinlichkeiten für belegte Tische k mit der jeweiligen PPD von H nach Aktualisierung von H mit den am Tisch k platzierten Objekte multiplizieren (Neal, 2000). Wir erhalten dann

$$P(t_{N+1} = k \mid o_{N+1}, t_{1:N}) = \begin{cases} F_k(o_{N+1}) \cdot \dfrac{n_k}{N + \alpha} & \text{für mit } n_k \text{ Gästen besetzten Tisch } k \\ F_{new}(o_{N+1}) \cdot \dfrac{\alpha}{N + \alpha} & \text{für unbesetzten Tisch } k \end{cases} \tag{20}$$

wobei die Likelihoods F_k bzw. F_{new} folgendermaßen definiert sind:

$$\begin{aligned} F_k(o_{N+1}) &= \int L(o_{N+1} \mid \theta) \cdot H(\theta \mid \{o_i \mid t_i = k\}) d\theta \\ F_{new}(o_{N+1}) &= \int L(o_{N+1} \mid \theta) \cdot H(\theta) d\theta \end{aligned} \tag{21}$$

Wie beim vorhergehenden ineffizienten Verfahren mit (18) verwenden wir nun Gibbs Sampling, diesmal jedoch auf (20) angewandt. Dabei sampeln wir nicht, wie ursprünglich, die

Clusterparametersets θ_i der Objekte o_i, sondern die Tischnummern t_i, welche die Indizes der zugeordneten Cluster darstellen.

2.4 Der Hierarchische Dirichlet Prozess

Für die Basisverteilung des DP wurde bisher stets eine parametrische Verteilung gewählt. Es ist jedoch auch möglich *H* selbst als eine DP verteilte Zufallsvariable zu definieren. Gleichzeitig lassen sich mehrere DPs formulieren, welche dieselbe DP-verteilte Zufallsvariable als Basisverteilung verwenden. Dieses Prinzip lässt sich beliebig fortführen, sodass eine Baumstruktur entsteht, in welcher die Knoten des Baumes DPs darstellen. Die zugehörige hierarchische Verteilung ist unter dem Namen Hierarchical Dirichlet Process bekannt, welcher von Teh et al. (2006) eingeführt wurde. Da in der gesamten Arbeit HDPs verwendet werden, welche eine 2-Ebenen umfassende Baumstruktur aufweisen, soll der HDP im Folgenden an einer Baumstruktur mit einem Eltern-DP mit Nummer 0 und einer beliebigen Anzahl *J* an Kind-DPs mit Nummern 1,…,*J* veranschaulicht werden:

$$\begin{aligned} G_0 &\sim DP(\gamma, H) \\ G_j &\sim DP(\alpha_j, G_0) \end{aligned} \tag{22}$$

Wollen wir einen HDP dieser Art als Prior für ein Hierarchisches *Dirichlet Prozess-Mischmodell* (HDPMM) verwenden, so gestaltet sich die Herangehensweise analog zur CRP Methode beim singulären DP, welche beim HDP *Chinese Restaurant Franchise*-Sampling (CRF) genannt wird. Die zugehörige Metapher und Prozedur ist folgendermaßen (Teh et al., 2006): Wir haben ein Franchisesystem bestehend aus *J* Restaurants, welche an jedem ihrer Tische eine Speise aus dem Franchisemenü servieren. Die Sitzplatzwahl eines in Restaurant *j* eintreffenden Gastes gestaltet sich dabei analog zum CRP, sodass ein Gast o_{jN+1} einen bereits besetzten Tisch t mit einer Wahrscheinlichkeit proportional zur Anzahl der an ihm sitzenden Gäste n_{jt} oder einen unbesetzten Tisch proportional zum Konzentrationsparameter α wählt.

$$P(t_{jN+1} = t \mid \boldsymbol{t}) = \begin{cases} \dfrac{n_{jt}}{N_j + \alpha} & \text{für mit } n_{jt} \text{ Gästen besetzten Tisch } t \\ \dfrac{\alpha}{N_j + \alpha} & \text{für unbesetzten Tisch } t \end{cases} \tag{23}$$

Wählt der neue Gast o_{jN+1} einen bereits besetzten Tisch, so erhält er die Speise k_{jt}, welche bereits am Tisch serviert wird. Setzt der Gast sich jedoch an einen unbesetzten Tisch, so kann er eine neue Speise aus dem Franchisemenü aussuchen. Dabei wählt der Gast o_{jN+1} eine Speise k mit einer

Wahrscheinlichkeit proportional zur Anzahl der Tische im gesamten Franchise m_k, an welchen die Speise k bereits gegessen wird oder eine bisher unservierte Speise proportional zu γ.

$$P(k_{jt} = k \mid \boldsymbol{k}) = \begin{cases} \frac{m_k}{M + \gamma} & \text{für Speise } k \text{ serviert an } m_k \text{ Tischen} \\ \frac{\gamma}{M + \gamma} & \text{für bisher unservierte Speise } k \end{cases} \tag{24}$$

Ähnlich wie beim Vorgehen in (20-21) können wir nun (23-24) heranziehen um Konditionalverteilungen des HDPMM zu formulieren und Gibbs Sampling darauf anzuwenden. Ein Sample aus dem HDPMM enthält dabei, im Gegensatz zu einem Sample aus dem DPMM, nicht nur eine Tischzuweisung für jede Beobachtung, sondern auch eine Speisezuordnung für jeden Tisch.
Analog zu (20) ergänzen wir zunächst in (23) jeweils die Likelihoods für einen bereits besetzten Tisch und einen neuen Tisch:

$$P(t_{jN+1} = t \mid o_{jN+1}, \boldsymbol{t}, \boldsymbol{k}) = \begin{cases} F_{k_{jt}}(o_{jN+1}) \cdot \frac{n_{jt}}{N_j + \alpha} & \text{für mit } n_{jt} \text{ Gästen besetzten Tisch } t \\ \sum_{k=1}^{K+1} P(k_{jt} = k \mid o_{jN+1}, \boldsymbol{k}) \cdot \frac{\alpha}{N_j + \alpha} & \text{für unbesetzten Tisch } t \end{cases} \tag{25}$$

Die Verteilung $P(k_{jt} = k \mid o_{jN+1}, \boldsymbol{k})$ entspricht dabei (24), ergänzt um die jeweilige Likelihood für eine bereits servierte und eine bisher unservierte Speise:

$$P(k_{jt} = k \mid o_{jN+1}, \boldsymbol{k}) = \begin{cases} F_k(o_{jN+1}) \cdot \frac{m_k}{M + \gamma} & \text{für Speise } k \text{ serviert an } m_k \text{ Tischen} \\ F_{new}(o_{jN+1}) \cdot \frac{\gamma}{M + \gamma} & \text{für bisher unservierte Speise } k \end{cases} \tag{26}$$

Dabei handelt es sich bei der Likelihood F_k um die selbe Likelihoodfunktion wie in (25). Sie ist zusammen mit F_{new} analog zu (21) formuliert, sodass F_k die PPD von H gegeben aller einer Speise k zugeordneten Beobachtungen darstellt und F_{new} die entsprechende Prior-Predictive-Distribution von H, wenn noch keine Beobachtungen vorliegen:

$$\begin{aligned} F_k(o_{jN+1}) &= \int L(o_{jN+1} \mid \theta) \cdot H(\theta \mid \{o_{ji} \mid k_{jt_{ji}} = k, j=1,..,J\}) d\theta \\ F_{new}(o_{jN+1}) &= \int L(o_{jN+1} \mid \theta) \cdot H(\theta) d\theta \end{aligned} \tag{27}$$

Da ein Sample aus dem HDPMM nicht nur Tischbelegungen sondern auch Speisezuordnungen enthält und wir diese ebenfalls im Gibbs Sampling Zyklus neu bestimmen, benötigen wir noch eine Konditionalverteilung für die k_{jt}. Wir stellen uns hierzu vor, dass der zu k_{jt} zugehörige Tisch t der letzte neu belegte Tisch war und die an ihm sitzenden Gäste (die Beobachtungen) die letzten

eintreffenden Gäste des Restaurants. Die Wahrscheinlichkeit, dass sich diese Gäste an einem eigenen gemeinsamen Tisch mit Speise k_{jt} zusammenfinden, war dann:

$$P(k_{jt} = k \mid \boldsymbol{o}_{jt}, \boldsymbol{k}) = \begin{cases} F_k(\boldsymbol{o}_{jt}) \cdot \frac{m_k}{M + \gamma} & \text{für Speise } k \text{ serviert an } m_k \text{ Tischen} \\ F_{new}(\boldsymbol{o}_{jt}) \cdot \frac{\gamma}{M + \gamma} & \text{für bisher unservierte Speise } k \end{cases} \quad (28)$$

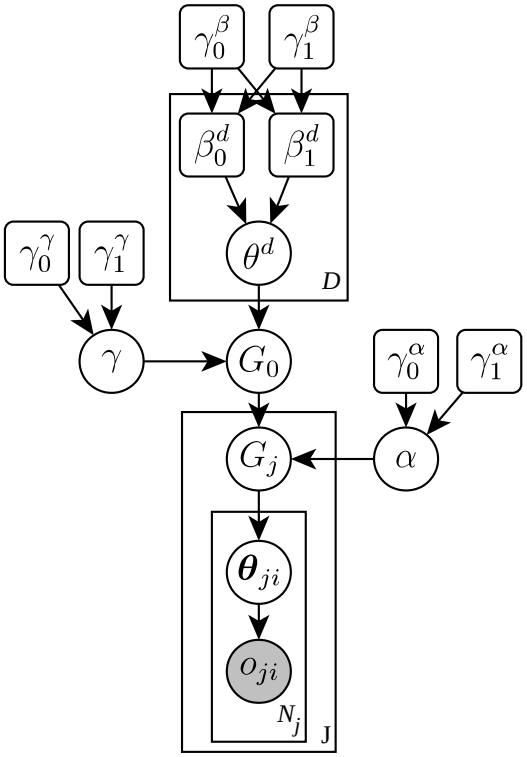
Es handelt sich somit im Wesentlichen um (26-27), wobei anstelle einer einzigen neuen Beobachtung die Likelihood aller Beobachtungen am Tisch *t* für Speise *k* berücksichtigt wird.

Um nun Samples aus dem HDMM zu generieren, gehen wir restaurantweise folgendermaßen vor: Wir bestimmen für jedes o_{ji} über (25) eine neue Tischbelegung. Wenn ein existierender Tisch gewählt wird, setzen wir $t_{ji} = t$. Soll hingegen ein neuer Tisch generiert werden, suchen wir einen bisher nicht verwendeten Index *t* und belegen t_{ji} entsprechend. Anschließend wählen wir über (26) eine Speise *k* für den neuen Tisch *t*. Wurde für jedes o_{ji} eine neue Tischbelegung ermittelt, suchen wir für jeden Tisch ein neues *k*, welches über (28) bestimmt wird. Sind die neuen Speisezuordnungen ebenfalls abgeschlossen und wurde dieses Verfahren für jedes Restaurant angewendet, liegt uns mit dem finalen (***t***, ***k***) ein Sample aus dem HDPMM vor.

2.5 Das URM

Beim URM handelt es sich um ein soeben besprochenes HDPMM. Abbildung 2 stellt das sogenannte generative Modell des URM dar. Es beschreibt, wie aus Sichtweise des Modells Beobachtungsdaten (= grau hinterlegter Kreis) zustande kommen. Kreise im Diagramm repräsentieren Zufallsvariablen, abgerundete Rechtecke stellen definierbare Parameter dar. Sind Variablen bzw. Parameter durch einen Kasten umrahmt, so bedeutet dies, dass die eingerahmten Knoten mehrfach vorkommen. Die Anzahl der Wiederholungen des Teilbereichs ist dann in der unteren rechten Ecke des Kastens angegeben. Pfeile im Diagramm stellen dar, welche Variablen bzw. Parameter für die Generierung von Realisierungen anderer Variablen verantwortlich sind. Die Rolle dieser Variablen bzw. Parameter lässt sich in der rechten Spalte ablesen, wo auch die Verteilung der jeweils betroffenen Zufallsvariable angegeben ist.

Wie man sehen kann, wurde die Basisverteilung *H* des HDP derart gewählt, dass ein Sample aus *H* einem *D*-Tupel $(\theta^1, ..., \theta^D) \sim H$ mit $\theta^d \sim \text{Beta}(\beta_0{}^d, \beta_1{}^d)$ entspricht. Dabei können die Parameter der Betaverteilungen wahlweise entweder direkt festgelegt oder über einen Gammaprior mit $\beta_0^d, \beta_1^d \sim \text{Gamma}(\gamma_0{}^\beta, \gamma_1{}^\beta)$ beschrieben werden. Bei Letzterem handelt es sich um einen Prior,



$$
\begin{aligned}
\gamma &\sim \text{Gamma}(\gamma_{0,}^{\gamma}\gamma_1^{\gamma}) \\
\alpha &\sim \text{Gamma}(\gamma_{0,}^{\alpha}\gamma_1^{\alpha}) \\
\beta_0^d,\beta_1^d &\sim \text{Gamma}(\gamma_{0,}^{\beta}\gamma_1^{\beta}) \\
\theta^d &\sim \text{Beta}(\beta_{0,}^{d}\beta_1^{d}) \\
\boldsymbol{\theta}=(\theta^1,\ldots,\theta^D) &\sim H \\
G_0 &\sim \text{DP}(\gamma, H) \\
G_j &\sim \text{DP}(\alpha, G_0) \\
\boldsymbol{\theta_{ji}} &\sim G_j \\
o_{ji} &\sim L(\boldsymbol{\theta_{ji}})
\end{aligned}
$$

Abbildung 2. Das Kategorisierungsmodell URM von Griffiths et al. (2007). *L* stellt eine multivariate Verteilung konstruiert aus *D* unabhängigen Bernoulliverteilungen dar.

welcher bei ganzzahligem Shape-Parameter γ_0 als eine Wahrscheinlichkeitsverteilung über die Summe von γ_0 unabhängigen negativ exponentialverteilten Zufallsvariablen mit Parameter γ_1 und Dichtefunktion

$$p_{\gamma_1}(x) = \gamma_1 \cdot e^{-\gamma_1 \cdot x} \quad \text{für } x \geq 0 \tag{29}$$

interpretiert werden kann (Amari & Misra, 1997). Die konkrete Dichtefunktion der Gammaverteilung lautet dabei:

$$p(x;\gamma_0,\gamma_1) = \frac{\gamma_1^{\gamma_0} \cdot x^{\gamma_0 - 1} \cdot e^{-x \cdot \gamma_1}}{\Gamma(\gamma_0)} \tag{30}$$

Wie in Abbildung 2 zu sehen ist, wird die Gammaverteilung nicht nur optional als Prior für die $\beta_{0,}^{d}\beta_1^{d}$ verwendet. Auch die Konzentrationsparameter α und γ des HDP werden über diese Verteilung beschrieben.

2.6 Sampling mit dem URM

Ein Sampling Verfahren für das URM wurde bereits mit dem CRF Sampling für HDPMMs in (23-28) vorgestellt. Es hat den Nachteil, dass einiges an Buchführung über die aktuelle Tisch- und Speisebelegung im CRF notwendig ist, damit das Sampling angewendet werden kann. Aus diesem Grund haben Teh et al. (2006) ein weiteres Verfahren vorgestellt, bei dem die Zuordnung der Beobachtungen jedes Knotens zu einer Speise *k* direkt durchgeführt wird. Dadurch sinkt der Variablenverwaltungsaufwand, was wiederum in einer softwaretechnisch weniger aufwändigen Implementation resultiert. Dieses als *Posterior-Sampling-by-Direct-Assignment* (PSDA) benannte Verfahren wurde für alle Modellierungen dieser Arbeit verwendet und soll im Folgenden vorgestellt werden:

Wir starten wie im CRF Verfahren mit der restaurantweisen Einsortierung der Beobachtungen. Diesmal wählen wir jedoch kein t_{ji} mit k_{jt} für jeden Gast o_{ji}, sondern direkt ein $z_{ji} = k_{jt_{ji}}$. Das unmittelbare Wählen eines z_{ji} entspricht somit einer direkten Zuordnung von Gast zu einer Speise im CRF.

Die Konditionalverteilung für die Zuweisung eines im Restaurant *j* eintreffenden Gastes o_{jN+1} zu einer Speise *k* ist dann folgendermaßen:

$$p(z_{jN+1} = k \mid o_{jN+1}, \boldsymbol{z}, \boldsymbol{b}) = \begin{cases} (n_{jk} + \alpha \cdot b_k) \cdot F_k(o_{jN+1}) & \text{für bereits servierte Speise } k \\ \alpha \cdot b_u \cdot F_{new}(o_{jN+1}) & \text{für bisher unservierte Speise} \end{cases} \quad (31)$$

n_{jk} bezeichnet die bisherige Anzahl der Gäste im Restaurant *j*, welche bereits Speise *k* zu sich nehmen, *α* ist der gemeinsame Konzentrationsparameter der Restaurants $G_{1,\ldots,J}$ und $F_k(o_{jN+1})$ bzw. $F_{new}(o_{jN+1})$ die jeweilige Likelihood aus (27). Für das URM werden diese Likelihoods für Gast o_{jN+1} mit Ausprägung o^d_{jN+1} in Dimension *d* bei Parametern β^d_0, β^d_1 der Betaverteilung für die *d*. Dimension, sowie der Anzahl an bisherigen, im gesamten CRF vorliegenden Beobachtungen O^d_{kv} von Ausprägungen *v* in Dimension *d* bei Gästen, welche Speise *k* verzehren, wie folgt berechnet:

$$F_k(o_{jN+1}) = \prod_{d=1}^{D} \left(\frac{\beta^d_{o^d_{jN+1}} + O^d_{ko^d_{jN+1}}}{\beta^d_0 + \beta^d_1 + O^d_{k0} + O^d_{k1}} \right)$$

$$F_{new}(o_{jN+1}) = \prod_{d=1}^{D} \left(\frac{\beta^d_{o^d_{jN+1}}}{\beta^d_0 + \beta^d_1} \right) \quad (32)$$

Es handelt sich somit bei $F_{new}(o_{jN+1})$ um die Multiplikation der Erwartungswerte (für $o^d_{jN+1} = 0$) bzw. von 1 subtrahierten Erwartungswerte (für $o^d_{jN+1} = 1$) der Betaverteilungen der *D* Dimensionen. Bei $F_k(o_{jN+1})$, welches für bereits servierte Speisen im Restaurant Franchise Anwendung findet, werden konsequenterweise aktualisierte Betaverteilungen Beta($\beta^d_0 + O^d_{k0}, \beta^d_1 + O^d_{k1}$) für jede Dimension *d* verwendet.

Die b_k und b_u aus (31) stellen den Wahrscheinlichkeitsanteil der Basisverteilung *H* (b_u), sowie der Cluster *k* (b_k) eines Samples aus der Posterior von G_0 dar und werden aus einer Dirichletverteilung gezogen (Teh & Jordan, 2010). Sie sind verteilt als:

$$(b_{1,..}, b_K, b_u) \sim \text{Dir}(m_{1,..}, m_K, \gamma) \tag{33}$$

Dabei bezeichnen die $m_1,\ldots,m_K$ die Anzahl der Tische an denen die Speisen 1,..,*K* bereits serviert werden. Dieses Vorgehen ergibt sich aus der Betrachtung der Dirichletverteilungen aus (14) und (33) als zwei aggregierte Verteilungen aus einer ursprünglichen Dir($\gamma H(C_1),\ldots,\gamma H(C_K), m_1,\ldots,.m_K$), wobei mit Aggregation die Anwendung der additiven Zusammenfassungsregel von Tupelkomponenten einer Ziehung sowie deren zugehörigen Verteilungsparameter gemeint ist (Frigyik et al., 2010). Nach Teh et al. (2006) gilt dabei folgender Zusammenhang zu einem Sample G_0^{Post} aus der Posterior von G_0:

$$\begin{aligned} G_u &\sim \text{DP}(\gamma, H) \\ G_0^{Post} &= \sum_{k=1}^{K} b_k \cdot \delta_{\theta_k} + b_u \cdot G_u \end{aligned} \tag{34}$$

Für das Resampling jedes dieser m_k müssen alle n_{jk} Gäste mit Speise *k* des aktuell betrachteten Restaurants *j*, das Restaurant kurzzeitig verlassen und dann entsprechend einem CRP nach (19) an eine neue Anzahl an Tischen *m* mit Speise *k* platziert werden. Das neue m_k entspricht dann dem alten m_k minus der ursprünglichen Anzahl an Tischen mit Speise *k* im Restaurant *j* plus der eben ermittelten Anzahl an neuen Tischen *m*.

Schließlich sind noch neue Konzentrationsparameter α und γ zu wählen. Diese werden nach folgendem Verfahren ermittelt, welches hier beispielhaft nur für α demonstriert wird. Das Verfahren für γ ist identisch. Zunächst werden die Werte zweier Hilfsvariablen w_j und s_j für jedes Restaurant *j* mit n_j Gästen über folgende Konditionalverteilungen bestimmt:

$$\begin{aligned} q(w_j|\alpha) &\sim \text{Beta}(\alpha+1, n_j) \\ q(s_j|\alpha) &\sim \text{Bernoulli}\left(\frac{n_j}{\alpha + n_j}\right) \end{aligned} \tag{35}$$

Danach wird α gezogen aus:

$$q(\alpha|\boldsymbol{w},\boldsymbol{s}) \sim \text{Gamma}(\gamma_0^{\gamma} + (\sum_{K=1}^{K} m_k) - (\sum_{j=1}^{J} s_j), \gamma_1^{\gamma} - (\sum_{j=1}^{J} \log(w_j))) \quad (36)$$

Um nun ein Sample aus dem URM über PSDA zu generieren, werden die Ziehungen aus den Zufallsvariablen in folgender Reihenfolge durchgeführt: Zunächst wird restaurantweise allen Gästen o_{ji} im Restaurant *j* über (31) eine neue Speise *k* zugeordnet und anschließend die m_k über das oben beschriebene CRP Resampling angepasst. Wurden diese zwei Teilschritte für jedes Restaurant *j* durchgeführt, werden für jedes Restaurant neue Betagewichte über (33) gezogen. Abschließend folgt die Wahl neuer Konzentrationsparameter aus der jeweiligen Gammaverteilung über (35-36). Das Vorgehen wiederholt sich für das nächste Sample mit den Werten der Variablen aus dem aktuellen Sample beginnend ab (31).

Teh et al. (2006) haben ein Softwarepaket bereitgestellt, welches die Berechnungen (31, 33-36) automatisiert. Es wurde von Griffiths et al. (2007) modifiziert, sodass die vorgestellte Basisverteilung *H* (siehe Abb. 2) und (32) verwendet werden. Zusätzlich wurde ein Gammaprior für die Parameter der Betaverteilungen $\beta_{0,}^{d}\beta_1^{d}$ definiert, welche über ein Metropolis-Hastings Sampling (Chib & Greenberg, 1995; Hastings, 1970) ermittelt werden. Hierfür werden zunächst für die Parameter $\beta_{0,}^{d}\beta_1^{d}$ einer Dimension *d* über eine sogenannte Proposal-Distribution neue Werte gezogen. Anschließend wird eine Bernoulliverteilung bestimmt, welche die Wahrscheinlichkeiten beschreibt, dass die neuen Parameterwerte angenommen bzw. abgelehnt werden. Führt eine darauffolgende Ziehung aus dieser Bernoulliverteilung zu einer Ablehnung, werden die alten Werte beibehalten, ansonsten werden die neuen Werte übernommen. Wie auch beim Gibbs Sampling führt diese Methode bei wiederholter Anwendung zu einer schrittweise besseren Approximation der Posteriorverteilung der Variablen $\beta_{0,}^{d}\beta_1^{d}$. Als Proposal-Distribution verwenden Griffiths et al. (2007) für jedes $\beta_v^d \; für \; v=0,1$ eine Log-Normalverteilung mit Median gleich dem noch aktuellen Parameterwert $\beta_{v^{old}}^{d}$ und einem Scale Parameter *σ*. Ein neuer Parameterwert $\beta_{v^{new}}^{d}$ wird dann vorgeschlagen über

$$\beta_{v^{new}}^{d} = e^{\log(\beta_{v^{old}}^{d}) + \sigma \cdot Z} \quad (37)$$

wobei *Z* eine Ziehung aus einer Standardnormalverteilung darstellt. Die alten und neuen Wertepaare $\boldsymbol{\beta_{old}^{d}} = (\beta_{0^{old}}^{d}, \beta_{1^{old}}^{d}), \boldsymbol{\beta_{new}^{d}} = (\beta_{0^{new}}^{0}, \beta_{1^{new}}^{d})$ bestimmen dann den Parameter *p* der Bernoulliverteilung. Die Wahrscheinlichkeit *p* setzt sich dabei zusammen aus einem Faktor, welcher sich aus der Wahl der Proposal-Distribution ergibt, der Verhältnisse der Wahrscheinlichkeiten (jeweils für *v* = 0,1) aus

einem Gammaprior ein $\beta^d_{v^{new}}$ und ein $\beta^d_{v^{old}}$ zu ziehen und des Verhältnisses der Wahrscheinlichkeiten die aktuell bestehenden Gäste-zu-Speise Zuordnungen mit $\boldsymbol{\beta}^{\boldsymbol{d}}_{\boldsymbol{new}}$ und $\boldsymbol{\beta}^{\boldsymbol{d}}_{\boldsymbol{old}}$ zu erhalten. Der Parameter wird folgendermaßen berechnet:

$$p(\boldsymbol{\beta}^{\boldsymbol{d}}_{\boldsymbol{old}}, \boldsymbol{\beta}^{\boldsymbol{d}}_{\boldsymbol{new}}) = \prod_{v=0}^{1} \left(\frac{\beta^d_{v^{new}}}{\beta^d_{v^{old}}} \cdot e^{(\gamma_0^\beta - 1) \cdot \log\left(\frac{\beta^d_{v^{new}}}{\beta^d_{v^{old}}}\right) + \frac{\beta^d_{v^{old}} - \beta^d_{v^{new}}}{\gamma_1^\beta}} \right) \cdot \prod_{k=0}^{K} \left(\prod_{v=0}^{1} \prod_{i=0}^{O^d_{kv}-1} \frac{\beta^d_{v^{new}} + i}{\beta^d_{v^{old}} + i} \cdot \prod_{i=0}^{O^d_{k0}+O^d_{k1}-1} \left(\frac{\sum_{v=0}^{1} (\beta^d_{v^{old}}) + i}{\sum_{v=0}^{1} (\beta^d_{v^{new}}) + i} \right)^2 \right) \tag{38}$$

Eine Ziehung aus einer Bernoulliverteilung mit diesem Parameter entscheidet dann über das nächste Parameterwertpaar $\boldsymbol{\beta}^{\boldsymbol{d}} = (\beta^d_0, \beta^d_1)$:

$$F(\boldsymbol{\beta}^{\boldsymbol{d}} = \boldsymbol{\beta} \mid \boldsymbol{\beta}^{\boldsymbol{d}}_{\boldsymbol{old}}, \boldsymbol{\beta}^{\boldsymbol{d}}_{\boldsymbol{new}}) = \begin{cases} p & \text{für } \boldsymbol{\beta} = \boldsymbol{\beta}^{\boldsymbol{d}}_{\boldsymbol{new}} \\ 1-p & \text{für } \boldsymbol{\beta} = \boldsymbol{\beta}^{\boldsymbol{d}}_{\boldsymbol{old}} \end{cases} \tag{39}$$

(37-39) werden nacheinander für alle *D* Dimensionen durchgeführt. Sollen die Betaverteilungen mit den Parametern $\beta^d_{0,}\beta^d_1$ immer symmetrisch sein ($\beta^d_0 = \beta^d_1$), so besteht die Möglichkeit eine entsprechende Einschränkung zu definieren. In diesem Fall wird immer $\beta^d_{1^{new}} = \beta^d_{0^{new}}$ gesetzt und der erste Multiplikationsblock in (38) nur für *v*=0 berechnet.

2.7 Durchführung einer Simulation mit dem URM

In der vorliegenden Arbeit wird das von Teh et al. (2006) veröffentlichte und von Griffiths et al. (2007) erweiterte Softwarepaket verwendet, welches im Weiteren als npBayes2.1G bezeichnet wird. npBayes2.1G ermöglicht es nach der Bereitstellung einer spezifischen Anzahl an Trainingsstimuli für jedes Restaurant *j* (= die Beobachtungen) Samples aus der Posterior des URM zu ziehen und die Generierungswahrscheinlichkeit von ausgewählten Teststimuli zu berechnen. Jedes Restaurant *j* repräsentiert dabei eine eigenständige Kategorie. Wenn nun *S* Samples aus dem URM gezogen wurden, kann die Kategorisierungswahrscheinlichkeit $p_j^{cat}(x)$ eines Teststimulus *x* für eine Kategorie *j* über dessen Generierungswahrscheinlichkeiten $p_{js}^{gen}(x)$ der Kategorien $j = 1,\ldots,J$ in jedem Sample *s* mit

$$
\begin{aligned}
p_{js}^{gen}(x) &= \sum_{k \in \{k \mid n_{sjk} > 0\}} \left(F_{sk}(x) \cdot \frac{n_{sjk}}{N+\alpha_s}\right) + \left(F_{snew}(x) \cdot \frac{\alpha_s}{N+\alpha_s}\right) \\
&= \sum_{k \in \{k \mid n_{sjk} > 0\}} \left(\prod_{d=1}^{D} \left(\frac{\beta_{x^d}^d + O_{skx^d}^d}{\beta_0^d + \beta_1^d + O_{sk0}^d + O_{sk1}^d} \right) \cdot \frac{n_{sjk}}{N+\alpha_s} \right) + \prod_{d=1}^{D} \left(\frac{\beta_{x^d}^d}{\beta_0^d + \beta_1^d} \right) \cdot \frac{\alpha_s}{N+\alpha_s}
\end{aligned}
\tag{40}
$$

durch Luces (1959) Auswahlregel bestimmt werden:

$$
p_j^{cat}(x) = \frac{\sum_{s=1}^{S} p_{js}^{gen}(x)}{\sum_{i=1}^{J} \sum_{s=1}^{S} p_{is}^{gen}(x)} \tag{41}
$$

Sind hingegen keine Kategorisierungen vorherzusagen, sondern dimensionale Ausprägungen, lässt sich ebenfalls Luces (1959) Auswahlregel für die Berechnung der Inferenzwahrscheinlichkeiten anwenden. Die Wahrscheinlichkeit, dass ein Stimulus x mit unbekannter Ausprägung x^d in der Dimension d einen Wert v aufweist, wird deshalb nach Vorliegen von S Samples folgendermaßen berechnet:

$$
p_j^{inf}(x^d = v) = \frac{\sum_{s=1}^{S} p_{js}^{gen}\left(x^{x^d = v}\right)}{\sum_{i=1}^{V} \sum_{s=1}^{S} p_{js}^{gen}\left(x^{x^d = i}\right)} \tag{42}
$$

Dabei ist mit $x^{x^d = i}$ die Stimulusvariante von x gemeint, in welchem die unbekannte Stimulusausprägung x^d den Wert i aufweist.

Eine klassische Simulation eines supervidierten Kategorisierungsexperiments mit zwei Antwortmöglichkeiten pro Trial verläuft dann trialweise folgendermaßen:

1. Es werden für npBayes2.1G alle in den vorherigen Trials dargebotenen Stimuli entsprechend ihrer Kategorienzugehörigkeit als Trainingsstimuli der jeweiligen Kategorie deklariert.
2. Der zu kategorisierende Stimulus im aktuellen Trial wird für jede Kategorie j als Teststimulus festgelegt.
3. Es wird (40) über npBayes2.1G für jedes Sample aus der Posterior des URM berechnet und dann (41) für jede Kategorie bestimmt. Ist man an der Wahrscheinlichkeit der Kategorisierungskorrektheit bzw. des Kategorisierungsfehlers interessiert, so erhält man diese über $p_j^{cat}(x)$ bzw. $1 - p_j^{cat}(x)$ wobei j der Index der korrekten Kategorie ist.

Im realen supervidierten Experiment wird nun ein Feedback gegeben. Für die Simulation heißt das, dass der aktuelle Teststimulus für nachfolgende Trials als Trainingsstimulus der korrekten Kategorie fungiert. Es wird deshalb für den nächsten Trial unmittelbar mit Schritt 1 fortgefahren.

Soll anstatt eines Experiments mit Kategorisierungsaufgaben, ein Experiment mit Inferenzaufgaben simuliert werden, sind trialweise in Schritt 2 Stimulusvarianten des Stimulus von Interesse mit jeder möglichen Ausprägung in der abgefragten Dimension als eigener Teststimuli zu deklarieren. Anschließend wird in Schritt 3 (42) statt (41) berechnet, wobei sich auch hier der Inferenzfehler über $1 - p_j^{inf}(x^d = v)$ ergibt.

Ist hingegen kein supervidiertes, sondern ein unsupervidiertes Experiment zu simulieren, wird nur eine Kategorie für alle Trainingsstimuli in Schritt 1 verwendet, wobei dieser Schritt üblicherweise auch nur einmalig in der Simulation eines solchen Experiments ausgeführt wird. Anschließend können beispielsweise die Schritte 2 und 3 abwechselnd wiederholt werden bis alle Inferenzaufgaben einer Testphase abgearbeitet sind, wobei hier natürlich aufgrund der Natur des Experiments ein Feedback ausbleibt.

Um oben beschriebene Simulationen durchführen zu können, sind neben den in Abbildung 2 beschriebenen Parametern für die Ausführung der Simulationsschritte (31-39) durch die Software npBayes2.1G noch fünf weitere Parameter festzulegen: das sogenannte Burn-In, die Anzahl der Samples, das sogenannte Thinning, die Anzahl der Iterationen des Metropolis-Hastings Samplers für (37-39) und der Scale Parameter der zugehörigen Proposal-Distribution. Ersteres bezeichnet die Anzahl an URM Samples, welche zwar berechnet jedoch verworfen werden. Der Grund hierfür ist, dass die willkürlich gewählten Anfangswerte der zu sampelnden Variablen eine ungünstige Startposition auf der Posterior darstellen könnten (Kruschke, 2010). In diesem Fall benötigt das PSDA einige Durchläufe um an repräsentativeren Stellen der Posterior zu gelangen. Verwirft man nun diese unrepräsentativen Durchläufe (= Burn-In) so vermeidet man auch eine anfängliche Verfälschung der approximierten Posterior. Im Anschluss an das Burn-In erfolgt die eigentliche Auswahl der relevanten URM Samples. Es werden dabei Samples $\times$ Thinning Gibbs Sampling Iterationen durchgeführt, wobei nur jedes Thinning-te Sample gespeichert wird. Diese Strategie erlaubt es die Autokorrelation zwischen den Samples zu verringern und somit eine aussagekräftigere Approximation bei gleicher Anzahl an Samples zu erhalten (Lee & Wagenmakers, 2014; Lynch, 2007). Die Verwendung von Thinning kann beispielsweise vorteilhaft sein, wenn das post-processing von Samples aufwendig ist und sich daher eine Erhöhung der Anzahl an Samples ohne Thinning zur Verbesserung der Approximation nicht lohnt (Link & Eaton, 2012).

Nicht immer ist man beim URM nur an den Generierungswahrscheinlichkeiten von Stimuli interessiert. So könnten wir auch (wie z.B. in Experiment 7) wissen wollen, welche konkrete Partition der Stimuli das Modell in einer Kategorie bevorzugt. Da npBayes2.1G die Beantwortung dieser Frage nicht unmittelbar erlaubt, wurde hierfür der nachfolgend beschriebene Ansatz gewählt.

2.8 Bestimmung des repräsentativen Clusterings

Um ein repräsentatives Clustering des URM zu erhalten, wird die mittlere der S Partitionen von Beobachtungen o_{ji} in K_s Cluster aus S URM Samples nach dem Verfahren von Huelsenbeck und Andolfatto (2007) bestimmt. Hierbei wird eine beliebige initiale Partition als vorläufige mittlere Partition definiert und iterativ optimiert, sodass diese eine möglichst niedrige mittlere quadrierte Distanz zu den Partitionen von Interesse aufweist. Das Verfahren endet, sobald für kein Item der vorläufigen mittleren Partition ein neuer Cluster gewählt werden kann, sodass die mittlere quadrierte Distanz nochmals sinkt. Der Ablauf des Algorithmus ist wie folgt:

Algorithm 1 Calculation of the mean partition

Input: partitions $P_1,\ldots,P_M$ of items $I_1,\ldots,I_N$
Output: mean partition *mp* of the partitions $P_1,\ldots,P_M$
  mp ← an arbitrary partition of the $P_1,\ldots,P_M$
  distance ← mean squared distance of *mp* to $P_1,\ldots,P_M$
  success ← true
  while *success* = true **do**
    success ← false
    for all *item* ∈ $\{I_1,\ldots,I_N\}$ **do**
      for all *cluster* ∈ *mp* **do**
        if item ∈ *cluster* **then**
          if $|mp| = N$ **then**
            continue with the iteration for the next *cluster* ∈ *mp*
          end if
          move *item* temporarily into an empty cluster
        else
          move *item* temporarily into *cluster*
        end if
        if mean squared distance of new temporarily *mp* to $P_1,\ldots,P_M <$ *distance* **then**
          keep new *mp* and save new *distance*
          success ← true
        else
          keep old *mp*
        end if
      end for

end for
end while
return *mp*

Die Distanz zwischen zwei Partitionen ist definiert als die Anzahl an Items, welche aus beiden Partitionen entfernt werden müssen, damit die Partitionen identisch sind (Gusfield, 2002). Ein erster Algorithmus zur Berechnung dieser Distanz wurde von Almudevar und Field (1999) vorgeschlagen, welcher jedoch aufgrund seiner rekursiven Konstruktion und einer exponentiellen Laufzeit (Gusfield, 2002) eine in der Praxis ungünstig lange Rechenzeit bei hohem Speicherbedarf aufweist (Berger-Wolf et al., 2007). Konovalov, Litow und Bajema (2005) konnten jedoch zeigen, dass die Berechnung der Partitionsdistanz auf das Lösen eines *Linear-Sum-Assignment-Problem* (LSAP) in polynomialer Zeit reduzierbar und dann in $O(n^3)$ durchführbar ist. Bei Letzterem handelt es sich um das Problem aus einer $n \times n$ Kostenmatrix n Zellen auszuwählen, sodass aus jeder Zeile und jeder Spalte nur eine Zelle verwendet wird und die Summe der n ausgewählten Kosten minimal ist (Burkard, Dell'Amico, & Martello, 2012).
Der verwendete Algorithmus, welcher entsprechend der Reduktion von Konovalov et al. (2005) die Distanz zweier Partitionen über die Lösung eines LSAP berechnet, ist folgendermaßen:

Algorithm 2 Calculation of the partition distance

Input: partitions P_1, P_2 of items $I_1,\ldots,I_N$
Output: distance between the partitions P_1, P_2
if $|P_1| < |P_2|$ **then**
Switch partition meanings of P_1 and P_2
end if
$M \leftarrow$ matrix of size $|P_1| \times |P_2|$ filled with zeros
for all $item \in \{I_1,\ldots,I_N\}$ **do**
$i \leftarrow$ cluster of *item* in P_1
$j \leftarrow$ cluster of *item* in P_2
$M_{i,j} \leftarrow M_{i,j} - 1$
end for
for all $i \in P_1$ **do**
for all $j \in P_2$ **do**
$M_{i,j} \leftarrow M_{i,j} + |i|$
end for
end for

return *minimum costs of a linear sum assignment problem with cost matrix M*

Das Zuweisungsproblem wurde über das Softwarepaket von Buehren (2014) gelöst, welches den Algorithmus von Munkres (1957) verwendet. Letzterer ist eine Verbesserung von $O(n^4)$ auf $O(n^3)$ Jonker & Volgenant, 1986; Riesen & Bunke, 2009) des ursprünglichen Algorithmus von Kuhn (1955), welcher unter dem Namen „Ungarische Methode“ bekannt ist. Die „Ungarische Methode“ lässt sich am einfachsten als mehrstufige Manipulation eines gerichteten bipartiten Graphen G_k mit Knotenmengen W und T mit $|W| \leq |T|$, welche die Arbeiter und die Aufgaben repräsentieren und der Kantenmenge $E_k = \{(w,t) \mid w \in W, t \in T, p_k(w) + p_k(t) = c_{w,t}\}$ erklären, wobei k die Anzahl der bereits durchlaufenen Stufen beschreibt (siehe auch Mills-Tettey, Stentz, & Dias, 2007; Suri, 2006). p_k stellt dabei eine Funktion dar, welche jedem Knoten der Menge $W \cup T$ ein sogenanntes Potential zuordnet. Die Potentiale sind derart gewählt, dass stets die Ungleichung $p(w) + p(t) \leq c_{w,t}$ gilt, wobei $c_{w,t}$ die entstehenden Kosten beschreiben, wenn Arbeiter w die Aufgabe t ausführt. Die Kosten jeder Kombination von $w \in W$ und $t \in T$ sind in der Kostenmatrix C zusammengefasst, in welcher jede Zelle in Zeile i und Spalte j die Kosten der Zuteilung von Arbeiter i und Aufgabe j benennt. Neben der Kantenmenge E_k des Graphen G_k ist eine weitere stufenspezifische Kantenmenge M_k definiert, welche alle Kanten aus E_k umfasst, die von einem Knoten $t \in T$ zu einem Knoten $w \in W$ gerichtet sind. Zusätzlich liegt in jeder Stufe k eine Knotenmenge vor, welche Knoten aus W und T enthält und folgendermaßen definiert ist: $R_k = \{w, v \mid w \in W$ ist nicht durch eine Kante in M_k bedeckt $\vee$ w ist Startpunkt eines Pfades entlang gerichteter Kanten in E_k, welcher zu einem Knoten $v \in W \cup T$ führt$\}$.

Der Prozess beginnt in Stufe 0, in welcher alle Knoten in W und T ein Potential von 0 besitzen und alle Kanten in E_0 von Knoten aus W nach Knoten aus T gerichtet sind. Anschließend werden pro Stufe drei bzw. vier der folgenden vier Schritte, beginnend ab Schritt 1 durchgeführt:

1. Überprüfe ob die Anzahl der Kanten in M_k gleich der Anzahl der Knoten in W ist. Wenn ja, dann stoppe und gebe M_k als Lösung aus, ansonsten gehe zu Schritt 2.
2. Überprüfe ob es einen Knoten $t \in T \cap R_k$ gibt, welcher nicht durch eine Kante in M_k überdeckt ist. Wenn ja, dann gehe zu Schritt 3, ansonsten zu Schritt 4.
3. Suche einen Pfad entlang gerichteter Kanten, welcher an einem Knoten $w \in W \cap R_k$ beginnt und zu einem Knoten $t \in T \cap R_k$ führt, welcher nicht durch eine Kante in M_k überdeckt ist. Wenn der Pfad gefunden ist, ändere die Richtung jeder Kante des Pfades, wodurch die Anzahl der gerichteten Kanten von T nach W um eins zunimmt. Aktualisiere die Kantenmenge M_k und definiere G_k mit E_k,

M_k und R_k als Mengen und Graphen der Stufe k+1. Gehe anschließend zu Schritt 1 und setze den Prozess in der neuen Stufe k+1 fort.

4. Berechne $\Delta = min(\{ c_{w,t} - p_k(w) - p_k(t) \mid w \in W \cap R_k, t \in T \setminus R_k \})$ und redefiniere die Potentialfunktion p_k der aktuellen Stufe für alle $w \in W \cap R_k$ und alle $t \in T \cap R_k$, sodass nun $p_k(w) = p_k(w) + \Delta$ und $p_k(t) = p_k(t) - \Delta$ gilt. Aktualisiere die Kantenmenge E_k sowie die Knotenmenge R_k und gehe anschließend zurück zu Schritt 3.

Der Prozess endet spätestens nach der Abarbeitung von $|W|$ Stufen, wonach $M_{|W|}$ eine Zuordnung von Arbeitern zu Aufgaben beschreibt, sodass keine zwei Arbeiter dieselbe Aufgabe erledigen und die Gesamtkosten durch die Zuordnung minimal sind.

Es wurden nun alle URM-spezifischen Berechnungen demonstriert, sodass im nächsten Abschnitt das Kategorisierungsmodell SUSTAIN vorgestellt werden kann.

2.9 SUSTAIN

Bei SUSTAIN handelt es sich im Wesentlichen um ein dreischichtiges neuronales Feedforward-Netzwerk, dessen Struktur in der Eingabe- und Zwischenschicht nicht fix ist, sondern sich durch fortschreitendes Training des Netzes und je nach präsentierten Stimuli bezüglich der Anzahl an Knoten und zugehörigen Kanten vergrößern kann (Lake, Zaremba, Fergus, & Gureckis, 2015; Love & Medin, 1998; Love et al., 2004). Bei SUSTAIN handelt es sich wie beim URM um ein sogenanntes Clustermodell. Das heißt, dass gelernte Stimuli entsprechend ihrer Ähnlichkeit zueinander als Gruppen (= Cluster) repräsentiert und Inferenzen von unbekannten Merkmalen eines neuen Stimulus anhand der aktuellen Partition durchgeführt werden (McDonnell & Gureckis, 2011). Als Eingabe erhält das Netz einen Stimulus, bei welchem eine Dimension z für die Verarbeitung ignoriert wird und prognostiziert dessen Ausprägung an den Ausgängen des Netzwerks. Training des Netzes findet ebenfalls mit angelegtem Stimulus statt. Handelt es sich um supervidiertes Lernen, dann wird die Ausprägung des Stimulus in der Dimension z mit dem prognostizierten Wert verglichen und bei Abweichung neue Knoten in der Zwischenschicht hinzugefügt. Bei unsupervidiertem Lernen wird Letzteres ebenfalls durch Nicht-Erreichen eines Vertrautheits-Threshold durchgeführt. Unabhängig vom Lernmodus werden immer zusätzliche Gewichtsanpassungen vorgenommen.

In Abbildung 3 ist das Prognosenetzwerk von SUSTAIN dargestellt. Kreise symbolisieren Knoten des Netzwerks und repräsentieren entweder Funktionen oder Konstanten. Funktionen sind mit griechischen Buchstaben und Konstanten mit Zahlen bzw. lateinischen Variablenbezeichnungen markiert. Die gerichteten Kanten zwischen den Knoten stellen die Art deren Beziehung dar. Hat eine Kante einen ausgefüllten Kreis am Pfeilanfang, so wird der Knotenwert des Knotens am Pfeilanfang mit der Konstante am Pfeilende gewichtet und dann weitergereicht. Fehlt dieser ausgefüllte Kreis wird der Wert des jeweiligen Knotens direkt als Parameter an die nachfolgende Funktion weitergeleitet. Funktionen mit dem griechischen Buchstaben Sigma (Σ) stellen sogenannte Propagierungsfunktionen dar. Funktionen, welche als Aktivierungsfunktionen bzw. Ausgabefunktionen bezeichnet werden, sind mit einem kleinen Phi (ϕ) bzw. mit einem kleinen

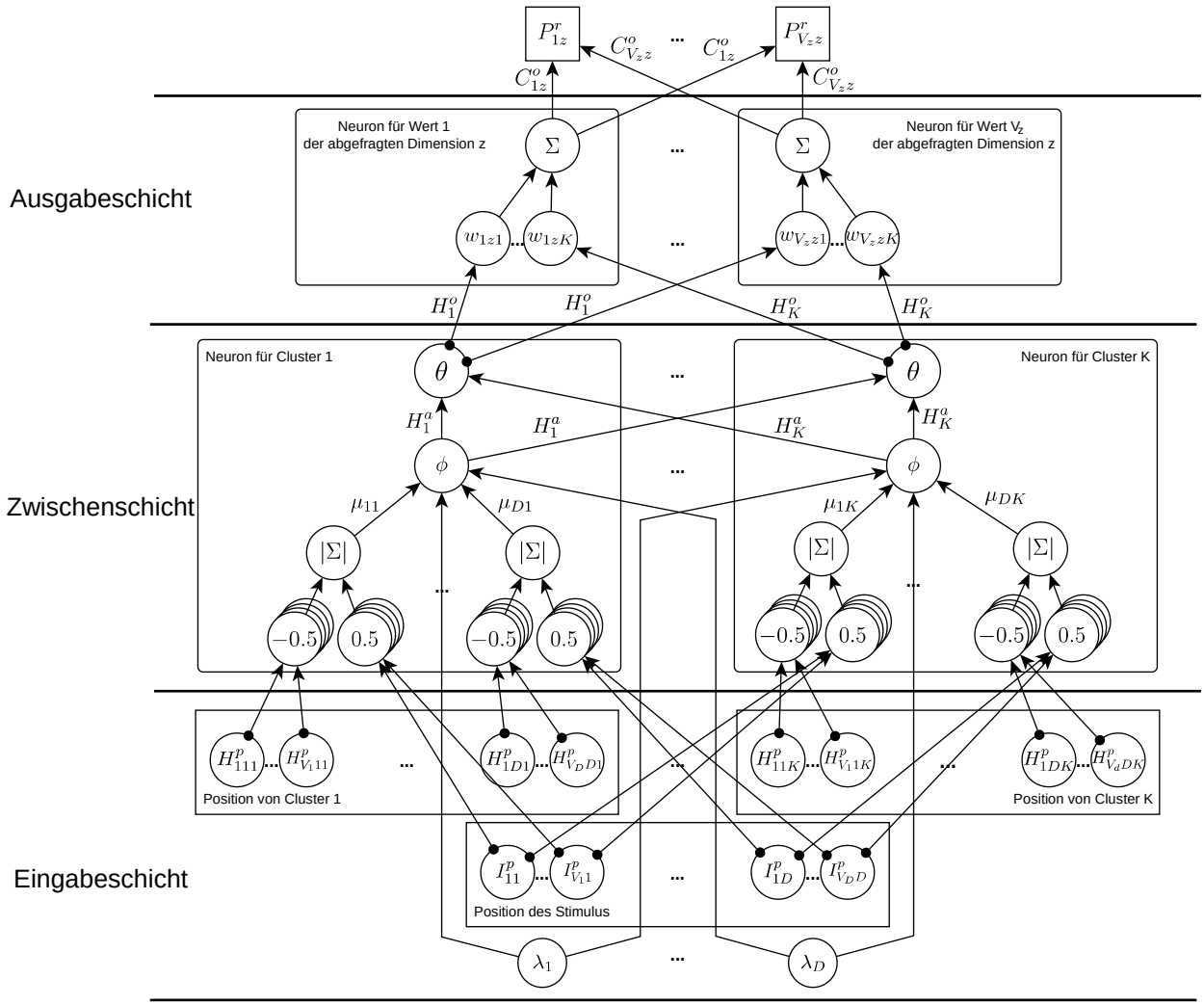


Abbildung 3. Das dreischichtige neuronale Feedforward-Netzwerk des Kategorisierungsmodells SUSTAIN von Love et al. (2004).

Theta (θ) gekennzeichnet. Die Quadrate oberhalb der Ausgabeschicht repräsentieren spezielle Knoten, welche Ausgabesignale des Netzwerks in Prognosewahrscheinlichkeiten übersetzen.

Eine typische Modellierungssituation mit sich wiederholenden Prognose-Training-Zyklen verläuft wie folgt (Gureckis, 2014; Love et al., 2004):

Zu Beginn sind keine Cluster definiert. Es befinden sich somit keine Knoten in der Zwischenschicht und die Eingabeschicht weist lediglich die Positionsinformationen I_{vd}^{p} für Stimulusausprägung $v = 1, \ldots, V_d$ auf den Dimensionen $d = 1, \ldots, D$ des angelegten Stimulus als Eingabewerte auf. Handelt es sich beispielsweise bei der ersten Dimension um die Dimension „Form“ mit drei möglichen Ausprägungen „Dreieck“, „Quadrat“ und „Kreis“, dann hat ein an das Netzwerk von SUSTAIN angelegter kreisförmiger Stimulus die Positionsinformationen $I_{11}^{p} = I_{21}^{p} = 0$ und $I_{31}^{p} = 1$, sodass nur der Positionsknoten für die Ausprägung „Kreis“ mit 1 markiert ist. Die Ausgangssignale C_{vz}^{o} werden bei nicht vorhandenen Cluster direkt auf 0 gesetzt und anschließend die Wahrscheinlichkeit P_{vz}^{r}, dass die unbekannte Dimension z des neuen Stimulus die Ausprägung v aufweist, berechnet über:

$$P_{vz}^{r} = \frac{e^{\left(m \cdot C_{vz}^{o}\right)}}{\sum_{i=1}^{V_z} e^{\left(m \cdot C_{iz}^{o}\right)}} \tag{43}$$

Bei m handelt es sich um einen Response-Parameter, welcher bereits vorhandene Unterschiede in den Ausgangssignalen C_{vz}^{o} verstärkt oder abschwächt, sodass größere oder geringere Sicherheit bei der Prognose der unbekannten Ausprägung in der Dimension z vorliegt. Wenn keine Cluster existieren und deshalb die C_{vz}^{o} auf 0 gesetzt werden, hat m keinen Einfluss. Alle P_{vz}^{r} haben dann mit $P_{vz}^{r} = V_z^{-1}$ *für* $v = 1, \ldots, V_z$ den gleichen Wert, welcher der Wahrscheinlichkeit entspricht, die richtige Ausprägung ohne Vorwissen zu erraten.

In der anschließenden Lernphase werden zunächst die Positionsinformationen des angelegten Stimulus als Positionsinformationen eines neu generierten ersten Clusters verwendet und die Kantenwerte im Prognosenetzwerk aktualisiert. Anschließend folgt eine Optimierung der Parameter und Gewichte des Netzwerks.

In einem ersten Aktualisierungschritt werden die Dimensionsdistanzen μ_{d1} zwischen den Positionen von Stimulus und erstem Cluster in den Dimensionen d (mit Ausnahme der unbekannten Dimension z) über die Positionsinformationen I_{vd}^{p} und H_{vd1}^{p} bestimmt:

$$\mu_{dk} = \frac{1}{2} \cdot \sum_{v=1}^{V_d} |I_{vd}^{p} - H_{vdk}^{p}| \tag{44}$$

μ_{dk} *für* $d \in \{1,..,D\} \setminus \{z\}$; $k = 1,...,K$ stellen dabei die Werte der mit den gewichteten Eingangsknoten parametrisierten Propagierungsfunktionen $|\Sigma|$ in der Zwischenschicht des Prognosenetzwerkes von SUSTAIN dar. Anschließend wird die Aktivierung H_1^a des ersten Clusters anhand der zuvor berechneten Distanzen μ_{d1} über die zugehörige Aktivierungsfunktion Φ in der Zwischenschicht ermittelt:

$$H_k^a = \frac{\sum_{d \in \{1,...,D\} \setminus \{z\}} (\lambda_d)^r \cdot e^{-\lambda_d \cdot \mu_{dk}}}{\sum_{d \in \{1,...,D\} \setminus \{z\}} (\lambda_d)^r} \tag{45}$$

Die λ_d stellen Parameter der Dichtefunktion (29) einer negativen Exponentialverteilung dar. Letztere werden in SUSTAIN für die Aktivierungsmodellierung von dimensionsspezifischen rezeptiven Feldern verwendet, welche als Laplace-Verteilungen mit Lageparameter gleich der Clusterposition repräsentiert sind. Bei den λ_d handelt es sich somit um Tuning-Parameter welche bestimmen, wie steil die Exponentialverteilung ausfällt. Vor Beginn des ersten Trainings weisen sie einen initialen Wert von 1 auf. Bei *r* handelt es sich um einen Aufmerksamkeitsparameter, welcher den Einfluss von Dimensionen entsprechend ihrem Tuning-Parameter reguliert. Ist *r* hoch, so haben Dimensionen mit hohem Tuning-Parameter größeren Einfluss. Bei $r = 0$ werden hingegen alle Dimensionen gleichbehandelt.

In SUSTAIN konkurrieren Cluster um den Einfluss auf die Prognose des unbekannten Dimensionswertes. Dabei bestimmt lediglich der am stärksten aktivierte Cluster wie groß die Ausgangssignale C_{vz}^o ausfallen. Hierzu wird zunächst die Ausgangsaktivierung H_k^o jedes Clusters über die zugehörige Ausgabefunktion θ in der Zwischenschicht bestimmt:

$$H_k^o = \begin{cases} \frac{(H_k^a)^\beta}{\sum_{i=1}^{K} (H_i^a)^\beta} \cdot H_k^a & \text{für Cluster mit stärkster Aktivierung} \\ 0 & \text{für jeden anderen Cluster} \end{cases} \tag{46}$$

β stellt dabei den sogenannten Lateral-Inhibition-Parameter dar, welcher die Stärke der Abschwächung (= Inhibition) der tatsächlichen Ausgangsaktivierung des gewinnenden Clusters durch konkurrierende Cluster beeinflusst. Ist β hoch, so ist die Inhibition durch die anderen Cluster nur gering. Die neuen Ausgangssignale C_{vz}^o werden schließlich folgendermaßen berechnet:

$$C_{vz}^{o} = \sum_{k=1}^{K} w_{vzk} \cdot H_{k}^{o} \tag{47}$$

Wenn ein neuer Cluster generiert wird, dann werden die zugehörigen w_{vdk} zunächst auf 0 gesetzt, wodurch auch alle $C_{vz}^{o} = 0 \quad für\, v = 1,\ldots,V_z$ sind.

Nach (47) sind die Kantenwerte im Prognosenetzwerk aktualisiert. Hierauf aufbauend kann nun eine Optimierung von Parametern und Gewichten im Netzwerk durchgeführt werden. Zunächst wird ein sogenanntes Teacher-Signal t_{vz} für jeden potentiellen Wert $v = 1,\ldots,V_z$ der unbekannten Dimension *z* ermittelt:

$$t_{vz} = \begin{cases} max\left(C_{vz}^{o},1\right) & \text{wenn } I_{vz}^{p} = 1 \\ min\left(C_{vz}^{o},0\right) & \text{wenn } I_{vz}^{p} = 0 \end{cases} \tag{48}$$

Danach folgt eine Anpassung der w_{vzk} über die Delta-Lernregel von Widrow und Hoff (1960) durch Addition von

$$\Delta w_{vzk} = \eta \cdot \left(t_{vz} - C_{vz}^{o}\right) \cdot H_{k}^{o} \tag{49}$$

auf den ursprünglichen Wert w_{vzk} . η ist hier die Lernrate, welche die Größe der Gewichtskorrektur pro Lernzyklus bestimmt. Anschließend folgt eine Korrektur der Position des gewinnenden Clusters *k* über die Anpassung der Positionsinformationen H_{vdk}^{p} entsprechend der sogenannten Kohonen (1982) Lernregel:

$$\Delta H_{vdk}^{p} = \eta \cdot \left(I_{vd}^{p} - H_{vdk}^{p}\right) \tag{50}$$

Analog zu (49) werden auch hier die berechneten Deltawerte auf die zugehörigen Positionsinformationen H_{vdk}^{p} aufaddiert. Die H_{vdk}^{p} lassen sich demnach als Wahrscheinlichkeiten interpretieren, dass ein Stimulus des Clusters *k* den jeweilige Wert *v* in der Dimension *d* aufweist (Love et al., 2004).

In einem abschließenden Schritt werden die Tuning-Parameter λ_d nach der sogenannten Gradient-Ascent-Methode korrigiert (siehe auch Kruse et al., 2013), indem die erste Ableitung der Dichtefunktion der negativen Exponentialverteilung der Dimension *d* des gewinnenden Clusters *c* an der Stelle λ_d , gewichtet durch die Lernrate η, zu λ_d hinzuaddiert wird:

$$\Delta \lambda_d = \eta \cdot e^{-\lambda_d \cdot \mu_{dk}} \cdot \left(1 - \lambda_d \cdot \mu_{dk}\right) \tag{51}$$

Der erste Prognose-Training-Zyklus ist damit abgeschlossen und ein erster Cluster befindet sich in der Zwischenschicht des Prognose Netzwerkes von SUSTAIN. Für jeden nachfolgenden Prognose-Training-Zyklus mit angelegten Stimulus werden (44-47) wiederholt und (43) für eine Prognose des unbekannten Dimensionswertes verwendet. Anschließend werden Teacher-Signale über (48)

erzeugt. Ist das Lernen supervidiert, dann wird ein neuer Cluster nur erstellt wenn das stärkste Ausgangssignal der C_{vz}^{o} nicht zum Wert *v* gehört, für welches das Teacher-Signal $t_{vz} = 1$ ist:

$$\text{wenn } max\left(C_{1z}^{o},\ldots,C_{V_z z}^{o}\right) \neq C_{vz}^{o} \text{ für v mit } t_{vz} = 1, \text{ dann erzeuge neuen Cluster} \quad (52)$$

Handelt es sich hingegen um unsupervidiertes Lernen, dann wird ebenfalls ein neuer Cluster erzeugt, wenn die Clusteraktivierung des gewinnenden Clusters einen Wert *τ nicht* überschreitet:

$$\text{wenn } H_k^a < \tau, \text{ dann erzeuge neuen Cluster} \quad (53)$$

Nach Erzeugen eines neuen Clusters werden (44-47) zur Aktualisierung der Kantenwerte im Prognosenetzwerk wiederholt und anschließend unabhängig vom Lernmodus (49-51) durchgeführt.

Nachdem nun beide Modelle sowie ihre Funktionsweise beschrieben wurden, soll im nächsten Abschnitt unterschiedliche Stärken beider Modelle aufgeführt, sowie das URM von anderen rationalen Kategorisierungsverfahren abgegrenzt werden.

2.10 Gegenüberstellung von URM, SUSTAIN und konkurrierenden Kategorisierungsverfahren

Das URM wurde als ein vereinigendes rationales Modell der Kategorisierung vorgeschlagen, welches eine Generalisierung von rationalen Modellen der zwei prominenten Klassifikationstheorien, der Exemplar- und Prototypensichtweise, darstellt (Griffiths et al., 2007). Erstere repräsentiert die Theorie, dass Klassifikationen nach Einzelvergleichen des einzusortierenden Objekts mit jedem Mitglied der Kategorie erfolgen, während bei Letzterer von einem Vergleich mit der zentralen Tendenz der Kategorie, dem sogenannten Prototypen, ausgegangen wird (Ross & Makin, 1999). Das URM ist dabei nicht nur in der Lage, sich entweder als klassisches Exemplar- oder Prototypenmodell zu verhalten, sondern kann, je nach Datenlage, auch hybride Repräsentationen aus exemplar- und protoypbasierten Subclustern pro Kategorie herausbilden. Dadurch lässt sich mit dem URM ein Strategiewechsel während des Lernprozesses simulieren, welchen zum Beispiel Smith und Minda (1998) in Abhängigkeit von der Größe und Art der zu lernenden Kategorien identifizierten. Die Autoren stellten in ihrem Experiment fest, dass zu Beginn des Lernprozesses ein Prototypenmodell die Kategorisierungsleistung der VP am besten erklärt, wohingegen zum Ende des Experiments ein Exemplarmodell plausibler erscheint. Gegenüber anderen Clusterverfahren wie beispielsweise dem *Rational Model of Categorization* (RMC; J. Anderson, 1991) oder SUSTAIN, welche ebenfalls hybride Repräsentationen erzeugen

Tabelle 1. Beispielhafte Objekte zur Demonstration einer Clusterreduzierung im URM.

	Objekte			
Dimensionen	0	1	1	0
	1	0	0	1
	1	1	0	0
	1	1	0	0

können, unterscheidet sich das URM darin, dass es nach jedem Informationszuwachs eine Verteilung über alle potentiell möglichen Kategorienstrukturen der aktuell vorliegenden Daten verwaltet, während das RMC und SUSTAIN eine Greedy-Strategie verfolgen und nur die aktuell lokal optimalste Partition behalten.

Im Vergleich zu SUSTAIN oder dem RMC ist das URM deshalb in der Lage die präferierte Anzahl der Cluster sowie deren Charakteristik während des Lernprozesses beliebig neu zu wählen. Hierzu gehört auch die Fähigkeit, die präferierte Anzahl der Cluster nach neuen Beobachtungen zu reduzieren. Dies kann beispielsweise dann angebracht sein, wenn ein stereotypisches Objekt einer Kategorie nach dem Erlernen einiger Spezialfälle präsentiert wird und eine Zusammenführung von zuvor generierten Subklassen nahelegt.

Nehmen wir beispielsweise vier Objekte, welche über Merkmalsausprägungen auf vier Dimensionen, entsprechend der Tabelle 1, beschrieben sind. Die vier Objekte kann man sich als Tiere zweier Fantasiekategorien, nennen wir sie ‚Mug' und ‚Zuf', vorstellen. Die Dimensionen könnten dabei die Charakteristiken Größe (0 = klein, 1 = groß), Farbe (0 = schwarz, 1 = weiß), Form (0 = rund, 1 = kantig) und Schwanz (0 = ohne, 1 = mit) repräsentieren. Das erste Objekt wäre dann ein kleines weißes kantiges Tier mit Schwanz. Präsentiert man nun diese vier Objekte einem entsprechend parametrisierten URM fünfmal, so präferiert das URM eine Gruppierung mit jeweils übereinstimmenden Objekten in einem eigenen Cluster. Die viermal fünf Objekte sind somit in genau vier Cluster organisiert, jeweils einen pro Objekttyp. Präsentiert man anschließend zehnmal die zwei stereotypischen Objekte der Subklassen (jeweils alle Merkmalsausprägungen auf 0 oder auf 1), so reduziert sich die präferierte Anzahl der Cluster im URM von vier auf zwei. Der Vorgang ist in Abbildung 4 dargestellt. SUSTAIN kann ein derartiges Verhalten nicht reproduzieren, da bereits generierte Cluster nicht wieder gelöscht werden können.

Ebenso kann SUSTAIN keine Clusterrestrukturierung vornehmen wie sie anhand der in Tabelle 2 dargestellten Stimuli beim URM demonstriert werden kann. In diesem Beispiel (siehe Abbildung 5) werden einem entsprechend parametrisierten URM die ersten zwei Stimuli jeweils fünfmal und der dritte Stimulus einmal präsentiert. Das URM präferiert daraufhin zwei Cluster, jeweils einen für

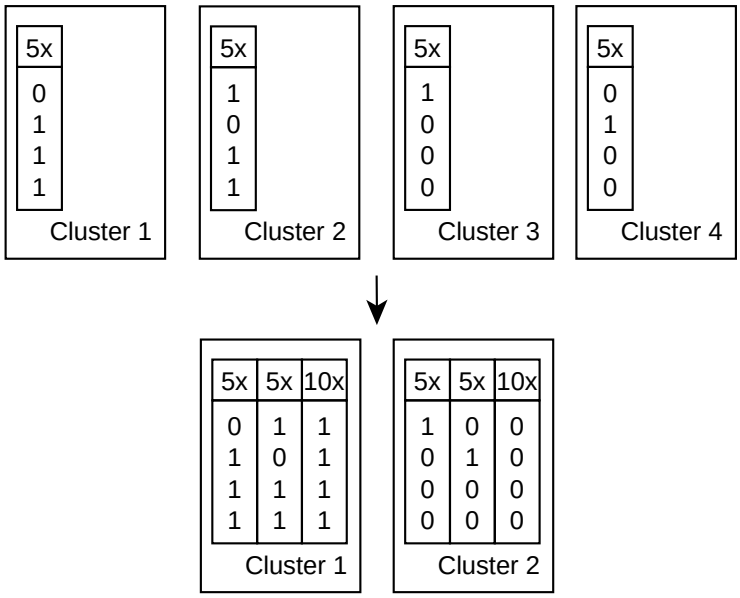


Abbildung 4. Beispiel einer Clusterreduzierung im URM.

jeden der ersten beiden Stimulustypen. Dabei wählt es eine Partition, in welcher der dritte Stimulus zusammen mit den fünf Kopien des ersten Stimulus in einen gemeinsamen Cluster liegt. Wird nun anschließend der vierte Stimulus ebenfalls fünfmal präsentiert, so werden diese fünf Kopien in das zweite Cluster, in welchem sich bereits die fünf Kopien des zweiten Stimulus befinden, positioniert. Aufgrund der Ähnlichkeit zum vierten Stimulus, verschiebt das URM zusätzlich den dritten Stimulus, welcher sich bisher noch in Cluster 1 befand, in den zweiten Cluster. Zu einer solchen Umstrukturierung ist SUSTAIN nicht in der Lage. Die Reihenfolge der Stimuluspräsentation ist deshalb entscheidend bei SUSTAIN, wobei die gleiche Anzahl an Stimuli in unterschiedlichen Präsentationsreihenfolgen zu einer voneinander abweichenden Anzahl und Charakteristik der generierten Cluster führen kann. Das Verhalten vom URM ist hingegen nicht reihenfolgensensitiv.

Tabelle 2. Beispielhafte Objekte zur Demonstration einer Clusterrestrukturierung im URM.

	Objekte			
Dimensionen	1	0	0	0
	1	0	0	0
	1	0	1	0
	1	0	1	1
	1	0	1	1
	1	1	1	1

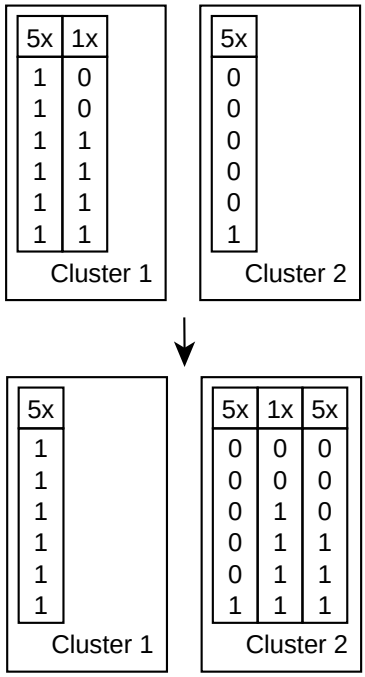


Abbildung 5. Beispiel einer Clusterrestrukturierung im URM.

Jedes beliebige Objekt könnte deshalb genauso gut das letzte Objekt in der Präsentationsreihenfolge sein, um zur gleichen Clusterpräferierung zu gelangen.

Neben dem URM existiert noch ein weiteres Kategorisierungsmodell, welches ebenfalls die oben genannten Fähigkeiten aufweist. Es handelt sich um das als *More Rational Model of Categorization* (MRMC) bezeichnete Verfahren von Sanborn, Griffiths und Navarro (2006), welches eine Verbesserung des RMC von Anderson (1991) darstellt. So wird bei Sanborn et al. (2006) anstelle des bisherigen inkrementellen lokalen Maximum A-Posteriori Sampling Verfahren, unter anderem Gibbs Sampling für die Inferenz der Clusterzugehörigkeit eines neuen Objekts vorgeschlagen. Das MRMC ist damit ebenso wie das URM fähig zur Clusterreduktion und -restrukturierung (McDonnell & Gureckis, 2011). Dennoch besteht ein Unterschied zwischen beiden Verfahren, welcher im Cluster-Sharing des RMC und somit auch im MRMC begründet liegt. Anders als im URM wird im MRMC nur ein DP für alle Kategorien verwendet, wobei die Kategorie der Stimuli als Ausprägung in einer zusätzlichen Dimension codiert ist (Sanborn et al., 2006; Sanborn, Griffiths, & Navarro, 2010). Dabei können Effekte nach dem Prinzip von Abbildung 4 auftreten, welche in diesem Zusammenhang nicht gewollt sein könnten. Beispielsweise könnte die in Abbildung 4 verwendeten Prototypen aus einer Kategorie A stammen und die vier vom jeweiligen Prototyp abweichenden Objekte aus einer Kategorie B. Während in diesem Beispiel das URM durch getrennte DPs für jede Kategorie die ursprüngliche Clusterstruktur der Kategorie B

beibehalten kann, würden im MRMC die Cluster der Kategorie B entsprechend der Abbildung 4 fusionieren. Das URM kann den Kategorien damit mehr Spielraum in der Wahl der Anzahl und Art der Subcluster geben, wohingegen das MRMC den Kategorien immer eine gemeinsame Clusterstruktur aufzwingt (Griffiths et al., 2007).

In manchen Modellierungsituationen kann jedoch Cluster Sharing ausdrücklich erwünscht sein, wobei man den Zwang einer gemeinsamen Clusterstruktur wie im RMC bzw. MRMC weiterhin vermeiden möchte. Das URM bietet deshalb zusätzlich die Möglichkeit einer graduellen Einstellbarkeit des Cluster Sharings. Hierfür ist der Konzentrationsparameter γ ausschlaggebend. Wählt man einen sehr großen Wert für γ, etwa 10^9 - 1, so findet praktisch keine gemeinsame Nutzung von Clustern über Kategorien hinweg statt. Ein sehr niedriges γ hingegen sorgt dafür, dass ausschließlich besetzte Cluster, unabhängig der Kategoriennutzung, als Ziel einer neuen Einsortierung eines Items in Frage kommen. Der Konzentrationsparameter α nimmt hierbei die Rolle eines Verstärkers ein. Ist γ sehr niedrig führt ein hohes α zu einer Tendenz Partitionen mit vielen Clustern zu erhalten. Ist γ sehr hoch, führt hingegen ein hohes α nicht zu einer hohen Neurekrutierung von Clustern, sondern zu einer stärkeren Nutzung bereits vorhandener Cluster, unabhängig davon ob die jeweiligen Cluster von der eigenen Kategorie besetzt sind oder nicht. Da im URM sowohl für γ als auch α Gammaprior definierbar sind, sampelt das URM die entsprechenden Konzentrationsparameter anhand der Beobachtungsdaten und erlaubt somit eine flexible Anpassung an die Situation.

Neben den aufgeführten Stärken des URM fehlt dem Modell jedoch, ebenso wie dem RMC und dem MRMC, ein bei einigen prominenten Modellen implementierter Mechanismus, welcher das Erlernen und die Nutzung dimensionsspezifischer Gewichtungen erlaubt. So sind neben SUSTAIN unter Anderem auch die beiden Exemplar-basierten Modelle, das mechanistische GCM (Nosofsky, 1986) und das konnektionistische A*ttention Learning Covering Map*-Modell (ALCOVE; Kruschke, 1992), dazu in der Lage, die aufgabenspezifische Wichtigkeit von Stimulusdimensionen zu identifizieren und bei der Prädiktion entsprechend zu berücksichtigen. Es wird davon ausgegangen, dass bestimmte Phänomene der menschlichen Kategorisierungsleistung ohne diesen Mechanismus nicht erklärbar sind (McColeman et al., 2014). Inwieweit diese These bestätigt werden kann, wird sich in den späteren Modellierungen zeigen.

2.11 Die Parameteroptimierung

Damit SUSTAIN und das URM ein bestimmtes Verhalten zeigen können, bedarf es eines geeigneten Parametersatzes. Diesen kann man erhalten, indem man entweder nach Kenntnis über die Auswirkung der einzelnen Parameter, die Werte direkt definiert oder man einen Ausgangsparametersatz systematisch variiert bis das gewünschte Verhalten des Modells (in bestmöglicher Weise) vorliegt. Eine systematische Variation kann beispielsweise ein Grid-Search darstellen (Zabinsky, 2003). Hierbei werden alle Parameter innerhalb eines definierten parameterspezifischen Wertebereichs mit einer festgelegten Schrittweite ausprobiert. Dieses Vorgehen kann jedoch, je nach Größe der Wertebereiche, Schrittweiten und Aufwand der Parameterevaluation, sehr zeitintensiv sein. Ein Modell mit drei Parametern und einer Suche im Wertebereich 1 – 100 mit Schrittweite 1 pro Parameter ergäbe dann beispielsweise eine Anzahl von 100^3 = 1.000.000 zu testenden Parametersätzen. Bei einem optimistischen Zeitaufwand von einer Sekunde pro Parameterevaluation wären das mehr als 11 Tage Rechenzeit. Um diesen Zeitaufwand zu reduzieren, besteht die Möglichkeit auf heuristische Suchverfahren zurückzugreifen. Hierzu zählen die sogenannten genetischen Algorithmen (Srinivas & Patnaik, 1994).

Bei dem in dieser Arbeit verwendete genetische Algorithmus handelt es sich um den Algorithmus GA der Global Optimization Toolbox der MATLAB Version R2012a. Er funktioniert folgendermaßen (The Mathworks, 2016): Zunächst werden N_I zufällige Parametersätze (= Individuen) erzeugt, welche zusammen die sogenannte Population bilden. Für jedes dieser Individuen wird anschließend ein Fitnesswert ermittelt, welcher in unserem Fall die Modellierungsgüte des jeweils zugehörigen Parametersatzes darstellt. Ausgehend von diesen Fitnesswerten wird dann aus der aktuellen Population (= aktuelle Generation) eine neue Generation an Individuen erzeugt. Einige der besten Individuen werden dabei direkt an die neue Generation weitergegeben, alle Übrigen werden durch sogenannte Mutation bzw. Crossover eines bzw. zweier Individuen der noch aktuellen Generation erzeugt, welche in diesem Kontext als Eltern des neuen Individuums fungieren. Handelt es sich bei einem neuen Individuum um eine Mutation, so wurden dessen Gene (= Parameterwerte) beispielsweise durch Addition von zufälligen normalverteilten Summanden verändert. Demgegenüber stellt bei einem Crossover jedes Gen eine Vererbung von einem zufällig gewählten Elternteil dar. Ist die neue Generation erzeugt, wird erneut ein Fitnesswert für jedes Individuum bestimmt und der Vorgang beginnt von vorne. Insgesamt werden in der Standardeinstellung 100 Generationen getestet.

2.12 Zugrundeliegende Maße der Modellvergleiche

Love et al. (2004) modellierten in ihrer Studie die oben angesprochenen sieben Experimente mit SUSTAIN. Dabei verwendeten sie ebenfalls einen genetischen Algorithmus, um die Parameter des Modells zu optimieren. Zur Berechnung der Fitnesswerte der Individuen wurde die Summe der Fehlerquadrate (SSE) verwendet, welche aus den Abweichungen der Modellvorhersage von der mittleren menschlichen Kategorisierungsleistung in dem jeweiligen Experiment resultiert. Um das URM und SUSTAIN bezüglich ihrer Modellierungsgüte bei den Experimenten vergleichen zu können, wurde deshalb ein analoges Vorgehen gewählt.

Die SSE stellt ein geeignetes Maß für die Bewertung eines Modells dar, sofern zwei Sachverhalte sichergestellt sind: 1. das Modell prognostiziert ein qualitativ korrektes Ergebnis und 2. die (niedrige) SSE ist kein Resultat der Flexibilität des Modells, das heißt es ist kein Modell, welches zu Overfitting neigt (Wills & Pothos, 2012). Der erste Punkt kann durch simples Inspizieren der Modellprognose bestätigt werden. Um dem zweiten Punkt hingegen zu genügen, bedarf es einer Modellflexibilitätsanalyse. In dieser Arbeit wird deshalb das hierfür vorgestellte gleichnamige Analyseverfahren von Veksler, Myers und Gluck (2015) verwendet. Bei diesem Verfahren wird eine Proportion Φ berechnet, welche das Verhältnis von Anzahl an modellierbaren Verhaltensmustern zu Anzahl an getesteten Parametersätzen darstellt. Je kleiner Φ ist, desto unflexibler ist das Modell und desto weniger neigt das Modell zu Overfitting. Eine niedrige SSE bei kleinem Φ kann dann als Hinweis auf ein gutes Modell für die Daten gesehen werden.

Φ wird berechnet, indem zunächst eine repräsentative Stichprobe S_P von Parametersätzen aus der Menge aller möglichen Parameterkombinationen ermittelt und die Prognosen des Modells unter Verwendung dieser Parametersätze erhoben werden. Anschließend wird die Anzahl der qualitativ unterschiedlichen Prognosen durch die Größe der Stichprobe S_P geteilt. Die konkrete Berechnung von Φ ist nach Veksler et al. (2015) folgendermaßen:

$$\Phi = \frac{N_{row}\left(unique\left(\left\lfloor \boldsymbol{P} \cdot \sqrt[N_{col}(\boldsymbol{P})]{N_{row}(\boldsymbol{P})} \right\rfloor\right)\right)}{N_{row}(\boldsymbol{P})} \tag{54}$$

$\boldsymbol{P}$ bezeichnet die Prognosematrix, deren Zeilenanzahl der Größe der Stichprobe S_P und deren Spaltenanzahl der Anzahl der modellspezifischen Verhaltensprognosen entspricht. Sagt beispielsweise ein Modell Werte auf zwei Verhaltensmaßen vorher, etwa mittlerer Fehleranteil und Anzahl der benötigten Lernblöcke, dann liefert das Modell für jeden Parametersatz zwei Prognosewerte und die Spaltenanzahl von $\boldsymbol{P}$ entspricht 2. Die Funktionen $N_{col}()$ bzw. $N_{row}()$ in

(54) ermitteln diese Spalten- bzw. Zeilenanzahl einer übergebenen Matrix. *unique*() entfernt dann aus dieser alle doppelt vorkommenden Prognosen und liefert die resultierende Matrix zurück.
Um eine möglichst repräsentative Stichprobe des Parameterraums zu erhalten, kann wie im Falle eines Grid-Searches, ein engmaschiges Netz von Parameterkombinationen verwendet werden. Das ist jedoch zum einen, wie bereits erwähnt, sehr rechenintensiv, und zum anderen nicht unbedingt optimal, um möglichst viele unterschiedliche Verhaltensprognosen zu erhalten. Bergstra und Bengio (2012) konnten beispielsweise zeigen, dass ein Random-Search gegenüber dem klassischen Grid-Search eine bessere Möglichkeit darstellt bei neuronalen Netzen einen geeigneten Parametersatz für vorliegende Daten zu finden. Sie erklären diesen Sachverhalt damit, dass häufig nur wenige Parameter eine Rolle spielen und somit bei Grid-Search viele weitestgehend irrelevante Parameterkombinationen getestet werden. Für die Modell-Flexibilitätsanalyse werden daher Stichproben verwendet, deren Parametersätze zufällig generierte Positionen im Parameterraum darstellen.
Während wir bei der Modellflexibilitätsanalyse ein approximiertes Φ erhalten, welches für $N \rightarrow \infty$ gegen das wahre Φ konvergiert und somit ein ausreichend großes *N* eine gute (erste) Einschätzung der Flexibilitätsunterschiede zweier Modelle erlaubt, sind insbesondere bei Modellbewertungen, welche auf der Grundlage weniger Prognosen getroffen werden, statistische Tests notwendig. Dies betrifft zum Beispiel die Frage, welches von zwei Modellen die bessere Prognose für einen experimentellen Befund aufweist. Warum hier kein einfacher Wertevergleich zielführend ist, ob der Grund hierfür Auswirkungen auf die Parameteroptimierung über GA hat und wie das Problem in dieser Arbeit gelöst wird, ist im Folgenden erklärt.

2.13 Statistische Auswertungen

Eine Modellprognose Ω^{θ} von URM oder SUSTAIN für ein spezifisches Parameterset θ stellt selbst eine Zufallsvariable dar. Der Hintergrund ist, dass bei beiden Modellen die Reihenfolge der Trials in jedem Experiment blockweise randomisiert ist und hierdurch je nach Modellierungswiederholung mit demselben Parametersatz unterschiedliche SSEs vorliegen. Weiterhin wird jede trialspezifische Vorhersage mit dem URM nur anhand einer Approximation der Posterior des HDPMM über 30 Samples bestimmt, was somit eine weitere Ursache für die Varianz von Ω^{θ}_{URM} darstellt.
Aufgrund der Erfahrung durch unzählige Wiederholungen der Parameteroptimierungsprozedur über GA für jedes der Experimente konnte festgestellt werden, dass trotz Varianz von Ω^{θ} eine SSE-Stabilität bei den ermittelten Parametersets gegeben ist, wobei allerdings die gefundenen besten

Parametersets zwischen den Wiederholungen der Optimierungsprozedur variieren. Es wird deshalb vermutet, dass viele Parametersets vergleichbar niedrige SSEs produzieren und die Überprüfung von 8000 Parametersets im Verlauf einer Fittingprozedur ausreichen, um eines dieser besten Parametersets zu ermitteln. Weiterhin wird aufgrund der Tatsache, dass für jede Prozedurwiederholung die initiale Population der Parametersets von GA zufällig über den gesamten relevanten Parameterraum verteilt ist und der Beobachtung, dass Wiederholungen der Optimierungsprozedur nicht zur Identifikation von besseren Parametersets führen, angenommen, dass kein globales Optimum übersehen wurde, welches zu deutlich besseren SSEs führt.

Wir wollen im Weiteren annehmen, dass der Erwartungswertunterschied und die Varianzen der SSE-Verteilungen $SSE\left(\Omega_{URM}^{\theta}\right)$ und $SSE\left(\Omega_{SUS}^{\theta}\right)$ der durch GA ermittelten Parametersets für das URM und SUSTAIN repräsentativ sind für den Unterschied der Erwartungswerte und die Varianzen der Parametersets beider Modelle, welche global betrachtet den wahren niedrigsten Erwartungswert aufweisen. Unter diesen Voraussetzungen können wir nun auf Basis der SSE-Verteilungen von Ω_{URM}^{θ} und Ω_{SUS}^{θ} eine Aussage bezüglich der Unterschiede zwischen den Modellen hinsichtlich der mittleren Prognosegüte für die jeweiligen Experimentaldaten treffen.

Hierzu wollen wir zunächst bewerten, wie wahrscheinlich ein gefundener Mittelwertsunterschied zweier Stichproben aus $SSE\left(\Omega_{URM}^{\theta}\right)$ und $SSE\left(\Omega_{SUS}^{\theta}\right)$ ist, wenn wir von einem wahren Unterschied von 0 ausgehen. Wenn beide SSE-Verteilungen normalverteilt wären und wir gleiche Varianzen annehmen könnten, wäre der *t*-Test die Methode der Wahl. Wenn die Varianzen hingegen klar voneinander abzuweichen scheinen, ist der Welch-Test vorzuziehen. Hierbei handelt es sich um einen approximativen Test, welcher eine praxisbezogene Lösung des Behrens-Fisher-Problems darstellt. Bei letzterem handelt es sich um das Problem einen exakten Signifikanztest für zwei vorliegenden Stichproben aus beiden Verteilungen mit garantiertem α Level zu bestimmen, wenn die Populationen von Interesse zwar normalverteilt sind, jedoch ihre Varianzen unbekannt und als nicht unbedingt identisch angesehen werden können (Dudewicz, Ma, Mai, & Su, 2007). Der russische Statistiker Juri Linnik konnte zeigen, dass ein solcher Test nicht existiert (Bather, 1996, p. 337), weshalb der Welch-Test nur ein approximativer Test darstellt. Es existiert eine Erweiterung dieses Problems, welches entsprechend als Generalisiertes Behrens-Fisher-Problem oder auch als Nonparametrisches Behrens-Fisher-Problem bezeichnet wird. Es liegt vor, wenn neben der Varianzenungleichheit zusätzlich keine Normalverteilungen angenommen werden können. Die ausgiebige Inspektion von Stichproben aus $SSE\left(\Omega_{URM}^{\theta}\right)$ und $SSE\left(\Omega_{SUS}^{\theta}\right)$ legen den Schluss nahe, dass dieses Problem hier vorliegt.

Ein oft in einer solchen Situation angewendeter *u*-Test ist streng genommen nicht korrekt, da für den *u*-Test ebenso wie für den *t*-Test Varianzengleichheit eine Anwendungsvoraussetzung darstellt (Zimmerman, 1987). Es besteht jedoch die theoretische Möglichkeit ein 2-Phasen Verfahren anzuwenden, ähnlich zu den vorgestellten Methoden für das klassische Behrens-Fisher-Problem in Dudewicz et al. (2007). Letztere Autoren präsentierten und begutachteten drei mittlerweile entwickelte Methoden, welche das Behrens-Fisher Problem über einen Umweg exakt lösen. Hierzu bedarf es zunächst zweier Stichproben, über welche eine Anzahl an weiteren notwendigen Ziehungen aus beiden Verteilungen bestimmt wird, damit die H_0 zugunsten einer H_1 mit angenommener wahrer Mittelwertsdifferenz von Δ bei einem festgelegten β mit garantiertem α Level abgelehnt werden kann (Dudewicz et al., 2007).

Der Grund warum wir theoretisch ein ähnliches Prinzip trotz Unkenntnis der Verteilungen und ihrer Varianzen auch bei $SSE\left(\Omega_{URM}^{\theta}\right)$ und $SSE\left(\Omega_{SUS}^{\theta}\right)$ anwenden können, sind zwei besondere Umstände: Realisierungen aus den SSE-Verteilungen stammen aus einem endlichen Intervall, für welche eine obere und untere Schranke angegeben werden kann und wir können jederzeit die aktuellen Stichproben durch beliebig viele neue Samples aus beiden Verteilungen vergrößern. Eine obere und untere Schranke der Intervallgrenzen lässt sich angeben, weil wir wissen, dass das beste mögliche Ergebnis eine SSE von 0 ist und das schlechteste eine SSE mit systematisch falsch beantworteten Trials im jeweiligen Experiment. Mit diesem Wissen können wir die maximal mögliche Varianz berechnen, welche eine Verteilung auf dem Intervall [0, SSE_{Max}] haben kann. Sie beträgt $2^{-2} \cdot SSE_{max}^{2}$ und liegt vor, wenn jeweils 50% Wahrscheinlichkeit den Werten 0 und SSE_{Max} zugeteilt sind, während alle anderen Werte Wahrscheinlichkeit 0 aufweisen. Wir können dann diese schlechtest mögliche Varianz für unsere unbekannten Verteilungen annehmen und berechnen, wie groß unsere zweiten Stichproben sein müssen, damit eine Mittelwertsdifferenz Δ_1 der initialen Stichproben signifikant ist, sofern die H_0 gilt. Haben wir nun jeweils diese Stichprobengrößen N durch Resampling erreicht, während weiterhin mindestens Mittelwertdsdifferenz Δ_1 vorliegt und können wir aufgrund des zentralen Grenzwertsatzes die Verteilung der Mittelwertsdifferenz durch eine Normalverteilung mit Erwartungswert 0 und Standardabweichung $\sqrt{2 \cdot 2^{-2} \cdot SSE_{max}^{2} \cdot N^{-1}} = \sqrt{(2N)^{-1}}SSE_{max}$ ausreichend approximieren, dann sind die Erwartungswerte der unbekannten Verteilungen unabhängig von ihren tatsächlichen Varianzen signifikant unterschiedlich mit ungefährem Typ I Fehler $\leq \alpha$. Wie sicher Letzteres ist hängt dabei von der Approximationsgüte der Verteilung der Mittelwertsdifferenz durch die Normalverteilung ab.

Theoretisch lässt sich also über diese Herangehensweise ein Urteil mit beliebiger Sicherheit bezüglich des wahren Typ I Fehler treffen. Praktisch jedoch ist das Verfahren aufgrund der hohen

Anzahl an benötigten Samples und den zeitlichen Kosten Samples aus $SSE(\Omega^{\theta}_{URM})$ und $SSE(\Omega^{\theta}_{SUS})$ zu ziehen, in seiner Brauchbarkeit begrenzt. Für Experiment 1 wurde beispielsweise ein Worst-Case SSE Intervall von [0, 80.1444] ermittelt mit einer maximal möglichen Varianz von 1605.78 für beide Verteilungen. Die Mittelwertsdifferenz von zwei ersten jeweils 100 Samples umfassenden Stichproben aus $SSE(\Omega^{\theta}_{URM})$ und $SSE(\Omega^{\theta}_{SUS})$ lag bei 0.117, was in einer benötigten Stichprobengröße von weit über 600000 resultierte um selbst bei der maximalen Varianz einen signifikanten Unterschied zu finden. Selbst wenn man den liberaleren Fall annimmt, dass die Modelle im Worst-Case durchgehend nur für jeden zweiten Trial eine richtige Antwort vorhersagen und somit die obere Grenze des SSE Intervalls auf 17.1834 sinkt, liegt die benötigte Stichprobengröße bei gleicher Mittelwertsdifferenz noch bei über 28000. Da die Generierung von 8000 Samples in Experiment 1 mit der verfügbaren Hardware bereits etwa 2 – 4 Tage in Anspruch nahm, wurde deshalb anstelle des vorgestellten Verfahrens eine alternative Methode gewählt. Es handelt sich dabei um den nonparametrischen Rang-Test von Brunner und Munzel (2000), welcher speziell für das generalisierte Behrens-Fisher-Problem konzipiert wurde und im Software Paket *nparcomp* für R enthalten ist (Konietschke, Placzek, Schaarschmidt, & Hothorn, 2015).

Das Verfahren funktioniert im Wesentlichen so, dass zunächst der relative Treatment Effekt *p* von zwei Datensätzen geschätzt wird, wobei *p* als

$$p = P(X_{11} < X_{21}) + \frac{1}{2}P(X_{11} = X_{21}) \tag{55}$$

definiert ist. Wenn $p = 0.5$ ist, dann tendieren Ziehungen X_{11} und X_{21} aus der Verteilung des ersten und zweiten Datensatzes zum gleichen Wert. Ist hingegen $p < 0.5$ bzw. $p > 0.5$, dann tendieren Ziehungen aus dem ersten Datensatz zu größeren bzw. kleineren Werten als Ziehungen aus dem zweiten Datensatz. Brunner und Munzel (2000) konnten einen erwartungstreuen und konsistenten Schätzer $\hat{p}$ für *p* bestimmen, welcher unter der H_0 asymptotisch verteilt ist als Normal(0.5, $\sigma_N \cdot (\sqrt{N})^{-1}$), wobei σ_N durch einen konsistenter Schätzer $\hat{\sigma}_N$ für σ_N ersetzt wird. Ein klassischer *p*-Wert zur Entscheidung über die Signifikanz ließe sich dann über das Quantil der zugehörigen Statistik W_N^{BF} mit

$$W_N^{BF} = \frac{\sqrt{N}(\hat{p} - \frac{1}{2})}{\hat{\sigma}_N} \tag{56}$$

in der Standardnormalverteilung ablesen. Um die Methode jedoch auch für kleine Stichproben möglichst akkurat zu halten, wird statt der Standardnormalverteilung eine *t*-Verteilung als Approximation der Verteilung von W_N^{BF} verwendet, wobei die Anzahl der Freiheitsgrade über die

Welch-Satterthwaite-Equation geschätzt wird (Brunner & Munzel, 2000; Welch, 1947). Die Approximationsqualität in einer Simulationsstudie bezeichnen Brunner und Munzel (2000) als vergleichbar zum Welch-Test in einer ähnlichen parametrischen Situation.

Um nun die beiden Modelle URM und SUSTAIN hinsichtlich ihrer Modellierungsgüte zu vergleichen, wurden für jedes Experiment 200 Samples aus $SSE\left(\Omega^{\theta}_{URM}\right)$ und $SSE\left(\Omega^{\theta}_{SUS}\right)$ generiert. Danach erfolgte die jeweilige Berechnung von $\hat{p}$ und dem zugehörigen p-Wert, sowie eines 95% Konfidenzintervall für $\hat{p}$. Für jedes Experiment werden Erstere zusammen mit den Mittelwerten und Varianzen der Stichproben berichtet und das zugehörige Konfidenzintervall einschließlich Density-Plots der Stichprobenverteilungen grafisch präsentiert. Um eine Prüfung auf Übereinstimmung des qualitativen Musters in den Experimentaldaten und der Modellprognosen zu ermöglichen, wird zusätzlich beispielhaft die Prognoserealisierung aus Ω^{θ} grafisch dargestellt oder im Text beschrieben, deren SSE von den 200 Samples die minimale quadratische Distanz zum SSE-Mittelwert aufweist. Sollten mehrere SSEs aus der Stichprobe die gleiche minimale Distanz zum Mittelwert aufweisen, wird die Realisierung gewählt, welche die minimale SSE unter diesen Alternativen aufweist.

Diese mittlere Prognoserealisierung reicht nach eigener Erfahrung aus um zu prüfen, ob das jeweilige Modell korrekte qualitative Vorhersagen trifft. So ist beispielsweise im Verlauf der Modellierung niemals beobachtet worden, dass bei einer Simulationswiederholung mit denselben Parametern trotz variierender SSEs ein qualitativ anderes Ergebnis entstand. Da dies jedoch theoretisch möglich ist, wird vermutet, dass ein solches Ereignis nur extrem selten auftritt. Sollte bei der Auswahl der beispielhaften Prognoserealisierung diese Situation vorliegen, so wird das in der Ergebnisdarstellung angesprochen.

3 Experimente

3.1 Genereller Ablauf

Alle Modellierungen der Experimente sowie die Flexibilitätsanalysen wurden in MATLAB R2012a unter OpenSuse Linux 13.2 durchgeführt. Prognosen des URM wurden mit dem Softwarepaket npBayes2.1G von Griffiths et al. (2007) und die von SUSTAIN mit der von Gureckis (2014) veröffentlichten Python-Version ermittelt. Für den genetischen Algorithmus zur Optimierung der Parameter des URM bzw. SUSTAIN wurde eine Populationsgröße von N_I = 80 pro Generation

Tabelle 3. Experimentspezifische Mengen zulässiger Parametersätze für SUSTAIN.

Experiment 1	$\{10^{-2}(r,\beta,m,\eta) \mid r,\beta,m \in \{1,\ldots,10^4\}, \eta \in \{1,\ldots,10^2\}\}$
Experimente 2 – 4	$\{10^{-2}(r,\beta,m,\eta) \mid r,\beta \in \{1,\ldots,2400\}, m \in \{1,\ldots,10^4\}, \eta \in \{1,\ldots,10^2\}\}$
Experimente 5 - 7	$\{10^{-2}(r,\beta,m,\eta) \mid r,\beta,m \in \{1,\ldots,2400\}, \eta \in \{1,\ldots,10^2\}\}$

festgelegt. Es wurden bei jeder Parameteroptimierung 100 Generationen getestet. Bei einer Prognose von SUSTAIN handelt es sich um den Mittelwert aus 30 Simulationswiederholungen mit dem über GA ermittelten Parametersatz. Für jede Prognose mit dem URM wurden ein Burn-In von 200, ein Sampling von 30 und ein Thinning von 10 verwendet. Die Werte wurden als Kompromiss zwischen Replizierbarkeit und Zeitaufwand pro Prognose gewählt. 50 Simulationswiederholungen für Experiment 1 mit einem zufällig gewählten Parametersatz wiesen mit diesen Einstellungen eine Standardabweichung der SSEs von 0.5% des Stichprobenmittelwerts bei einer Ausführungszeit von 70 Sekunden pro Simulation auf einer i7-3630qm CPU auf. Für den Metropolis-Hastings Sampler zur Ziehung der $\beta_{0,}^{d}\beta_{1}^{d}$ wurden pro Sampling fünf Iterationen durchgeführt. Die Log-Normalverteilung, welche als Proposal-Distribution verwendet wurde, hatte ein Scaling von 0.5.

Für das URM wurde zunächst getestet, welche Priorvariante (z.B. Betaparameter mit oder ohne Gammaprior) den besten Fit in Experiment 1 ermöglicht und diese Variante, sofern keine Modifizierungen am Modell notwendig waren, für die weiteren Experimente beibehalten. Die getesteten Varianten sind in Tabelle 4 dargestellt und werden im Abschnitt zum Experiment 1 näher erläutert.

Bei den Modell-Flexibilitätsanalysen wurden für jedes Experiment 128000 zufällige Parametersätze aus einer Menge zulässiger Parametersätze des jeweiligen Modells gezogen. Sofern keiner der für das jeweilige Experiment benötigten Parameter $p_{1,\ldots,n}$ eine Lernrate darstellte, wurde für das URM $\{10^{-2}(p_{1,\ldots}, p_n) \mid p_1,\ldots,p_n \in \{1,\ldots,10^5\}\}$ als Menge der zulässigen Parametersätze gewählt. Stellte hingegen ein Parameter p_i eine Lernrate dar, so wurde stattdessen die Menge $\{10^{-2}(p_{1,\ldots}, p_i \cdot 10^{-1},\ldots,p_n) \mid p_1,\ldots,p_n \in \{1,\ldots,10^5\}\}$ verwendet.

Bei SUSTAIN wurden die zulässigen Parametersätze je nach Experiment angepasst, da die Potenzierungen in (45) und (46) sehr schnell zu Laufzeitfehlern aufgrund zu großer Zahlen führten. Tabelle 3 zeigt die entsprechenden Mengen.

Es folgen nun die sieben Experimente, deren Befunde mit dem URM und SUSTAIN prognostiziert werden sollen. Zunächst wird für jedes Experiment der thematische Hintergrund erläutert, der Ablauf des Experiments erklärt und anschließend die Ergebnisse in der jeweiligen Studie benannt.

Anschließend werden die Modellierungsergebnisse mit dem URM und SUSTAIN aufgeführt und miteinander verglichen.

3.2 Experiment 1: Lernschwierigkeit von sechs Kategorisierungsregeln

Shepard et al. (1961) untersuchten in ihrer einflussreichen Studie determinierende Faktoren des Lernprozesses bei Identifikations- und Klassifikationsaufgaben. Erstere bezeichnen Aufgaben, bei welchen einer Reihe von Items individuelle Namen zugeordnet sind und diese Namen nach Präsentation der jeweiligen Items wiedergegeben werden sollen. Sind in Aufgaben für mehrere Items hingegen identische Namen zugeordnet, sodass die Anzahl der Namen kleiner ist als die Anzahl der Items, dann handelt es sich um Klassifikationsaufgaben. Im Zentrum des Interesses stand die Frage nach der Ursache der typischerweise geringeren Schwierigkeit von Klassifikations- gegenüber Identifikationsaufgaben. Shepard et al. (1961) vermuteten, dass entweder die Möglichkeit zur Fokussierung auf wenige gemeinsame bzw. unterscheidende Eigenschaften der Items beim Klassifikationslernen einen Verarbeitungsvorteil liefert oder Identifikationslernen durch die größere Anzahl an Antwortmöglichkeiten, welche ebenfalls erinnert werden müssen, beeinträchtigt ist. Ein direkter Vergleich der Aufgabentypen wurde dabei durch die Tatsache erschwert, dass beide Aufgabentypen unterschiedliche sogenannte Chance-Levels aufweisen. Liegen beispielsweise acht Items vor, deren Zuordnungen zu einer von zwei Kategorien bei einer Klassifikationsaufgabe zu erlernen sind, ist die Wahrscheinlichkeit ein Item nur durch Raten richtig einzuordnen 50%. Werden dagegen dieselben acht Items in einer Identifikationsaufgabe verwendet, bei welcher jedes Item in eine eigene Kategorie gehört, reduziert sich die Wahrscheinlichkeit richtig zu raten auf 12.5%. Shepard et al. (1961) präsentierten daher eine Möglichkeit die zur Lösung einer Klassifikationsaufgabe zwingend zu betrachtende Anzahl an Eigenschaften eines Items, bis auf das Level einer Identifikationsaufgabe anheben zu können und gleichzeitig sowohl das Chance-Level als auch die Anzahl der Antwortmöglichkeiten konstant zu halten. Die Grundidee bestand darin durch Variation der Klassenzugehörigkeit einer definierten Anzahl an Items die Schwierigkeit systematisch zu beeinflussen. Shepard et al. (1961) verwendeten dazu acht Items, welche auf drei Dimensionen jeweils eine von zwei mögliche Ausprägungen besaßen. Immer vier dieser acht Items gehörten zur Kategorie A, die übrigen vier Items zur Kategorie B. Unter dieser Voraussetzung existierten $\frac{8!}{(4!)^2}$ Möglichkeiten vier aus acht Items für eine Kategorie A zu bestimmen. Shepard et al. (1961) stellten dabei fest, dass sich die 70 Varianten in sechs Typen mit jeweils bezüglich der

Itemeigenschaften gleichen Zuordnungsregeln gruppieren ließen. Diese sechs Basistypen bzw. -regeln unterschieden sich hinsichtlich der zwingend zu betrachtenden Anzahl an Itemeigenschaften zur Lösung der Klassifikation und Shepard et al. (1961) zeigten mit ihnen, dass sie sich auch wie vermutet in ihrer Aufgabenschwierigkeit unterschieden.

Die Arbeit von Shepard et al. (1961) hatte großen Einfluss auf die Kategorisierungsforschung und deren Befunde zur Aufgabenschwierigkeit der sechs Regeltypen wurde zum wichtigsten Benchmark für formale Modelle der menschlichen Kategorisierung (Kurtz et al., 2013). Dieser Status wurde nochmals untermauert als Nosofsky et al. (1994) eine Replikation und Erweiterung der ursprünglichen Befunde veröffentlichten. Die Autoren wollten damit den über die Jahre zunehmend komplexer werdenden Kategorisierungsmodellen eine reichhaltigere Datenbasis zur Verfügung stellen und anstelle von Endprodukten der Kategorisierung eine Modellierung des gesamten Kategorisierungsprozesses ermöglichen. Ihre Arbeit resultierte in der Erhebung von charakteristischen Lernkurven der sechs Regeltypen, welche in Abbildung 6 dargestellt sind. Wie in Shepard et al. (1961) sind bei Nosofsky et al. (1994) acht Items zwei Kategorien zuzuordnen, welche sich auf drei Dimensionen bei zwei möglichen Ausprägung pro Dimension unterscheiden. Die Autoren verwendeten solche wie in Abbildung 6 B dargestellten Stimuli. Insgesamt waren pro Kategorisierungsregel maximal 400 Kategorisierungen (= Trials) durchzuführen, sofern eine VP nicht im Laufe des Experiments 32 aufeinanderfolgende Trials ohne Fehler absolvierte. In diesem Fall wurde das Experiment erfolgreich vorzeitig beendet.

Die Reihenfolge der Stimuli wurde blockweise randomisiert. In den ersten beiden Blöcken mussten

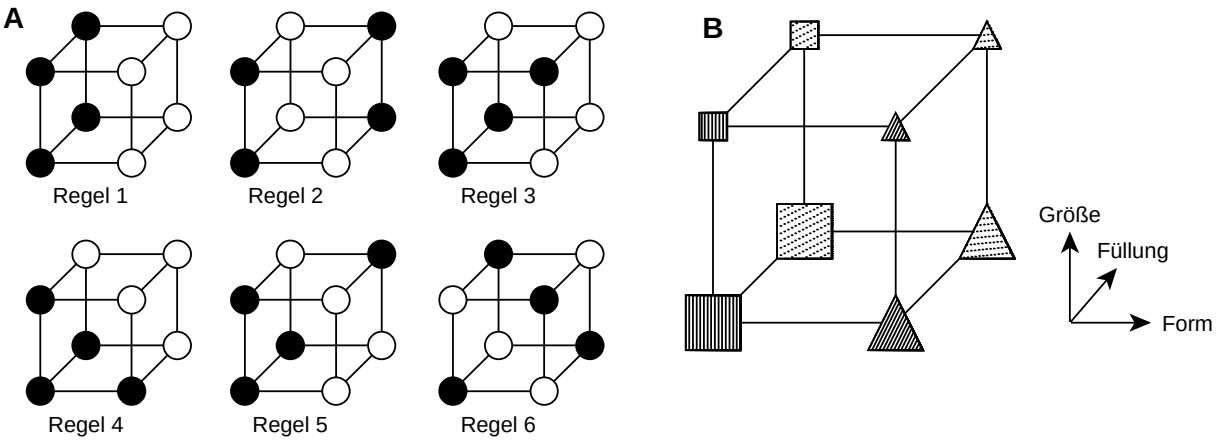


Abbildung 6. Kategorisierungregeln und Stimuli aus Nosofsky et al. (1994). Die ursprüngliche Version dieser Grafik ist aus Glassen und Nitsch (2015), hat hier jedoch leicht angepasste Stimuli in B. (A) Struktur der sechs Kategorisierungsregeln. Gleichfarbige Knoten symbolisieren Items derselben Kategorie. (B) Schematische Darstellung der verwendeten Items.

nacheinander acht Stimuli kategorisiert werden, wobei jeder Stimulus aus Abbildung 6 B genau einmal präsentiert wurde. Anschließend folgten Blöcke mit jeweils 16 Stimuli und zwei Präsentationen pro Stimulus. Nach jeder Einordnung wurde ein Feedback gegeben, ob der Stimulus korrekt kategorisiert wurde.

Insgesamt wurden 120 Versuchspersonen (VP) getestet, wobei jede VP zwei Kategorisierungsregeln lösen musste. Das Design resultierte demnach in 40 VP pro Kategorisierungsregel. Die für die Regeln relevanten Dimensionen wurden für jede VP randomisiert. Dies entspricht einer zufälligen Rotation des in Abbildung 6 B dargestellten Würfels, wobei die Farben der Knoten der in A relevanten Regel die jeweilige Kategorienzugehörigkeit der Stimuli des rotierten Würfels symbolisieren.

Für jeden Block und jeder Kategorisierungsregel wurde abschließend der durchschnittliche Fehleranteil der Kategorisierungen der 40 VP ermittelt, wobei die ersten zwei Blöcke à acht Stimuli als ein zusammenhängender erster Block à 16 Stimuli ausgewertet wurden. Die zugehörigen Lernkurven über die Blöcke 1 bis 16 sind in Abbildung 7 oben dargestellt.

3.2.1 Modellierungsergebnisse für Experiment 1

Um mit dem URM die mittleren Fehleranteile aus Nosofskys et al. (1994) Experiment mit einem möglichst guten Fit zu modellieren, wurden zunächst exploriert welche der Parameter $\beta_{0,}^{d}\beta_{1}^{d}$, α und γ des URM direkt bzw. über einen Gammaprior zu spezifizieren sind. Die durchgeführten Modellierungen weichen dabei von denen in Glassen und Nitsch (2015) ab, welche die Ergebnisse von Nosofskys et al. (1994) Experiment so interpretierten, dass die unter Block 1 und 2 aufgeführten Fehleranteile ihrer Ergebnistabelle, tatsächlich denen der realen Blöcke entsprachen. Wie weiter oben erklärt, wurden die ersten beiden kurzen Blöcke jedoch als ein erster großer Block zusammengefasst. Gureckis (2014) Modellierung mit SUSTAIN folgt ebenfalls dieser Interpretation.

Hiervon ausgehend wurden die Modellierungsvarianten aus Tabelle 4 getestet, wobei sich die Variante 4 hinsichtlich des Mittelwerts der 200 Samples umfassenden Stichprobe aus $SSE\left(\Omega_{URM}^{\theta}\right)$ mit $\overline{SSE}_4$ = 0.154 gegenüber ihren Alternativen $\overline{SSE}_1$ = 0.208, $\overline{SSE}_2$ = 0.157, $\overline{SSE}_3$ = 0.188 behaupten konnte. Die Lernkurven der mittleren Prognoserealisierung der Variante 4 über die Blöcke 1 bis 16 sind in Abbildung 7 unten dargestellt.

Love et al. (2004) modellierten mit SUSTAIN ebenfalls Nosofskys et al. (1994) Lernkurven, verzichteten jedoch auf eine Angabe der SSE. Es wurde deshalb zur Berechnung der SSE die

Tabelle 4. Getestete Priorvarianten in Experiment 1.

Variante	Prior		
1	α	~	Gamma($\gamma_{0,}^{\alpha}\gamma_{1}^{\alpha}$)
	γ	=	α
	$\beta_{0,}^{d}\beta_{1}^{d}$	~	Gamma($\gamma_{0,}^{\beta}\gamma_{1}^{\beta}$)
2	α	~	Gamma($\gamma_{0,}^{\alpha}\gamma_{1}^{\alpha}$)
	γ	=	α
	$\beta_{0,}^{d}\beta_{1}^{d}$	=	β
3	α	~	Gamma($\gamma_{0,}^{\alpha}\gamma_{1}^{\alpha}$)
	γ	~	Gamma($\gamma_{0,}^{\gamma}\gamma_{1}^{\gamma}$)
	$\beta_{0,}^{d}\beta_{1}^{d}$	~	Gamma($\gamma_{0,}^{\beta}\gamma_{1}^{\beta}$)
4	α	~	Gamma($\gamma_{0,}^{\alpha}\gamma_{1}^{\alpha}$)
	γ	~	Gamma($\gamma_{0,}^{\gamma}\gamma_{1}^{\gamma}$)
	$\beta_{0,}^{d}\beta_{1}^{d}$	=	β

Modellierung mit der Python Version von Gureckis (2014) und den bei Love et al. (2004) angegebenen Parametern wiederholt. Sie liegt mit einem Mittelwert von 0.080 der 200 Samples aus $SSE\left(\Omega_{SUS}^{\theta}\right)$ deutlich unter der SSE des URM. Die zugehörigen prognostizierten Lernkurven der mittleren Prognoserealisierung sind in Abbildung 8 unten dargestellt.

Wie zu sehen ist, kann SUSTAIN die Lernkurven von Nosofsky et al. (1994) deutlich besser vorhersagen als das URM, was die Frage aufwirft, welche zentralen Unterschiede zwischen den Modellen für dieses Ergebnis eine Rolle spielen. Einer dieser potentiell wichtigen Unterschiede ist ein Mechanismus für dimensionsbezogene selektive Aufmerksamkeit in SUSTAIN, welche typischerweise bei regelbasierter Kategorisierung eine Rolle spielt (Ashby, 2013; Best, Yim, & Sloutsky, 2013; Folstein, Palmeri, & Gauthier, 2013) und von einigen Forschern als ursächlich für die relative Leichtigkeit der Kategorisierungsregel 2 gegenüber den Regeln 3,4 und 5 gesehen wird (Kurtz et al., 2013). So kann zwar das URM die schwierig zu modellierende geringere Schwierigkeit der Kategorisierungsregel 2 gegenüber der Regel 4 korrekt vorhersagen und sich damit beispielsweise vom RMC absetzen (J. Anderson, 1991; Kurtz et al., 2013), scheitert jedoch gleichzeitig an der Vorhersage einer leichteren Regel 2 verglichen mit der Regel 3. So liegt der durchschnittliche Fehleranteil in den bereits gezogenen 200 Samples aus Ω_{URM}^{θ} bei 0.092 mit einer Varianz von 1.34e-06 für die Kategorisierungsregel 2, während der mittlere Fehleranteil für Regel 3 mit 0.09 und Varianz 8.19e-07 leicht geringer ausfällt. Eine Schätzung des relativen Treatment

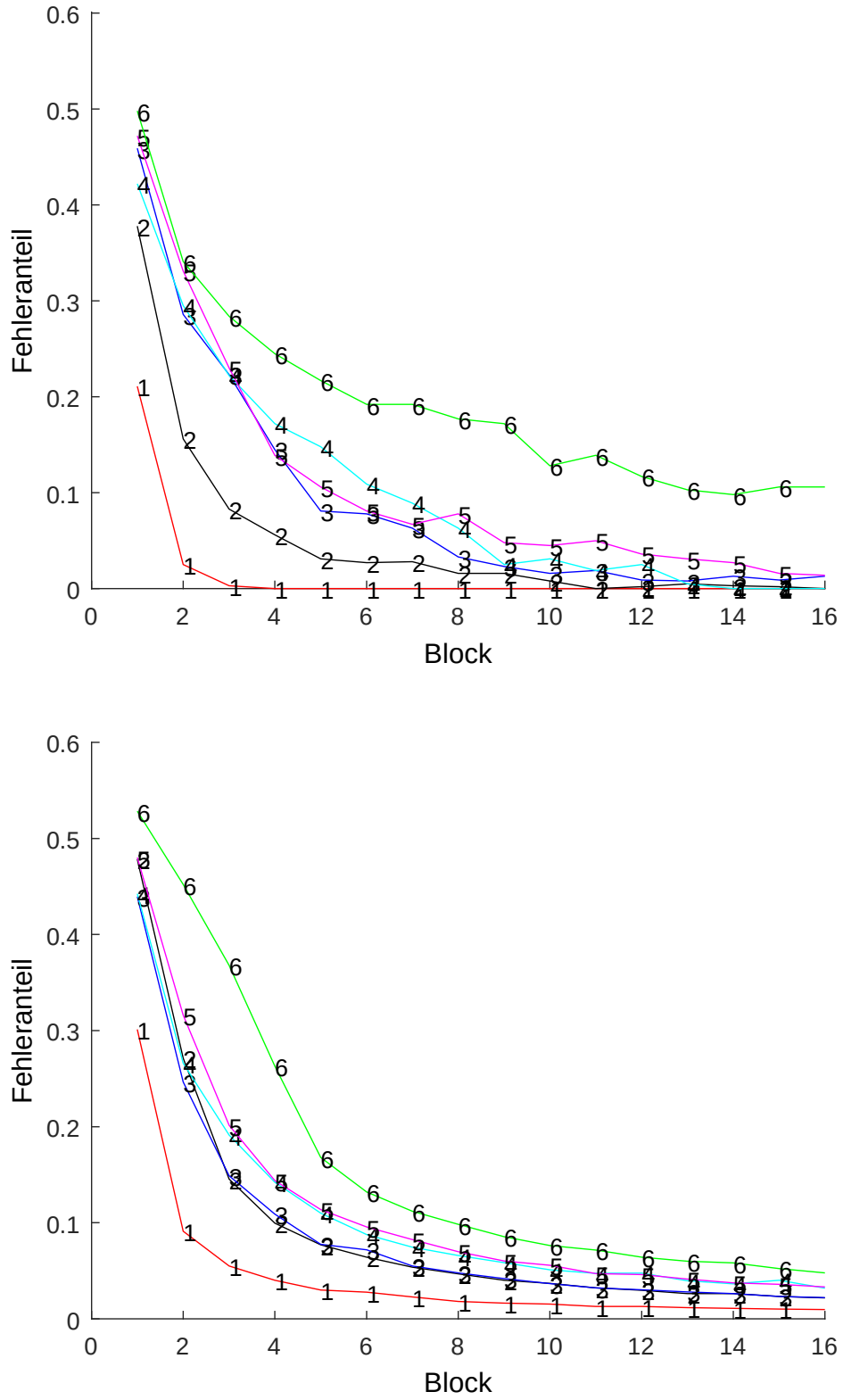


Abbildung 7. (Oben) Lernkurven der Kategorisierungsregeln 1 – 6 aus Nosofsky et al. (1994). (Unten) Lernkurven der mittleren Prognoserealisierung des URM (Variante 4).

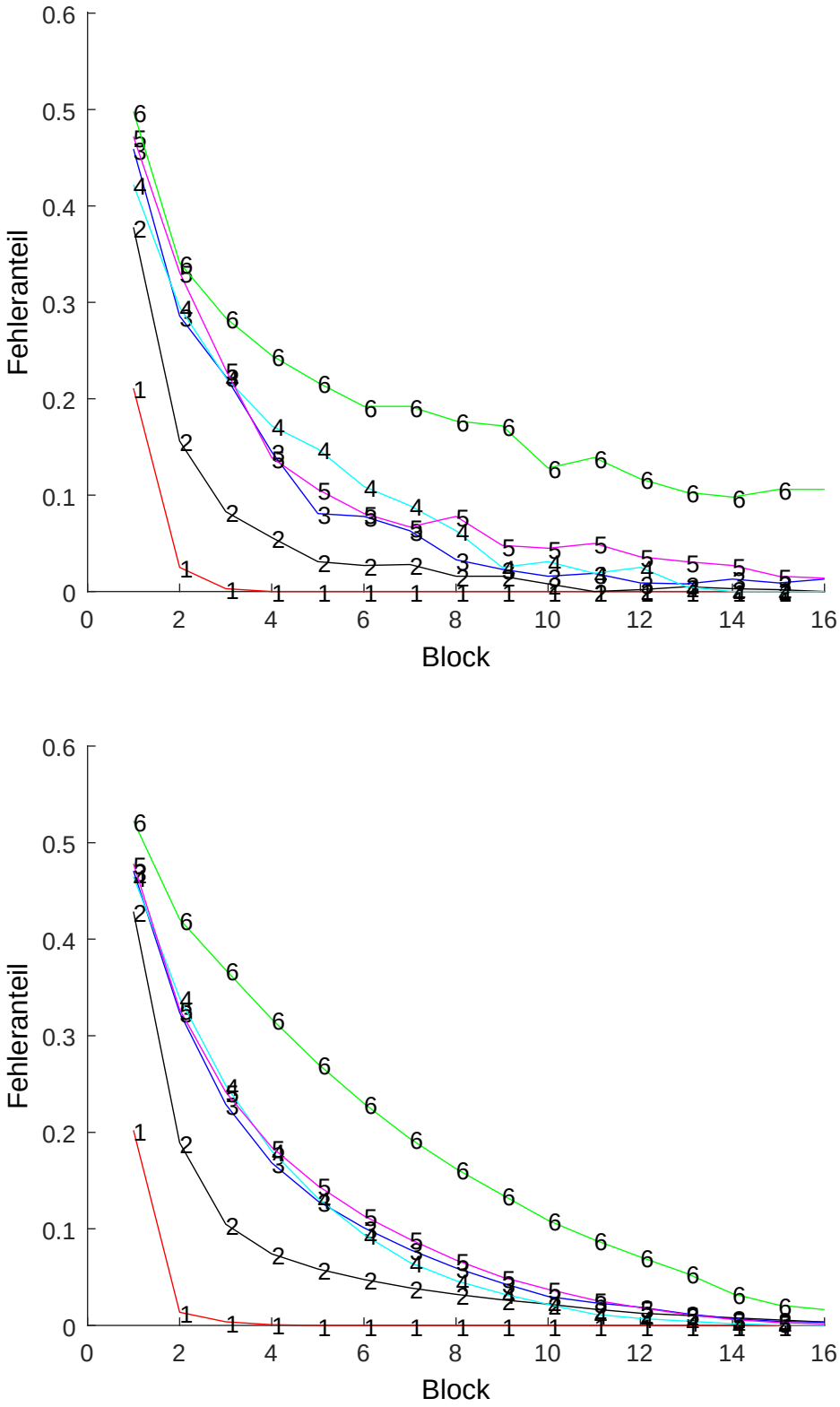


Abbildung 8. (Oben) Lernkurven der Kategorisierungsregeln 1 – 6 aus Nosofsky et al. (1994). Die Grafik ist aus Glassen und Nitsch (2015). (Unten) Lernkurven der mittleren Prognoserealisierung von SUSTAIN (Gureckis, 2014; Love et al., 2004).

Effekts ergab eine signifikant leichtere Regel 3 mit einem $\hat{p}$ = 0.097 und einem p-Wert < 0.001 mit 95% Konfidenzintervall [0.068, 0.127].

Um das URM und SUSTAIN unter ähnlicheren Modellierungsbedingungen gegenüberstellen zu können, wurde daher ersteres Modell mit einem vergleichbaren Mechanismus ausgestattet. Hierzu wurde die Likelihood aus (32) geändert in

$$\begin{aligned} F_k(o_{jN+1}) &= \prod_{d=1}^{D} \left(\frac{\beta^d_{o^d_{jN+1}} + O^d_{k o^d_{jN+1}}}{\beta^d_0 + \beta^d_1 + O^d_{k0} + O^d_{k1}}\right)^{Att^d} \\ F_{new}(o_{jN+1}) &= \prod_{d=1}^{D} \left(\frac{\beta^d_{o^d_{jN+1}}}{\beta^d_0 + \beta^d_1}\right)^{Att^d} \end{aligned} \tag{57}$$

und die Kategorisierungsfehlerfunktion E^{Cat} des zu kategorisierenden Stimulus $\boldsymbol{x}$, mit der korrekten Kategorie j^{corr} und der Dimensionsgewichte $Att^1,\ldots,Att^D$ für S Samples bestimmt:

$$\begin{aligned} E^{Cat}(x, j^{corr}, Att^1, \ldots, Att^D) &= \frac{\sum_{s=1}^{S} \sum_{j \in \{1,\ldots,J\} \setminus \{j^{corr}\}} \sum_{k \in \{k | n_{sjk} > 0\}} F_{sk}(x)}{\sum_{s=1}^{S} \sum_{j=1}^{J} \sum_{k \in \{k | n_{sjk} > 0\}} F_{sk}(x)} \\ &= \frac{\sum_{s=1}^{S} \sum_{j \in \{1,\ldots,J\} \setminus \{j^{corr}\}} \sum_{h \in \{k | n_{sjk} > 0\}} \prod_{d=1}^{D} \left(\frac{\beta^d_{\boldsymbol{x}^d} + O^d_{sk\,\boldsymbol{x}^d}}{\beta^d_0 + \beta^d_1 + O^d_{sk0} + O^d_{sk1}}\right)^{Att^d}}{\sum_{s=1}^{S} \sum_{j=1}^{J} \sum_{k \in \{k | n_{sjk} > 0\}} \prod_{d=1}^{D} \left(\frac{\beta^d_{\boldsymbol{x}^d} + O^d_{sk\,\boldsymbol{x}^d}}{\beta^d_0 + \beta^d_1 + O^d_{sk0} + O^d_{sk1}}\right)^{Att^d}} \end{aligned} \tag{58}$$

Anschließend können die Dimensionsgewichte $Att^1,\ldots,Att^D$ ausgehend von einer initialen uniformen Gewichtung $Att^1,\ldots,Att^D$ = 1, pro Trial nach Vorliegen eines Feedbacks dimensionsweise über Gradient-Descent mit einer Lernrate L angepasst werden:

$$Att^d = Att^d - L \cdot \frac{d}{d\,Att^d} E^{Cat}(\boldsymbol{x}, j^{corr}, Att^1, \ldots, Att^D) \tag{59}$$

Das modifizierte URM, im Weiteren als URM^{Att} bezeichnet, konnte in einer erneuten Modellierung vom Experiment 1 einen niedrigeren Mittelwert von 0.094 mit einer Varianz von 0.0002 in den 200 Samples aus $SSE(\Omega^{\theta}_{URM^{Att}})$ erreichen. Die Lernkurven der mittleren Prognoserealisierung sind in Abbildung 9 unten dargestellt. Dabei ist zu sehen, dass sich die Lernkurve 2 nun deutlich von den Lernkurven 3, 4 und 5 absetzt und somit dieser von Love et al. (2004) als schwierig zu modellierender Befund vom modifizierten URM vorhergesagt werden kann. Der durchschnittliche

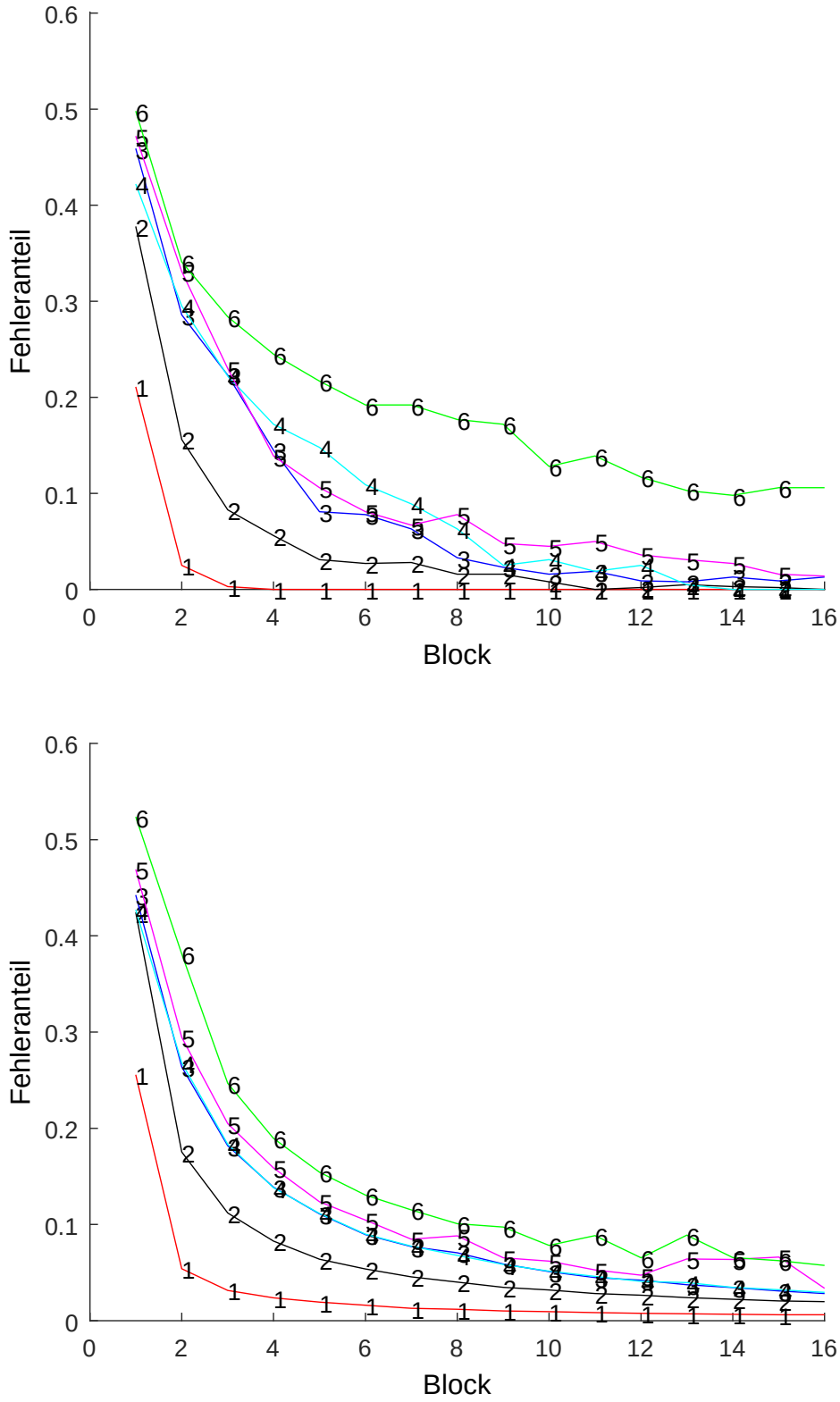


Abbildung 9. (Oben) Lernkurven der Kategorisierungsregeln 1 – 6 aus Nosofsky et al. (1994). Die Grafik ist aus Glassen und Nitsch (2015). (Unten) Lernkurven der mittleren Prognoserealisierung des URM^{Att}.

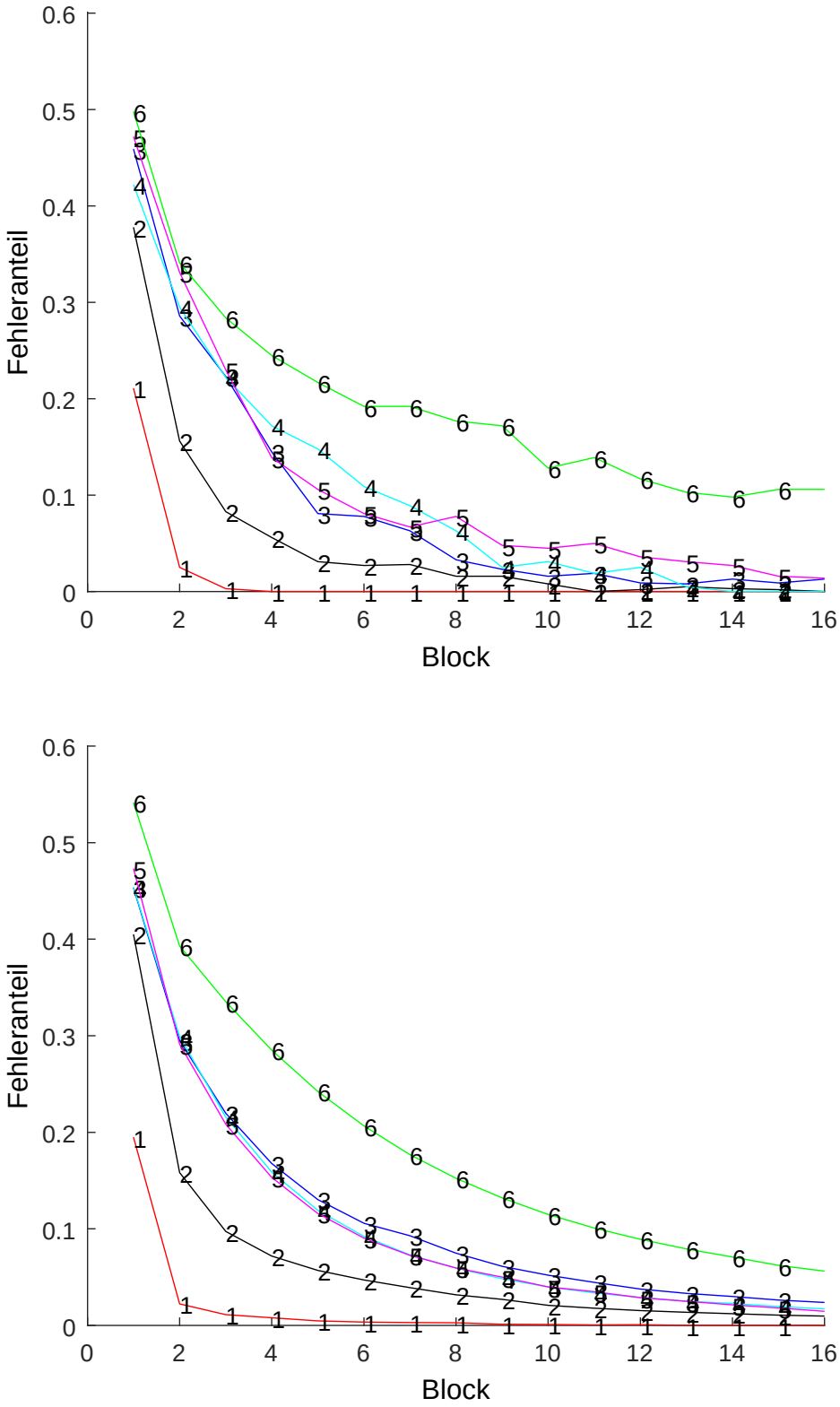


Abbildung 10. (Oben) Lernkurven der Kategorisierungsregeln 1 – 6 aus Nosofsky et al. (1994). Die Grafik ist aus Glassen und Nitsch (2015). (Unten) Lernkurven der mittleren Prognoserealisierung von SUSTAIN nach Korrektur des Modellierungscodes von Gureckis (2014).

Fehleranteil bei Kategorisierungsregel 2 hatte in den bereits gezogenen 200 Samples aus Ω^{θ}_{URM} einen Mittelwert von 0.075 und eine Varianz von 7.11e-07 und bei Kategorisierungsregel 3 einen Mittelwert von 0.107 mit Varianz von 1.72e-06. Ein Vergleich der durchschnittlichen Fehlerhäufigkeiten bei beiden Regeln über die Brunner und Munzel (2000) Methode ergab einen geschätzten relativen Treatment Effekt von $\hat{p}$ = 0.999 mit einem *p*-Wert < 0.001 und 95% Konfidenzintervall [0.992, 1.006].

Trotz dieser Modifikation liegt noch keine ausreichende Vergleichbarkeit beider Modellierungen vor. Gureckis (2014) und somit vermutlich auch Love et al. (2004) verwendeten nämlich in ihrer Modellierung eine subblockweise ausbalancierte Randomisierung über jeweils 8 Trials. Das trifft entsprechend dem Artikel von Nosofsky et al. (1994) jedoch nur auf die ersten beiden Blöcke zu, in welchen jeder Stimulus nur einmal auftritt. In allen nachfolgenden Blöcke à 16 Trials treten hingegen alle Stimuli zweimal in einer zufälligen Reihenfolge auf, wodurch auch Stimulusreihenfolgen möglich sind, in welchen innerhalb der ersten 8 Trials vier Doppelpräsentationen vorliegen. Letzteres kann in der Simulation von Gureckis (2014) nicht auftreten.

Der Code von Gureckis (2014) wurde daher in Bezug auf die Randomisierung korrigiert und eine erneute Modellierung durchgeführt. Dabei zeigte sich sogar eine Verkleinerung des Mittelwerts der 200 Samples aus $SSE(\Omega^{\theta}_{SUS})$ von ursprünglich 0.080 auf 0.045 mit einer neuen Varianz von 3.35e-05. Die Lernkurven der mittleren Prognoserealisierung finden sich in Abbildung 10 unten.

Einen abschließenden Vergleich von URM^{Att} und SUSTAIN mit korrigierter Randomisierung ergab einen geschätzten relativen Treatment Effekt von $\hat{p}$ = 0.001 mit einem p-Wert < 0.001 für die jeweils 200 Samples aus $SSE(\Omega^{\theta}_{URM^{Att}})$ und $SSE(\Omega^{\theta}_{SUS})$. SUSTAIN weist somit die klar niedrigere mittlere SSE auf. Abbildung 11 zeigt das 95% Konfidenzintervall der Schätzung von $\hat{p}$ sowie die geschätzten Dichten der SSE-Verteilungen.

Tabelle 5. Ermittelte Parameter für URM und URM^{Att} für Experiment 1.

	β	γ_0^{α}	γ_1^{α}	γ_0^{y}	γ_1^{y}	L
URM	0.22865	0.36139	0.19081	0.31168	0.40134	-
URM^{Att}	0.34109	6.162	0.35838	0.19159	0.27436	3.6553

Tabelle 6. Ursprüngliche und nach der Codekorrektur ermittelte Parameter für SUSTAIN für Experiment 1.

	r	β	m	η
SUSTAIN	9.01245	1.252233	16.924073	0.092327
SUSTAIN^{Corr}	2.1481	0.5596	4.3117	0.91243

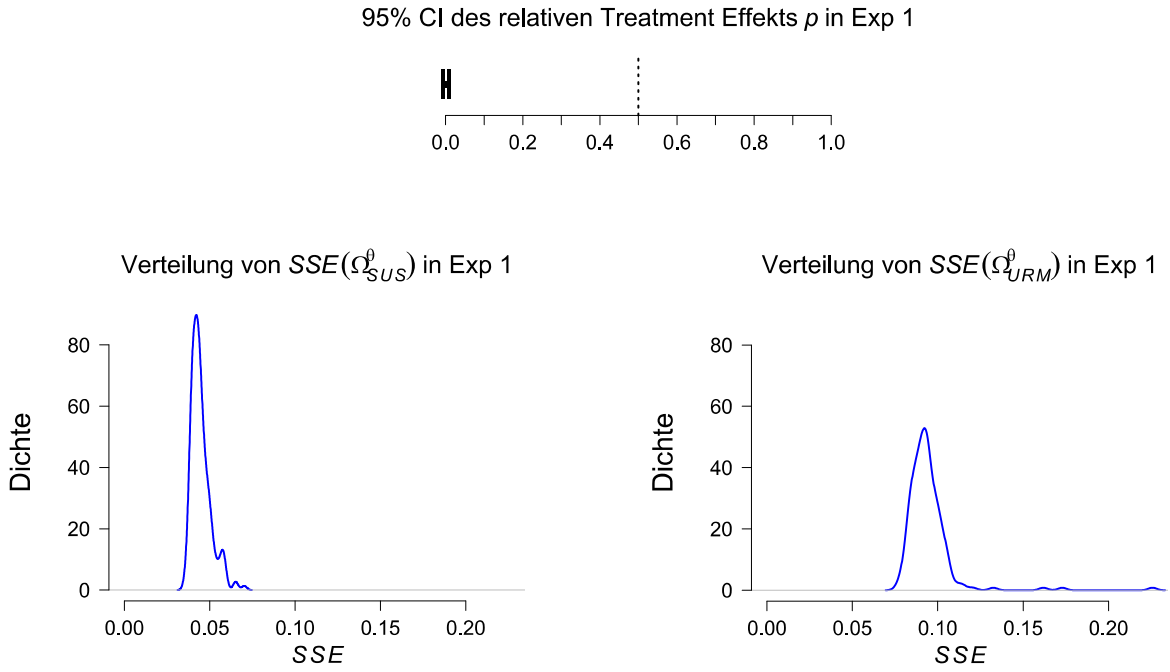


Abbildung 11. (Oben) Relativer Treatment Effekt von URM^{Att} und SUSTAIN in Experiment 1. (Unten) Anhand der Stichproben geschätzte Dichte der SSE-Verteilungen von URM^{Att} und SUSTAIN in Experiment 1.

In den Tabellen 5 und 6 sind die besten gefundenen Parameter für URM, URM^{Att} sowie die alten und neuen besten Parameter für SUSTAIN aufgeführt.

Durch eine Modell-Flexibilitätsanalyse des URM^{Att} und SUSTAIN bezüglich der möglichen Prognosen für Experiment 1 wurde für Ersteres ein $\phi_1^{URM^{Att}} = 0.00078$ und für Letzteres eine leicht größere Flexibilität mit $\phi_1^{SUS} = 0.00135$ ermittelt. SUSTAIN ist somit leicht flexibler als das URM in Experiment 1.

3.3 Experiment 2: Identifikations vs. Klassifikationsschwierigkeit bei komplexen Stimuli

Wie bei Experiment 1 bereits angesprochen, weisen Identifikationsaufgaben typischerweise eine höhere Schwierigkeit auf als Klassifikationsaufgaben. Dies gilt jedoch nicht generell. Medin et al. (1983) konnten zeigen, dass eine Kategorisierungsaufgabe, welche ungefähr dem Problemtyp 4 des Experiments 1 entspricht, schwieriger zu lernen war als eine Identifikationsaufgabe mit den

gleichen Stimuli. Hierbei war ausschlaggebend, dass die Stimuli nicht nur auf den für die Kategorisierung potentiell relevanten Dimensionen variierten sondern auch jeweils eigene sogenannte idiosynkratische Informationen besaßen. Konkret handelte es sich um Fotografien von Frauen aus alten College Jahrbüchern, welche sich hinsichtlich der Haarfarbe, der Haarlänge, der Art des Lächelns und der Farbe des Oberteils systematisch unterschieden. Zusätzlich variierten die Bilder, aufgrund der Natur der Stimuli, unsystematisch auf vielen anderen Dimensionen, beispielsweise der Nasenform, der Augenfarbe oder der Ausprägung der Wangenknochen der abgebildeten Frauen. Medin et al. (1983) verzichteten in ihrem Artikel auf eine Präsentation von Stimulusbeispielen, weshalb in der vorliegenden Arbeit ebenfalls keine grafischen Darstellungen aufgeführt sind.

Im Experiment der Autoren hatten 64 VP die Aufgabe die Zuordnung von neun Frauen zu zwei Familien Asch und Boyd anhand ihrer Fotografien zu erlernen (= Nachnamen-Bedingung). Die logische Struktur der Zugehörigkeit zu den jeweiligen Familien ist in Tabelle 7 aufgeführt. Die konkrete Bedeutung der Dimensionsausprägungen und somit die für eine VP verwendeten Fotografien wurde zwischen den VP randomisiert. Pro Übungsblock wurden den VP neun Fotografien in randomisierter Reihenfolge präsentiert. Dabei folgte auf jede Kategorisierungsentscheidung ein Feedback über die Korrektheit der Zuordnung. Das Experiment endete, wenn zwei aufeinanderfolgende fehlerfreie Blöcke bearbeitet wurden, oder bei Nicht-Erreichen dieses Kriteriums nach maximal 16 Blöcken.

Eine weitere Gruppe von 32 VP hatte die Aufgabe anstelle des Familiennamens die Vornamen der neun Frauen zu erlernen (= Vornamen-Bedingung). Diese waren in zufälliger Zuordnung von

Tabelle 7. Logische Struktur der Stimulusattribute auf den für die Kategorisierung potentiell relevanten Dimensionen in Medin et al. (1983).

	Vorname	Familie	Haarfarbe	Oberteilfarbe	Art des Lächelns	Haarlänge
Stimuli	N1	Asch	1	1	1	0
	N2	Asch	1	0	1	0
	N3	Asch	1	0	1	1
	N4	Asch	1	1	0	1
	N5	Asch	0	1	1	1
	N6	Boyd	1	1	0	0
	N7	Boyd	0	1	1	0
	N8	Boyd	0	0	0	1
	N9	Boyd	0	0	0	0

Vorname zu Frau die Namen Anne, Joan, Kate, Emma, Mary, Sara, Ruth, Cora und Lucy. Analog zur VP-Gruppe mit der Nachnamen-Bedingung hatten die 32 VP zwei aufeinanderfolgende fehlerfreie Blöcke zu absolvieren oder maximal 16 Blöcke zu bearbeiten.

Während nun die Nachnamen-Bedingung in etwa der Kategorisierungschwierigkeit des Lernproblems 4 aus Experiment 1 entspricht, handelt es sich bei der Vornamen-Bedingung um eine Identifikationsaufgabe bei welcher die Kategorie der einzelnen Stimuli analog zum Lernproblem 6 nicht nach einer Regel hergeleitet werden kann, sondern erinnert werden muss (Love et al., 2004). Anders als im Experiment 1 zeigte sich jedoch, dass die Vornamen- gegenüber der Nachnamen-Bedingung leichter zu erlernen war. VP der Letzeren konnten das Abbruchkriterium von zwei aufeinanderfolgenden fehlerfreien Blöcken im Durchschnitt nach 9.7 Blöcken erreichen, während VP der Ersteren bereits nach durchschnittlich 7.1 Blöcken abschlossen. Hierbei hatten die VP der Nachnamen-Bedingung eine durchschnittliche Kategorisierungskorrektheit von 87% und die der Vornamen-Bedingung von 84% (Love et al., 2004).

3.3.1 Modellierungsergebnisse für Experiment 2

Um die zwei Bedingungen aus Medin et al. (1983) mit dem URM zu modellieren, musste zunächst eine Modifikation am Modell vorgenommen werden. Dies war notwendig, da das URM von Griffiths et al. (2007) nur binäre Dimensionen unterstützt und somit idiosynkratische Informationen von neun Stimuli nicht codiert werden können. Die Dimensionsausprägungen eines Stimulus wurden deshalb anstatt aus einer Beta-Bernoulli- aus einer Dirichlet-Kategorialverteilung mit $\theta^d \sim \text{Dir}(\beta^d_{0,\dots},\beta^d_V)$ gezogen, sodass nun auf jeder Dimension V unterschiedliche Ausprägungen möglich sind. Die Likelihood $F_k(o_{jN+1})$ bzw. $F_{new}(o_{jN+1})$ aus (32) wurde entsprechend generalisiert:

$$F_k(o_{jN+1}) = \prod_{d=1}^{D}\left(\frac{\beta^d_{o^d_{jN+1}}+O^d_{ko^d_{jN+1}}}{\sum_{v=1}^{V}(\beta^d_v+O^d_{kv})}\right)$$
$$F_{new}(o_{jN+1}) = \prod_{d=1}^{D}\left(\frac{\beta^d_{o^d_{jN+1}}}{\sum_{v=1}^{V}\beta^d_v}\right) \quad (60)$$

Analog zur Modellierung von Love et al. (2004) mit SUSTAIN wurde nachfolgend für die Modellierung mit dem modifizierten URM, im Weiteren URM^{Cat} genannt, die in Tabelle 7

aufgeführte logische Struktur um eine zusätzliche Dimension erweitert und jedem Stimulus eine einzigartige Nummer zwischen 0 und 8 zugewiesen, welche die charakteristische idiosynkratische Information des jeweiligen Stimulus repräsentierte.

Mit dem URM^{Cat} wurden anschließend per GA die besten Parameter ermittelt, wobei bewusst auf ein zusätzliches Fitting der durchschnittlichen Kategorisierungskorrektheit zugunsten einer bestmöglichen Vorhersage der relativen Anzahl an Blöcken verzichtet wurde. Letztere wurde für das URM^{Cat} folgendermaßen ermittelt: Zunächst wurde anhand der trialweisen Kategorisierungsfehler im Experiment die Wahrscheinlichkeit jedes Blocks berechnet, der letzte bearbeitete Block zu sein, wenn das Stoppkriterium zwei fehlerfrei bearbeitete konsekutive Blöcke sind. Für diese Wahrscheinlichkeitsverteilung über die Blöcke 1 bis 16 wurde schließlich der Erwartungswert bestimmt, welcher somit die erwartete Anzahl an benötigten Böcken in der jeweiligen Bedingung darstellt.

Es zeigte sich schließlich in den 200 Samples aus $SSE\left(\Omega^{\theta}_{URM^{Cat}}\right)$ eine mittlere SSE von 0.82 mit einer Varianz von 2.06, wobei die mittlere Prognoserealisierung bei 9.6 Blöcken für die Nachnamenbedingung und bei 6.2 Blöcken für die Vornamenbedingung lag. Die ungefittete durchschnittliche Kategorisierungskorrektheit wurde dabei mit 0.70 und 0.53 für die Nachnamen- bzw. Vornamenbedingung prognostiziert.

Love et al. (2004) modellierten beide Bedingungen mit SUSTAIN, indem Sie, wie bereits oben erwähnt, der in Tabelle 7 aufgeführten logischen Struktur eine weitere Dimension hinzufügten, in welcher jeder Stimulus stellvertretend für die idiosynkratischen Informationen eine eindeutige Ausprägung bekam. Zusätzlich führten sie einen weiteren Parameter ein (= $\lambda_{distinct}$), welcher den initialen Wert des Tuning-Parameters λ_d für diese Dimension deklariert. Damit überließen sie es der Parameteroptimierung, welche Bedeutung der Dimension für einen bestmöglichen Fit zukommt. Love et al. (2004) erhielten abschließend eine Prognose von SUSTAIN von 9.7 benötigten Übungsblöcken und 85% durchschnittliche Kategorisierungskorrektheit für die Nachnamen-Bedingung und 7.2 Blöcke mit 87% durchschnittlicher Korrektheit für die Vornamen-Bedingung.

Eine erneute Parameteroptimierung und Modellierung mit SUSTAIN ergab hingegen eine durchschnittliche SSE von 0.93 der 200 Samples aus $SSE\left(\Omega^{\theta}_{SUS}\right)$ bei einer Varianz von 0.71. Die mittlere Prognoserealisierung wies eine benötigte Anzahl von Blöcken von 8.8 bzw. 6.9 für die Nachnamen bzw. Vornamenbedingung auf, wobei die ungefittete prognostizierte durchschnittliche Kategorisierungskorrektheit für beide Bedingungen bei 0.76 lag. Für den geschätzten relativen Treatment Effekt zwischen der URM^{Cat}- und SUSTAIN-Modellierung wurde ein $\hat{p}$ =0.61 mit einem p-Wert < 0.001 ermittelt. Dies spricht für einen klar besseren Fit der Prognose des URM^{Cat}.

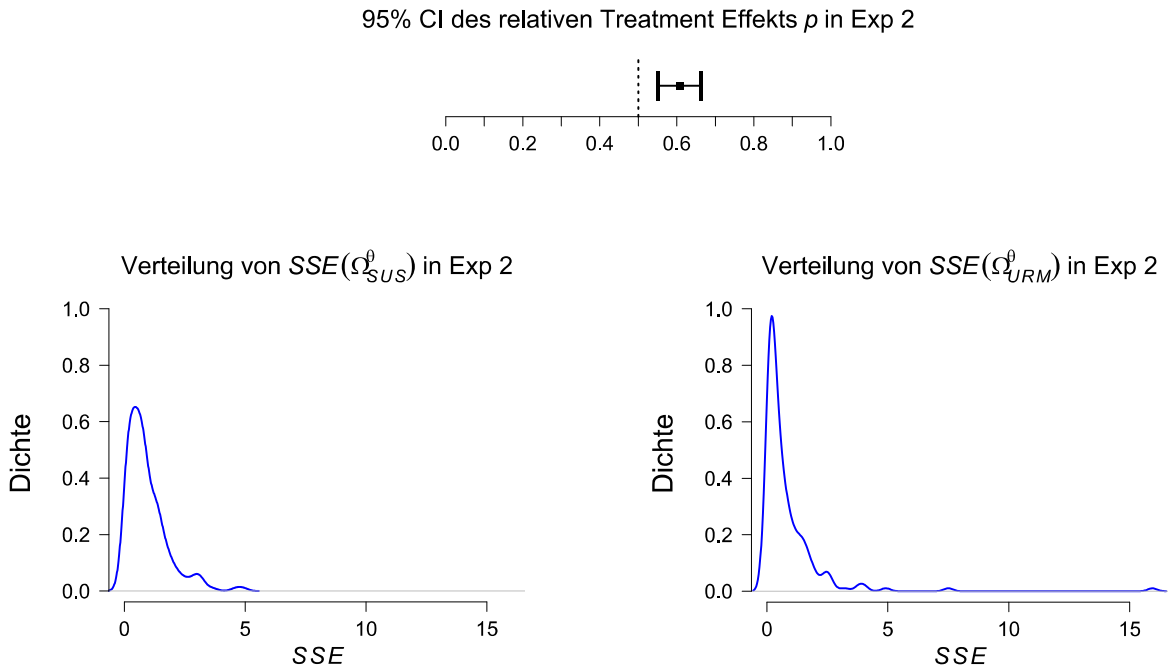


Abbildung 12. (Oben) Relativer Treatment Effekt von URM^{Cat} und SUSTAIN in Experiment 2. (Unten) Anhand der Stichproben geschätzte Dichte der SSE-Verteilungen von URM^{Cat} und SUSTAIN in Experiment 2.

Abbildung 12 zeigt die geschätzte Dichte der SSE-Verteilungen, sowie das Konfidenzintervall von $\hat{p}$. Die ermittelten Parameter der Lösungen für das URM^{Cat} und SUSTAIN sind in Tabelle 8 und 9 aufgeführt.

Eine Modell-Flexibilitätsanalyse bezüglich der möglichen Prognosen für Experiment 2 ergab für das URM^{Cat} ein $\phi_2^{URM^{Cat}} = 0.00157$ und für SUSTAIN ein $\phi_2^{SUS} = 0.01557$, was somit auf ein um ca. den Faktor 10 flexibleres SUSTAIN hindeutet, qualitativ verschiedene Befunde vorherzusagen.

Tabelle 8. Ermittelte Parameter für das URM^{Cat} für Experiment 2.

	β	γ_0^{α}	γ_1^{α}	γ_0^{γ}	γ_1^{γ}
URM^{Cat}	0.31133	3.5027	0.015403	0.42298	1.0419

Tabelle 9. Ursprüngliche und in einer wiederholten Modellierung ermittelten Parameter für SUSTAIN für Experiment 2.

	r	β	m	η	$\lambda_{distinct}$
SUSTAIN	4.349951	5.925613	15.19877	0.0807908	5.213135
SUSTAIN^{Rep}	3.4912	3.0409	5.1894	1	2.4425

3.4 Experimente 3 und 4: Inferenz- vs. Klassifikationsschwierigkeit bei linear separablen und nicht-separablen Kategorien

In den bisherigen Experimenten wurden Kategorien ausschließlich über Klassifikationsaufgaben erlernt. Hierbei handelt es sich jedoch nicht um die einzige Möglichkeit Kategorienwissen zu erlangen. Yamauchi und Markman (1998) stellten sich die Frage, ob die mentale Struktur einer zu lernenden Kategorie abhängig ist von der mit ihr zu bewältigenden Aufgabe. Sie untersuchten deshalb in ihrer Studie die Transferleistung von VP, welche Klassifikationsaufgaben zu lösen hatten mit VP, welche Inferenzaufgaben bearbeiteten. VP der ersteren Gruppe, welche im Weiteren als Klassifikationsgruppe bezeichnet wird, hatten die Aufgabe, wie in den vorhergehenden Experimenten Stimuli in eine von zwei Kategorien einzuordnen, wobei sie nach jeder Kategorisierungsentscheidung Feedback über die Korrektheit der Klassifikation erhielten. VP der Gruppe, welche Inferenzaufgaben zu lösen hatten, im Folgenden als Inferenzgruppe bezeichnet, sollten hingegen in jedem Trial die Ausprägung eines Stimulus in einer spezifizierten Dimension angeben, wenn alle anderen Dimensionsausprägungen und die Kategorienzugehörigkeit des Stimulus bekannt waren. Wie in der Klassifikationsbedingung wurde auch hier nach jedem Trial Feedback über dir Korrektheit der Entscheidung gegeben. Yamauchi und Markman (1998) testeten jeweils 24 VP pro Gruppe. Jede dieser VP hatte zunächst in einer Trainingssitzung maximal 30 Blöcke à 8 Stimuli mit einem Stimulus pro Trial zu bearbeiten bzw. bis in drei aufeinanderfolgenden Blöcken mindestens 90% korrekte Antworten abgegeben wurden. Anschließend folgte, nach einer 10 minütigen Füllphase mit einer für das Experiment irrelevanten Aufgabe, eine Testphase, in welcher die Transferleistung jeder VP anhand von Klassifikations- und Inferenzaufgaben mit den Stimuli aus der Lernphase zuzüglich der Prototypenstimuli beider Kategorien erfasst wurde. Letztere stellen Stimuli mit Ausprägungen in jeder Dimension dar, welche unter den Mitgliedern der Kategorie am häufigsten vorkommen. Bei den Stimuli handelte es

Tabelle 10. Logische Struktur der Stimuli aus Yamauchi und Markman (1998).

	Kategorie A				Kategorie B			
	Form	Größe	Farbe	Position	Form	Größe	Farbe	Position
Stimuli	1	1	1	0	0	0	0	1
	1	1	0	1	0	0	1	0
	1	0	1	1	0	1	0	0
	0	1	1	1	1	0	0	0

sich um rote oder grüne geometrische Figuren (Dreieck oder Kreis) in zwei Größen (groß oder klein), welche je nach Stimulus entweder links oder rechts am Bildschirm angezeigt wurden. Die logische Struktur der Stimuli aus der Lernphase ist in Tabelle 10 aufgelistet. VP der Inferenzgruppe bearbeiteten Trials wie sie in Abbildung 13b dargestellt sind. VP der Klassifikationsgruppe erhielten hingegen Trials nach dem Muster von Abbildung 13a.

Die Reihenfolge der präsentierten Stimuli war randomisiert, ebenso die Reihenfolge der zu inferierenden Dimensionsausprägungen. In jedem Klassifikationsblock wurde jeder Stimulus genau einmal und in jedem Inferenzblock jede Dimension mindestens einmal präsentiert. Die Abfragereihenfolge der Dimensionen wurde dabei so gewählt, dass im gesamten Experiment für jeden Stimulus alle Dimensionen mit Ausnahme der Dimension abgefragt wurden, in welcher der Stimulus vom Prototypenstimulus der Kategorie abweicht. Für Kategorie A und dem zugehörigen Prototypenstimulus mit den logischen Ausprägungen [1, 1, 1, 1] weichen demnach die Dimensionsausprägungen mit dem Wert 0 von der Ausprägung des Prototypenstimulus ab. Ebenso weichen die Ausprägungen mit Wert 1 von der entsprechenden Dimensionsausprägung des Prototypenstimulus [0, 0, 0, 0] für Kategorie B ab. Die Abweichungen eines Stimulus vom Prototypen werden im Weiteren als Ausprägungsabweichungen bezeichnet. Sie sind in Tabelle 10 auf der Gegendiagonalen der Strukturmatrix der jeweiligen Kategorie einsehbar.

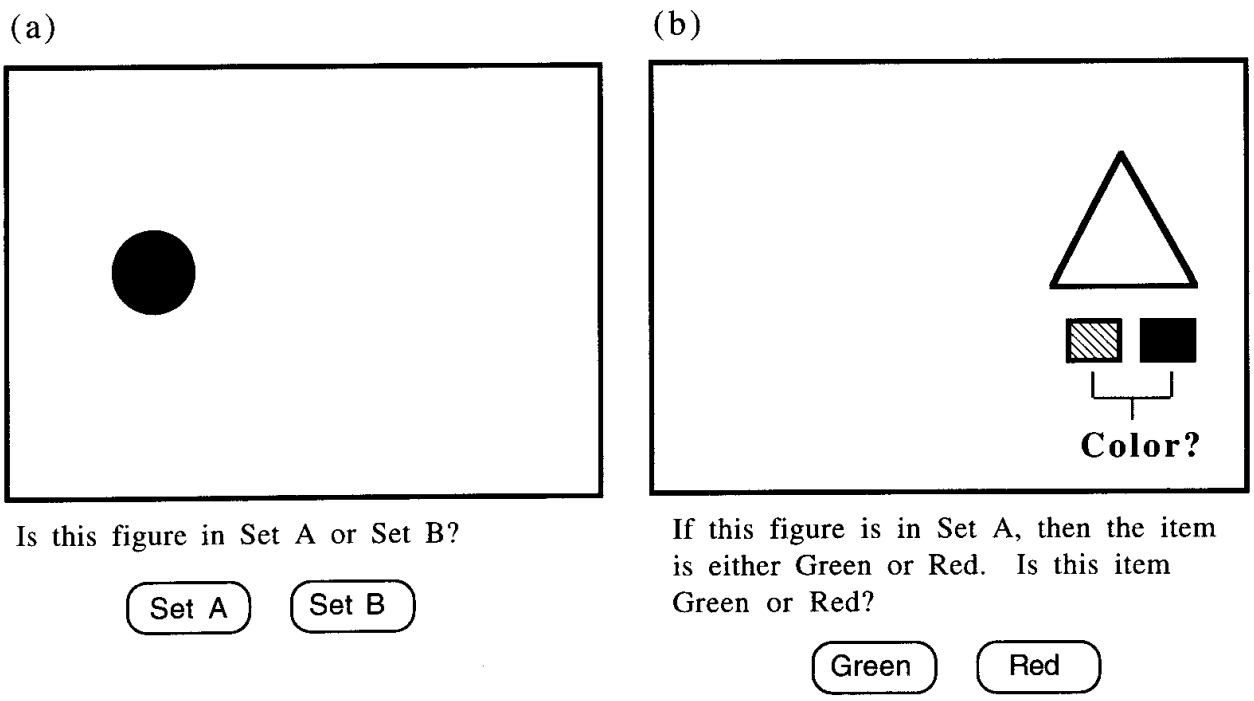


Abbildung 13. Beispielhafte Trials von VP mit Klassifikationsaufgaben (a) und Inferenzaufgaben (b) aus Yamauchi und Markman (1998). Nachgedruckt mit Genehmigung von Elsevier.

Es zeigte sich in der Testphase, dass die mentale Kategorienrepräsentation der VP eine Struktur aufweisen muss, welche hinsichtlich der jeweiligen Kategorisierungsaufgabe im Training optimiert ist. So konnten VP der Klassifikationsgruppe 92% der Stimuli aus der Lernphase korrekt klassifizieren, während VP der Inferenzgruppe nur 77% korrekt zuordneten. Demgegenüber konnten VP der letzteren Gruppe 94% der Inferenzaufgaben in der Testphase korrekt lösen, während die VP der Klassifikationsgruppe nur 81% korrekt beantworteten. Dass sich die mentale Kategorienstruktur zwischen den VP beider Gruppen unterschied, konnte zudem durch Betrachtung der Klassifikationsleistung bei den Prototypenstimuli und der Inferenzleistung bei Stimulusdimensionen mit Ausprägungsabweichungen bestätigt werden. Während die Klassifikation der Prototypenstimuli für VP beider Gruppen mit jeweils 96% korrekter Antworten gleich leicht war, gaben VP der Inferenzgruppe beim Inferieren der Ausprägung in Stimulusdimensionen mit Ausprägungsabweichungen signifikant häufiger die Dimensionsausprägung des Prototypen an (in 86% vs. 64% der Trials). Yamauchi und Markman (1998) schlossen daraus, dass die mentalen Strukturen von Kategorien, welche über Inferenzaufgaben erlernt wurden, Codierungen der prototypischen Ausprägungen der Kategorienmitglieder darstellen. Mentale Strukturen von Kategorien, welche über Klassifikationsaufgaben angeeignet wurden, beinhalten nach Yamauchi und Markman (1998) hingegen Informationen über Regeln, Ausnahmen oder konkrete Exemplare. Neben diesen Befunden stellten Yamauchi und Markman (1998) noch eine weitere Besonderheit der beiden Aufgabentypen fest. So benötigten VP aus der Inferenzgruppe durchschnittlich nur 6.5 Trainingsblöcke um das Kriterium von 90% korrekten Trails in drei aufeinander folgenden Blöcken zu erfüllen, während VP aus der Klassifikationsgruppe hierzu durchschnittlich 12.3 Blöcke benötigten.

Yamauchi et al. (2002) überprüften in ihrer Studie Yamauchi und Markmans (1998) Postulat zu den aus beiden Lernmodi resultierenden mentalen Kategorienstrukturen, indem sie das Experiment mit einer nicht separablen Kategorienstruktur wiederholten. Die logische Kategorienstruktur von Yamauchi et al. (2002) ist in Tabelle 11 dargestellt. Die visuelle Darstellung der Stimuli entspricht der aus dem Experiment von Yamauchi und Markman (1998). Wie zu sehen ist, unterscheiden sich der zweite Stimulus von Kategorie A und der dritte Stimulus von Kategorie B nur in der Ausprägung der Dimension ‚Größe'. Dabei ist der Erstgenannte dem Stimulus in der ersten Zeile der Kategorie B mit nur einer Abweichung ähnlicher als der Zweitgenannte, welcher Abweichungen auf zwei Dimensionen aufweist. Wie bei erneuter Betrachtung der Definition von linearer Separabilität in (1) erkennbar ist, kann hier, ganz egal wie wir das Gewicht w_2 und der Threshold θ wählen, für diese drei Stimuli (1) nicht gelten. Die Kategorien sind somit nicht linear separabel.

Tabelle 11. Logische Struktur der Stimuli aus Yamauchi et al. (2002).

	Kategorie A				Kategorie B			
	Form	Größe	Farbe	Position	Form	Größe	Farbe	Position
Stimuli	1	1	1	1	1	1	0	1
	1	1	0	0	0	1	1	0
	0	0	1	1	1	0	0	0

Da in jeder Stimulusdimension von Kategorie A die 1 am häufigsten vorkommt, stellt [1, 1, 1, 1] der Prototypenstimulus dieser Kategorie dar. Entsprechend handelt es sich bei [1, 1, 0, 0] um den Prototypenstimulus der Kategorie B, welchem zudem eine besondere Rolle im Experiment von Yamauchi et al. (2002) zukommt. So ist aus Tabelle 11 ersichtlich, dass es sich nicht nur um den Prototypen der Kategorie B handelt sondern zusätzlich um ein Mitglied aus Kategorie A. Wenn nun nach der Annahme von Yamauchi und Markman (1998) durch Inferenzaufgaben erworbene mentale Kategorienstrukturen primär Prototypenwissen kodieren, sollte sich die Klassifikation des Prototyps der Kategorie B als ein Stimulus von Kategorie A unter dieser Struktur als schwierig erweisen (Yamauchi et al., 2002). VP, welche hingegen die Kategorien über Klassifikationsaufgaben trainierten, sollten nach Yamauchi et al. (2002) keine beeinträchtigte Leistung aufweisen.

Das Experiment von Yamauchi et al. (2002) wurde analog zum Experiment von Yamauchi und Markman (1998) mit 48 VP, jeweils 24 VP in der Klassifikations- und der Inferenzgruppe, durchgeführt. Die Trainingsphase bestand demnach aus maximal 30 Blöcken à 6 Stimuli, in welchen jeder Stimulus aus Tabelle 11 einmal präsentiert wurde. Wie bei Yamauchi und Markman (1998) lag das vorzeitige Abbruchkriterium bei 90% korrekt gelösten Klassifikations- bzw. Inferenzaufgaben über die Dauer von drei aufeinanderfolgenden Blöcken. Die Struktur der Testphase entsprach ebenfalls der aus Yamauchi und Markman (1998), in welcher von jeder VP jeder Stimulus einmal klassifiziert und jede Dimensionsausprägung jedes Stimulus einmal inferiert werden sollte. Die Leistung beider Gruppen in der Testphase entsprach schließlich der eingangs beschriebenen Erwartung von Yamauchi et al. (2002).

Während die Klassifikationsgruppe die Klassifikationsaufgaben der Testphase zu 94% korrekt löste, erreichte die Inferenzgruppe lediglich eine Korrektheit von 70%. Der Unterschied zwischen den Gruppen trat bei Einzelbetrachtung der Klassifikationsleistung der Prototypenstimuli nochmals deutlicher hervor. So klassifizierte die Klassifikationsgruppe die Prototypenstimuli von Kategorie A und B mit 88% und 92% annähernd gleich gut, während die Inferenzgruppe eine starke Divergenz in der Kategorisierungsleistung zwischen dem Prototypenstimulus der Kategorie A (83% korrekt) und Kategorie B (46% korrekt) aufwies. Dieser Befund ist in Übereinstimmung mit der Annahme

von Yamauchi et al. (2002), dass die über Inferenzaufgaben erworbene mentale Kategorienstruktur Informationen über den Prototypenstimulus der Kategorie repräsentiert.
Bei Betrachtung der Inferenzleistung beider Gruppen in der Testphase ergab sich ein ähnliches Bild. So inferierte die Klassifikationsgruppe 85% der Dimensionsausprägungen korrekt, während die Inferenzgruppe nur 68% Korrektheit erreichte. Gleichzeitig zeigte sich in der Klassifikationsgruppe, nicht aber in der Inferenzgruppe, eine verglichen mit den Dimensionen ‚Form' und ‚Größe' höhere Korrektheit bei der Inferierung von Farbausprägungen. Die Klassifikationsgruppe löste hierbei 81% bzw. 80% der Form- bzw. Größenausprägungen und 88% der Farbausprägungen korrekt. Die entsprechenden Werte der Inferenzgruppe lagen bei 72%, 68% und 73%. Da die Stimulusausprägung in der Farbdimension gegenüber der Ausprägung in den Dimensionen ‚Form' und ‚Größe' ein besserer Indikator für die Gruppenzugehörigkeit darstellt (siehe Tabelle 11), spricht dies nach Yamauchi et al. (2002) für die Annahme, dass über Klassifikationsaufgaben erworbene mentale Kategorienstrukturen Regeln, Ausnahmen oder konkrete Exemplare kodieren und die Aufmerksamkeit während des Lernprozesses auf prädiktivere Dimensionen lenken.
Neben der Bestätigung der Annahmen von Yamauchi und Markman (1998) konnten Yamauchi et al. (2002) zudem ein qualitativ konträres Muster bei der Lerngeschwindigkeit von Klassifikations- und Inferenzaufgaben feststellen, wenn anstelle der ursprünglichen linear trennbaren eine nicht linear separable Kategorienstruktur verwendet wird. Wie oben beschrieben, zeigte sich bei einer linear trennbaren Kategorienstruktur ein deutlicher Vorteil für die Lerngeschwindigkeit bei Inferenzaufgaben. Liegt jedoch eine nicht linear separable Struktur vor, wie in Yamauchi et al. (2002) geschehen, dreht sich das ursprünglich bei Yamauchi und Markman (1998) ermittelte Muster, sodass nun das Abbruchkriterium bei Klassifikationsaufgaben früher erreicht wird als bei Inferenzaufgaben (10.4 vs. 27.4 benötigte Blöcke). Die Modellierung dieses Phänomens soll im Folgenden betrachtet werden.

3.4.1 Modellierungsergebnisse für die Experimente 3 und 4

Um mit dem URM beide Experimente mit einem Parametersatz zu modellieren und dabei die unterschiedliche Verarbeitung der Stimuli bei Klassifikations- und Inferenzaufgaben zu berücksichtigen, wurde für jede der beiden Aufgabentypen eine eigene URM Instanz mit eigenen Gammaverteilungen für die Konzentrationsparameter α und γ verwendet. Dieses Vorgehen wurde gewählt, da nach den Ergebnissen von Yamauchi und Markman (1998) bzw. Yamauchi et al. (2002) bei Inferenzaufgaben im Wesentlichen eine Prototypenstruktur und bei Klassifikationsaufgaben

vermehrt eine Exemplar-orientierte Struktur der Kategorien aufgebaut wird. Durch aufgabenspezifische URM-Instanzen lässt sich dann ein individuelles Fitting der Konzentrationsparameter erreichen, sodass für Inferenzaufgaben eine Partition mit wenigen Clustern (= Prototypenstruktur) und für Klassifikationsaufgaben eine Partition mit vielen Clustern (= Exemplarstruktur) gebildet werden kann. Um die Anzahl der benötigten Blöcke für einen gegebenen Parametersatz zu bestimmen, wurde in analoger Weise zum Vorgehen in Experiment 2 der erwartete letzte Block anhand der Verteilung der Stoppwahrscheinlichkeiten berechnet.

Hierbei zeigte sich für das URM eine mittlere SSE der 200 Samples aus $SSE\left(\Omega_{URM}^{\theta}\right)$ von 450.15 bei einer Varianz von 11.68. Die mittlere Prognoserealisierung lag bei 27.6 und 10.3 benötigten Blöcken in der Inferenz- und der Klassifikationsbedingung in Experiment 3 und bei 27.5 und 10.4 Blöcken in Experiment 4. Wie zu erkennen ist, konnte das URM das Phänomen der konträren Aufgabenschwierigkeiten bei linearer und nichtlinearer Kategorienstruktur somit qualitativ nicht korrekt prognostizieren.

SUSTAIN erreicht nach Love et al. (2004) mit einer SSE von 3.69 eine nahezu perfekte (und qualitativ korrekte) Vorhersage der VP-Daten. In einer erneuten Parametersuche konnte dieses Fitting jedoch nicht repliziert werden. So war die mittlere Prognoserealisierung von SUSTAIN mit 16.5 und 11.3 benötigten Blöcken in der Inferenz- und Klassifikationsbedingung in Experiment 3 sowie 17.3 und 13.0 Blöcke in Experiment 4 ebenfalls qualitativ nicht korrekt, allerdings lag die mittlere SSE der 200 Samples aus $SSE\left(\Omega_{SUS}^{\theta}\right)$ mit 211.09 deutlich unter der des URM. Die Varianz der SSE-Stichprobe war mit 1423.80 hingegen deutlich größer. Für den geschätzten relativen Treatment Effekt wurde ein $\hat{p}$ = 0.001 mit einem *p*-Wert < 0.001 ermittelt, welcher zunächst für eine klar niedrigere SSE von SUSTAIN spricht, jedoch ohne qualitativ korrekte Vorhersage

Tabelle 12. Ermittelte Parameter für das URM für die Experimente 3 und 4 für Inferenz- ($URM_{Inference}$) und Klassifikationsaufgaben ($URM_{Classification}$).

	β	γ_0^{α}	γ_1^{α}	γ_0^{γ}	γ_1^{γ}
$URM_{Inference}$	0.98561	0.0044036	0.050683	0.015218	0.0086207
$URM_{Classification}$	0.98561	1.1961	1.9309	2.2401	0.39593

Tabelle 13. Ursprüngliche und in einer wiederholten Modellierung ermittelte Parameter für SUSTAIN für die Experimente 3 und 4.

	r	β	*m*	η	λ_{label}
SUSTAIN	1.016924	3.97491	6.514972	0.1150532	12.80691
$SUSTAIN^{Rep}$	1.65037	11.1111	1.45835	1	0.733429

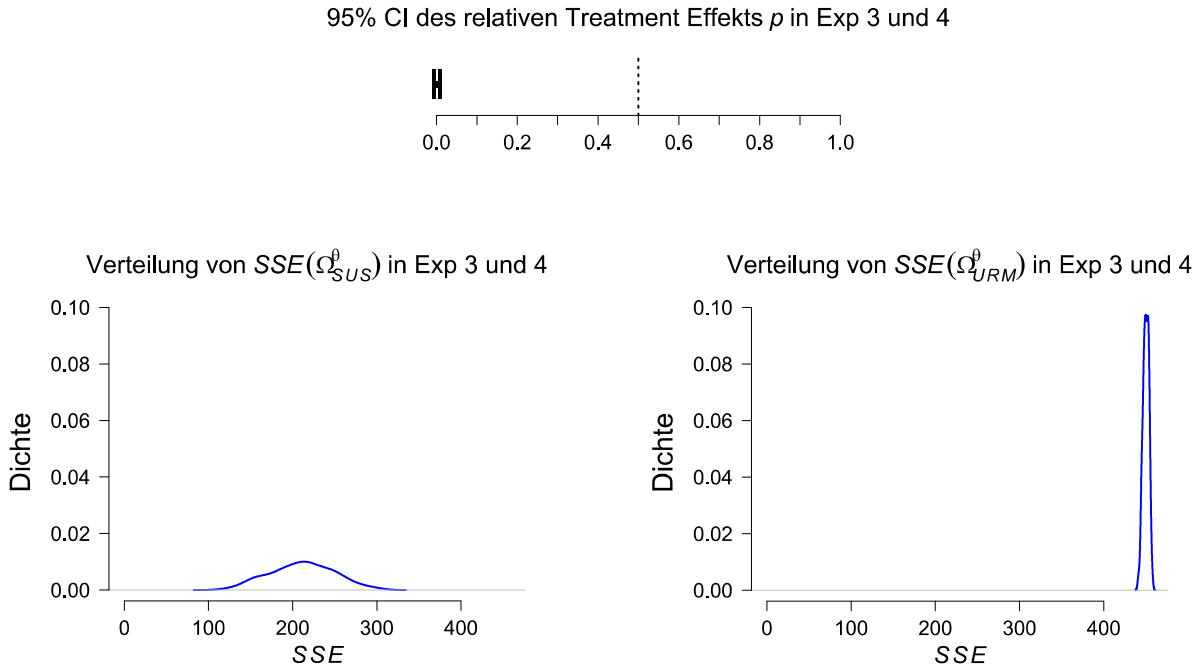


Abbildung 14. (Oben) Relativer Treatment Effekt von URM und SUSTAIN in den Experimenten 3 und 4. (Unten) Anhand der Stichproben geschätzte Dichte der SSE-Verteilungen von URM und SUSTAIN in den Experimenten 3 und 4.

höchstens Rückschlüsse auf die Flexibilität des Modells erlaubt. In Abbildung 14 sind die geschätzten Dichten der SSE-Verteilungen und das Konfidenzintervall von $\hat{p}$ einsehbar.
Eine Modellflexibilitätsanalyse ergab für das URM ein $\phi_{3,4}^{URM} = 0.00374$ und für SUSTAIN ein $\phi_{3,4}^{SUS} = 0.01730$, was somit tatsächlich auf ein mehr als den Faktor 4 flexibleres SUSTAIN hindeutet. Die ermittelten Parameter für das URM finden sich in Tabelle 12, die alten und neuen Parameter für SUSTAIN in Tabelle 13.

3.5 Experimente 5 und 6: Inferenzakkuratheit bei niedrig und hoch interkorrelierenden Stimulusattributen

Im Verlauf des Lebens begegnen Menschen nicht ausschließlich Lernherausforderungen, bei welchen ein wohlwollender Unterstützer korrektes bzw. überhaupt ein Feedback zur eigenen Klassifikation eines neuartigen Stimulus bzw. Ereignis abgibt. Tatsächlich findet der größere Anteil

der Wissensakquisition eines Menschen, anders als bei den bisher betrachteten Experimenten, unsupervidiert statt (Gureckis & Love, 2003; Love, 2002).

Diese Tatsache und die Feststellung von Billman und Knutson (1996), dass zwischen allen von Menschen potentiell konstruierbaren Kategorien Unterschiede in der Kohärenz, Natürlichkeit und Informativität und somit in ihrer Nützlichkeit bestehen, ließen die Autoren vermuten, dass der unsupervidierte menschliche Kategorisierungsprozess einen Bias zur Formung von sogenannten „guten“ Kategorien aufweisen könnte. So seien Kategorien wie ‚Fahrrad‘ oder ‚Vogel‘ intuitiv nützlich, Kategorien wie ‚Dinge die rot sind‘ oder ‚Dinge die größer sind als eine Brotbox‘ hingegen weniger sinnvoll. Eine zentrale Frage deren Beantwortung weitere Rückschlüsse auf den menschlichen Kategorisierungsmechanismus zulassen sollte, ist deshalb nach Billman und Knutson (1996) die nach der gemeinsamen Charakteristik von „guten“ Kategorien.

Die Autoren führen dabei an, dass für die Beantwortung prinzipiell nur zwei mögliche Erklärungsvarianten in Frage kommen können: Ähnlichkeit und Theorie. Nach Ersterer sind Kategorien „gut“, wenn ihre Mitglieder bezüglich eines formalen Merkmalsvergleichs untereinander ähnlich sind. Nach der zweiten Sichtweise lassen sich hingegen Kategorien als „gut“ betrachten, wenn ihre Mitglieder einen zentralen gegenseitigen Zusammenhang in Bezug auf eine übergeordnete Theorie aufweisen. Billman und Knutson (1996) präsentieren schließlich eine mögliche, beider Erklärungsvarianten genügende Charakterisierung von „guten“ Kategorien anhand des Korrelationale-Systeme-Ansatz. Dieser besagt, dass Kohärenz und Struktur in Kategorien durch Zusammenfassung von Beobachtungen entsteht, sodass die kodierten Informationen auf unterschiedlichen Abstraktionsebenen (die Systeme) möglichst korrelieren (Barsalou & Billman, 1989; Billman & Knutson, 1996). Beispielsweise wird eine hohe Korrelation auf der Attributsebene und der Attributswerteebene angestrebt, sodass möglichst Beobachtungen zu Kategorien zusammengefasst werden, welche viele gemeinsame Attribute aufweisen und die mittlere Korrelationsstärke von Attributswerten bezüglich Werte der anderen Attribute maximal ist. Die Höhe der Korrelation einer Kategorie auf der Attributsebene wird dabei als *Attribute Systematicity* und die auf der Attributswerteebene als *Value Systematicity* bezeichnet (Barsalou & Billman, 1989). Um diesen Ansatz experimentell zu überprüfen, wurde Letztere von Billman und Knutson (1996) zwischen zwei Bedingungen variiert und die Auswirkung auf die Güte der Kategorienakquisition bei unsupervidiertem Training erhoben. Hierzu verwendeten die Autoren Stimuli, welche außerirdische Lebewesen in ihrem jeweiligen Habitat darstellten und sich entlang der sieben ternären Dimensionen Kopf, Körper, Textur, Schwanz, Beine, Habitat und Tageszeit unterschieden. Ein schematisches Beispiel eines Stimulus ist in Abbildung 15 dargestellt. 120 VP sahen in ihrem Experiment vier Blöcke à 27 solcher Stimuli, welche in einer zufälligen Reihenfolge präsentiert

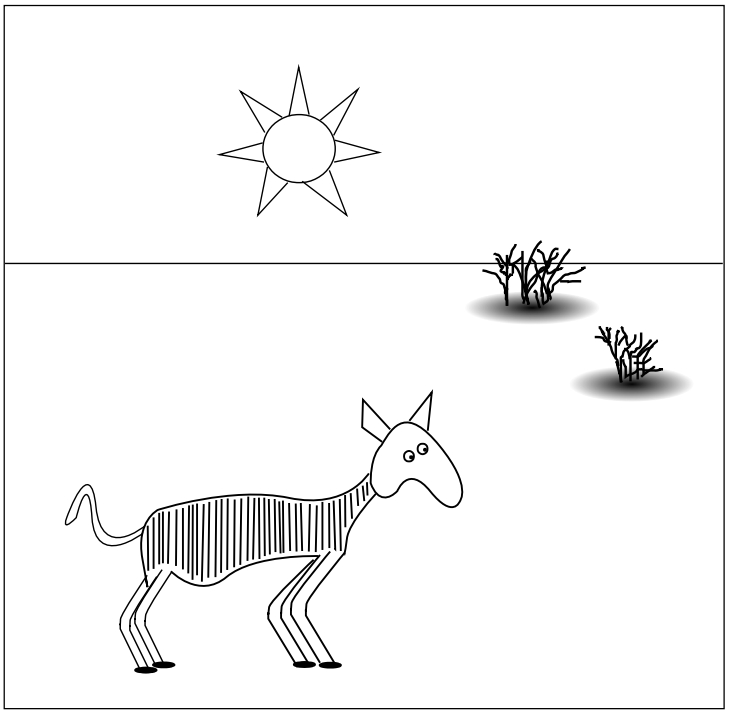


Abbildung 15. Schematisches Beispiel eines Stimulus aus Billman und Knutson (1996).

wurden und von der VP beliebig lange inspiziert werden konnten. Dabei waren jeweils 60 VP der isolierten und der strukturierten Bedingung zugewiesen. Bei Ersterer korrelierten immer zwei, bei Letzterer immer vier Attribute der Stimuli perfekt miteinander, während alle übrigen Attribute völlig unkorreliert waren. Tabelle 14 zeigt die logische Stimuluskonfiguration der beiden Bedingungen.

Die letztendlich korrelierenden Attribute wurde zwischen den VP der isolierten Bedingung derart gewählt, dass bei jeweils 10 VP eine von sechs vorab definierten Korrelationsregeln, z.B. die Korrelation der Attribute Kopf und Zeit, vorlag. In der strukturierten Bedingung erfolgte die Zusammenstellung der korrelierenden Attribute indem jede der sechs Regeln aus der isolierten Bedingung in fünf verschiedenen Konstellationen mit zwei weiteren Attributen kombiniert wurde.

Tabelle 14. Logische Stimuluskonfiguration des zweiten Experiments von Billman und Knutson (1996).

Bedingung	Isoliert	Strukturiert
Stimulus-typen	11XXXXX	1111XXX
	22XXXXX	2222XXX
	33XXXXX	3333XXX

Bemerkung. Von jedem Stimulustyp wurden neun Instanzen gezeigt, wobei die X jeweils durch Werte ersetzt waren, sodass die zugehörigen Attribute mit keinen anderen Attributen korrelierten.

Hieraus ergab sich schließlich eine individuelle Kombination von korrelierenden Attributen für jeweils 2 von 60 VP innerhalb der strukturierten Bedingung.

Nach der vier Blöcke umfassenden Trainingsphase folgte die Testphase, in welcher jede VP einen 45 Trials umfassenden Forced-Choice-Test zu absolvieren hatte. In diesem wurden zwei Stimuli nebeneinander präsentiert und die VP musste entscheiden, welcher der Stimuli nach ihrer Auffassung am besten zu den Stimuli aus der Trainingsphase passt. VP aus der strukturierten Bedingung erhielten in diesem Test Stimuli, welche dieselbe Korrelationsstruktur aufwiesen wie in der Trainingsphase. VP der isolierten Bedingung bearbeiteten Tests, für deren Stimuli eine Konstellation korrelierender Attribute aus der strukturierten Bedingung ausgewählt wurde, in welcher ebenfalls die Korrelation aus der Trainingsphase der VP vorkam. Die jeweilige Auswahl erfolgt derart, dass immer genau zwei VP der isolierten Bedingung Stimuli mit derselben Konstellation korrelierender Attribute im Test bearbeiteten.

Geprüft wurden im Forced-Choice-Test jeder VP der Lernerfolg bezüglich der bzw. einer der erlernten Korrelationsregeln, sodass insgesamt von vier VP der Lernerfolg für einer der sechs vorab definierten Korrelationsregeln erhoben wurde. Lediglich 15 der 45 Items des Forced-Choice-Test prüften dabei die Korrelationsregel von Interesse während die übrigen 30 Items ausschließlich als Füllitems fungierten. Damit in jedem Trial der Testphase ausschließlich der Lernerfolg von genau

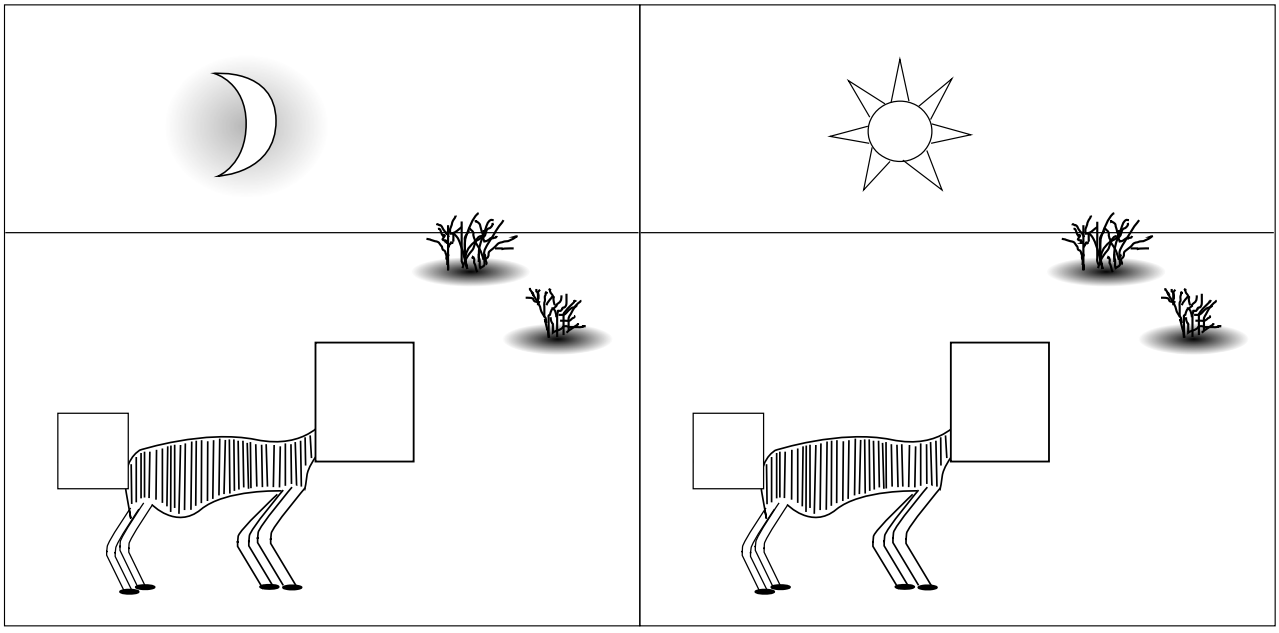


Abbildung 16. Schematisches Beispiel eines Trials im Forced-Choice-Test zur Überprüfung der Trainingserfolgs aus Billman und Knutson (1996).

einer Korrelationsregel geprüft werden konnte, wurden die übrigen korrelierenden und somit ebenfalls informativen Attribute ausgeblendet. Abbildung 16 zeigt einen schematischen Beispieltrial des Forced-Choice- Tests.

Bei der Auswertung der Ergebnisse des Experiments zeigt sich, dass VP der strukturierten Bedingung einen höheren Lernerfolg erzielten als VP der isolierten Bedingung. Erstere Gruppe beantworteten 73% der Forced-Choice-Aufgaben korrekt, während VP der letzteren Gruppe nur 62% korrekt lösten. Dieser Befund einzeln betrachtet, stützt noch nicht die Annahme von Billman und Knutson (1996), dass das menschliche Kategorisierungssystem einen Bias hinsichtlich korrelierender Systeme aufweist. So besteht die Möglichkeit, dass nicht die Interkorrelation ursächlich war für den höheren Lernerfolg, sondern die reine Anzahl an Korrelationen, die Anzahl der in Korrelationen beteiligten Attribute oder die durch die Korrelationsregeln erzeugte Redundanz (Billman & Knutson, 1996).

Aus diesem Grund führten die Autoren ein weiteres Experiment durch, in welchem sie zwischen zwei Bedingungen die Anzahl der in Korrelationen beteiligten Attribute und damit einhergehend die Redundanz in den Stimulusinformationen variierten, die Gesamtanzahl der Korrelationen jedoch konstant hielten. Die strukturierte Bedingung folgte dabei dem aus dem vorhergehenden Experiment bekannten Prinzip von interkorrelierenden Attributen. Dieses Mal wurde jedoch die Anzahl der beteiligten Attribute auf drei reduziert. In der als orthogonal bezeichneten Bedingung verteilten sich stattdessen die drei Korrelationen nicht auf drei Attribute sondern auf drei distinkte Attributspaare (siehe Tabelle 15). Hierdurch stieg gegenüber der strukturierten Bedingung sowohl die Anzahl der an Korrelationen beteiligten Attribute, als auch die Redundanz in den Informationen. Billman und Knutson (1996) erwarteten nun weiterhin einen Lernvorteil in der strukturierten

Tabelle 15. Logische Stimuluskonfiguration des dritten Experiments von Billman und Knutson (1996).

Bedingung	Orthogonal			Strukturiert
Stimulustypen	111111X	221111X	331111X	111XXXX
	111122X	221122X	331122X	222XXXX
	111133X	221133X	331133X	333XXXX
	112211X	222211X	332211X	
	112222X	222222X	332222X	
	112233X	222233X	332233X	
	113311X	223311X	333311X	
	113322X	223322X	333322X	
	113333X	223333X	333333X	

Bemerkung. Von jedem Stimulustyp wurden drei (orthogonale Bedingung) bzw. neun Instanzen (strukturierte Bedingung) gezeigt, wobei die X jeweils durch Werte ersetzt waren, sodass die zugehörigen Attribute mit keinen anderen Attributen korrelierten.

Bedingung. Sie trainierten daher jeweils 24 VP pro Bedingung mit denselben Stimuli aus dem letzten Experiment und prüften in einem anschließenden Forced-Choice-Test mit 54 Trials den Lernerfolg bei insgesamt neun vorab ausgewählten Korrelationsregeln.

Eine ausbalancierte Aufteilung der VP zu unterschiedlichen Korrelationskonfigurationen in der jeweiligen Bedingung sorgte dafür, dass über das gesamte Experiment der Lernerfolg bei neun verschiedenen Korrelationsregeln erhoben werden konnte, jede VP im Forced-Choice-Test jedoch nur bezüglich der in der Trainingsphase beobachteten drei Korrelationen geprüft wurde. Das Prinzip des Forced-Choice-Test entsprach dem des Tests aus dem letzten Experiment. Demnach waren auch diesmal in jedem Trial zwei Attribute der Stimuli ausgeblendet, sodass nur der Lernerfolg bezüglich einer ausgewählten Korrelationsregel entscheidend war. In den Forced-Choice-Tests der strukturierten Bedingung handelte es sich bei den ausgeblendeten Attributen um das dritte prädiktive Attribut der interkorrelierenden Attribute und um ein weiteres zufällig ausgewähltes irrelevantes Attribut. Im Test der orthogonalen Bedingung wurde jeweils eines der Attribute aus den übrigen nicht erfragten Korrelationspaaren maskiert.

Nach Auswertung der Ergebnisse zeigte sich auch in diesem Experiment ein größerer Lernerfolg in der strukturierten Bedingung (Billman & Knutson, 1996). So wählten VP dieser Bedingung im Forced-Choice-Test in 77% der Fälle den korrekten Stimulus, in der orthogonalen Bedingung hingegen nur 66%. Da sowohl die Anzahl der in Korrelationen beteiligten Attribute, als auch die Redundanz in den Stimulusinformationen in der orthogonalen Bedingung höher war als in der strukturierten Bedingung lässt sich die Ursache für den höheren Lernerfolg entsprechend der Annahme von Billman und Knutson (1996) in der hohen Interkorrelation der Attribute der strukturierten Bedingung vermuten. Die Befunde deuten somit auf einen Verarbeitungsbias des menschlichen Kategorisierungssystems bezüglich korrelationaler Systeme hin.

3.5.1 Modellierungsergebnisse für die Experimente 5 und 6

Da beide Experimente Stimuli mit ternären Attributen verwendeten, war die Modellierung mit dem klassischen URM von Griffiths et al. (2007) nicht möglich. Es wurde deshalb URM^{Cat} aus Experiment 2 verwendet, welches Stimuli mit beliebig vielen Attributsausprägungen erlaubt. Es konnte dabei ein Parametersatz identifiziert werden, welcher eine durchaus beachtliche mittlere SSE von 0.004 mit einer Varianz von 1.7e-05 in der Stichprobe der 200 Samples aus $SSE\left(\Omega^{\theta}_{URM^{Cat}}\right)$

aufwies. Die mittlere Prognoserealisierung lag bei 0.76 und 0.64 durchschnittlicher Auswahlkorrektheit in der strukturierten und isolierten Bedingung bei Experiment 5 und bei 0.76 und 0.71 in der strukturierten und orthogonalen Bedingung bei Experiment 6.

Love et al. (2004) berichten eine Prognose von SUSTAIN, welche eine leicht schlechtere SSE als die des URM^{Cat} aufweist. Dieser Befund konnte in einer erneuten Modellierung nicht repliziert werden. Tatsächlich unterbietet SUSTAIN nochmals die für das URM^{Cat} ermittelte mittlere SSE mit dem besten gefundenen Parametersatz. So lag die mittlere SSE der 200 Samples aus $SSE\left(\Omega_{SUS}^{\theta}\right)$ bei 0.00057, wobei die Varianz nur 1.7e-07 betrug. Die mittlere Prognoserealisierung wies mit 0.75 und 0.62 bei der strukturierten und isolierten Bedingung in Experiment 5 und mit 0.77 und 0.65 bei der strukturierten und orthogonalen Bedingung in Experiment 6 beinahe einer Punktlandung auf. Die Schätzung des relativen Treatment Effekts über die Brunner und Munzel (2000) Methode ergab ein $\hat{p}$ =0.007 mit einem *p*-Wert < 0.0001, welcher eine klar niedrigere SSE für SUSTAIN bestätigt. In Abbildung 17 sind das zugehörige Konfidenzintervall sowie die geschätzten Dichten der SSE-Verteilungen dargestellt.

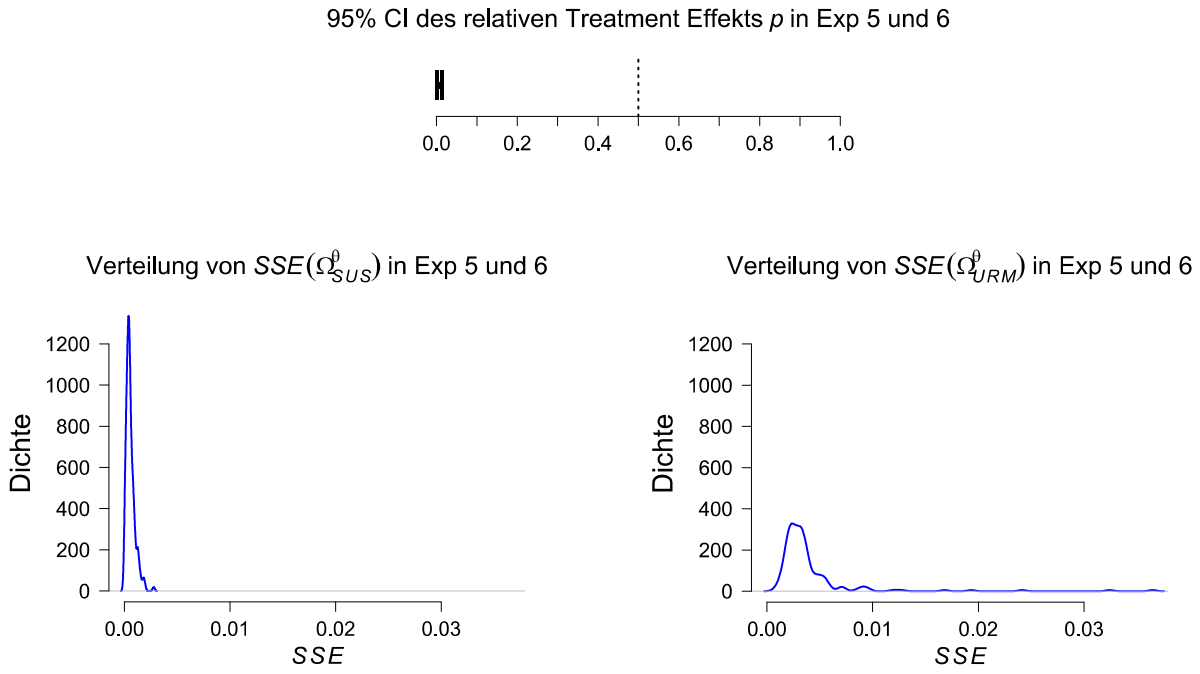


Abbildung 17. (Oben) Relativer Treatment Effekt von URM^{Cat} und SUSTAIN in den Experimenten 5 und 6. (Unten) Anhand der Stichproben geschätzte Dichte der SSE-Verteilungen von URM^{Cat} und SUSTAIN in den Experimenten 5 und 6.

Tabelle 16. Ermittelte Parameter für das URM ^{Cat} für die Experimente 5 und 6.

	β	γ_0^{α}	γ_1^{α}	γ_0^{γ}	γ_1^{γ}
URM^{Cat}	0.038091	1.7238	0.012406	0.39742	0.013002

Tabelle 17. Ursprüngliche und in einer wiederholten Modellierung ermittelte Parameter für SUSTAIN für die Experimente 5 und 6.

	r	β	m	η
SUSTAIN	9.998779	6.396300	1.977312	0.096564
SUSTAIN^{Rep}	0.26569	3.6888	3.2668	1

Eine Modellflexibilitätsanalyse ergab für das URM ein $\phi_{5,6}^{URM^{Cat}} = 0.00204$ und für SUSTAIN ein $\phi_{5,6}^{SUS} = 0.00639$. Die ermittelten Φ deuten somit erneut auf ein, hier um den Faktor > 3, flexibleres SUSTAIN hin. Die gefundenen besten Parameter für beide Modelle sind in Tabelle 16 und 17 aufgelistet.

3.6 Experiment 7: Unsupervidiertes Sortieren von Stimuli in zwei Gruppen

Die Experimente 5 und 6 legen den Schluss nahe, dass Menschen dazu neigen Kategorien zu konstruieren, deren Mitglieder eine Struktur interkorrelierender Attributswerte aufweisen und somit Stimuli zusammenfassen, welche in den gleichen Attributen gleiche Ausprägungen zeigen. Das ist jedoch nicht immer der Fall. So konnten Medin et al. (1987) belegen, dass Menschen bei einer unsupervidierten Aufgabe, in der sie zehn Stimuli in genau zwei Kategorien sortieren sollten, eine eindimensionale Sortierungsregel verfolgen, selbst wenn die logische Attributsstruktur der Stimuli eine Organisation nach Familienähnlichkeit erlaubt. Medin et al. (1987) legten hierzu einer nicht näher genannten Anzahl an VP zehn Zeichnungen von Cartoon-ähnlichen Tieren vor, welche sich auf den vier Dimensionen Kopfform (eckig oder rund), Anzahl an Beinen (vier oder acht), Körpermarkierungen (Punkte oder Streifen) und Schwanzlänge (kurz oder lang) unterschieden. Dabei war die Auswahl der Attributswerte für jeden Stimulus so gewählt, dass zwei gleich große Gruppen mit jeweils hoher Familienähnlichkeit zwischen ihren Gruppenmitgliedern vorlagen. Die logische Attributsstruktur innerhalb der zwei Gruppen ist in Tabelle 18 dargestellt.

Die Konkretisierung dieser Struktur wurde zwischen den VP randomisiert, sodass Stimuli mit einer logischen 1 auf einer Dimension, z.B. der Kopfform, für VP *A* als eckiger Kopf für VP *B* jedoch als

Tabelle 18. Logische Struktur der Stimuli aus Medin et al. (1987).

	Gruppe 1				Gruppe 2			
	Kopf	Beine	Körper	Schwanz	Kopf	Beine	Körper	Schwanz
Stimuli	1	1	1	1	0	0	0	0
	1	1	1	0	0	0	0	1
	1	1	0	1	0	0	1	0
	1	0	1	1	0	1	0	0
	0	1	1	1	1	0	0	0

runder Kopf realisiert wurde. Eine beispielhafte schematische Repräsentation der zehn Tiere ist in Abbildung 18 dargestellt und entsprechend ihrer Familienähnlichkeit in zwei Gruppen sortiert.

Jede VP hatten nun die Aufgabe die Tiere, welche sie in einer zufälligen Reihenfolge erhielt, nach Belieben in zwei gleichgroße Gruppen einzuteilen. Dabei wurde betont, dass die Forschergruppe daran interessiert sei, welche Partitionierung aus der Sicht der VP am natürlichsten sei, es jedoch nicht die eine korrekte Lösung sondern viele Möglichkeiten gebe zwei gleichgroße Gruppen zu bilden und es der VP überlassen sei, für welche dieser Varianten sie sich entscheide.

Die Autoren der Studie erwarteten, dass eine Anordnung nach Familienähnlichkeit eine häufige Strategie sein würde. Tatsächlich wählte jedoch niemand dieses Vorgehen. Jede VP bevorzugte stattdessen eine Sortierung der Tiere entlang einer individuell gewählten Dimension. Tabelle 19 veranschaulicht beispielhaft eine solche Sortierung entlang der Dimension Kopf.

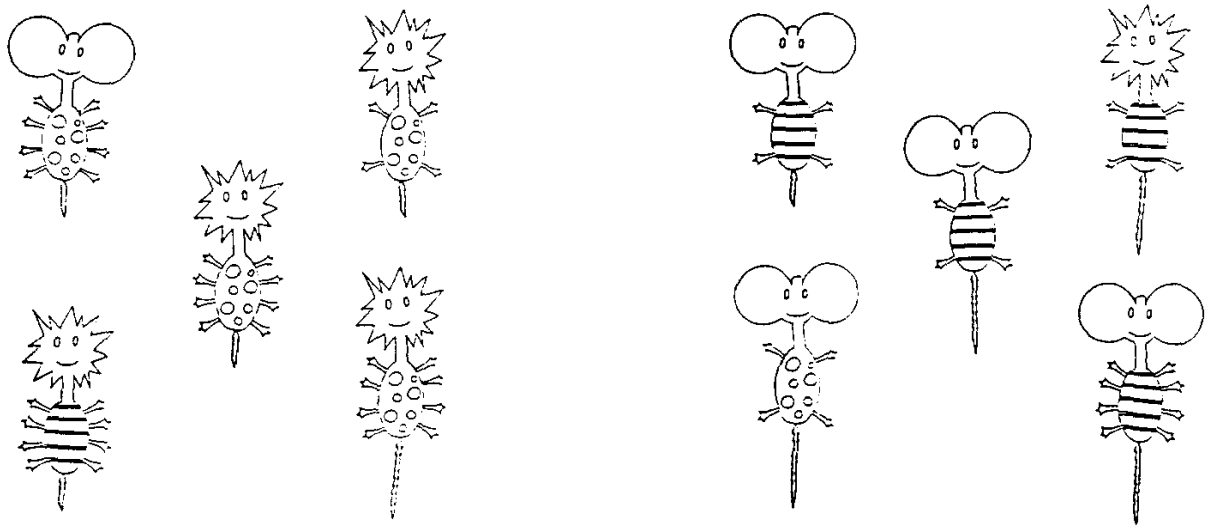


Abbildung 18. Darstellung der zehn Stimuli aus Medin et al. (1987). Nachgedruckt mit Genehmigung von Elsevier.

Tabelle 19. Beispiel für eine eindimensionale Sortierung der Stimuli aus Medin et al. (1987).

	Gruppe 1				Gruppe 2			
	Kopf	Beine	Körper	Schwanz	Kopf	Beine	Körper	Schwanz
Stimuli	1	1	1	1	0	0	0	0
	1	1	1	0	0	0	0	1
	1	1	0	1	0	0	1	0
	1	0	1	1	0	1	0	0
	1	0	0	0	0	1	1	1

3.6.1 Modellierungsergebnisse für Experiment 7

Um das Experiment mit dem URM zu modellieren, wurde das Modell zunächst modifiziert, sodass nur maximal zwei Cluster gebildet werden können. Anschließend konnte eine analoge Modellierung zu der von Love et al. (2004) mit SUSTAIN durchgeführt werden, bei welcher die zeitlich uneingeschränkte Stimulusbegutachtung der VP über vorausgehende 10 Trainingsblöcke simuliert und anschließend für einen gegebenen Parametersatz die Partition der zehn Stimuli prognostiziert wurde.

Die für die Parameteroptimierung beim URM verwendete Fitnessfunktion bestimmte dabei eine SSE zwischen der Lösung der VP aus Medin et al. (1987) und der mittleren Partition der 30 Samples aus dem URM in Form einer quadrierten Partitionsdistanz für den jeweils gegebenen Parametersatz. Eine richtige (und beste) Lösung wäre demnach eine quadrierte Partitionsdistanz von 0.

Es zeigte sich, dass mit dem besten gefundenen Parametersatz beinahe immer die Gruppierung aus Tabelle 18 bevorzugt wurde, welche eine Partitionsdistanz von 2 aufweist. Die 200 Samples umfassende Stichprobe aus $SSE\left(\Omega_{URM}^{\theta}\right)$ hatte einen Mittelwert von 3.98 und eine Varianz von 0.26. Die mittle Prognoserealisierung lag entsprechend bei einer Partitionsdistanz von 2, was einer quadratischen Abweichung von 4 entspricht.

Love et al. (2004) berichten, dass SUSTAIN die Befunde aus Medin et al. (1987) korrekt vorhersagte und zwar mit denselben Parametern aus den Experimenten 5 und 6. Weder mit diesen Parametern noch mit dem über GA besten gefundenen Parametersatz konnte das Resultat aus Love et al. (2004) repliziert werden. Tatsächlich sind die Vorhersagen von SUSTAIN weitgehend übereinstimmend mit der des URM. So weist die Stichprobe der 200 Samples aus $SSE\left(\Omega_{SUS}^{\theta}\right)$ einen Mittelwert von 3.91 und ebenfalls eine Varianz von 0.26 auf. Die mittlere Prognoserealisierung liegt

dabei identisch zu der des URM bei einen Partititonsdistanz von 2, wobei die Berechnung dieser Distanz mit der gleichen Methode wie beim URM erfolgte. Bei einem Vergleich der Stichproben mit der Brunner und Munzel (2000) Methode wurde deshalb ein geschätzter relativer Treatment Effekt von $\hat{p}$ = 0.49 mit einem p-Wert = 0.204 ermittelt. Die SSEs beider Modelle weichen somit nicht signifikant voneinander ab. Abbildung 19 zeigt das Konfidenzintervall der Schätzung sowie die geschätzten Dichten der SSE-Verteilungen. Die für Experiment 7 besten gefundenen Parameter jedes der Modelle sowie die alten Parameter aus Love et al. (2004) für SUSTAIN finden sich in Tabelle 20 bzw. 21. Eine Modellflexibilitätsanalyse ergab für das URM ein $\phi_7^{URM} = 4.69\text{e-}05$ und für SUSTAIN ein $\phi_7^{SUS} = 2.34\text{e-}05$, was somit für ein in diesem Experiment doppelt so flexibles URM spricht.

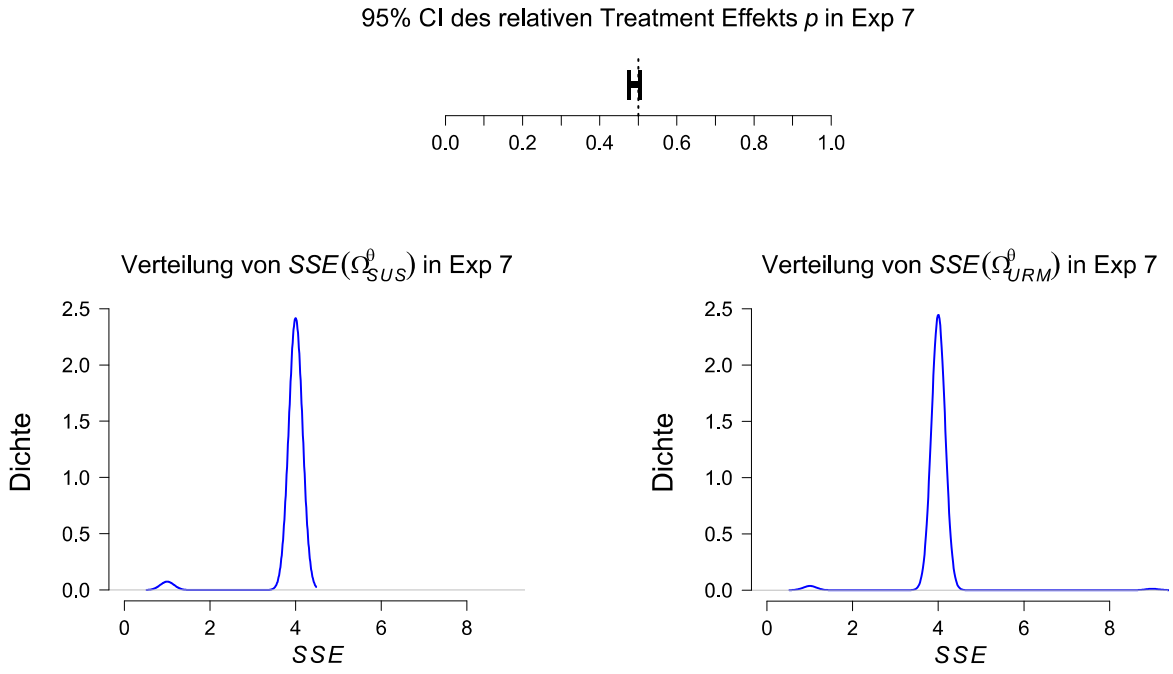


Abbildung 19. (Oben) Relativer Treatment Effekt von URM und SUSTAIN in Experiment 7. (Unten) Anhand der Stichproben geschätzte Dichte der SSE-Verteilungen von URM und SUSTAIN in Experiment 7.

Tabelle 20. Ermittelte Parameter für das URM für Experiment 7.

	β	γ_0^{α}	γ_1^{α}	γ_0^{γ}	γ_1^{γ}
URM	0.32393	3.5876	0.41621	0.71494	0.48417

Tabelle 21. Ursprüngliche und in einer wiederholten Modellierung ermittelte Parameter für SUSTAIN für Experiment 7.

	r	β	m	η
SUSTAIN	9.998779	6.396300	1.977312	0.096564
$SUSTAIN^{Rep}$	1.1836	0.03082	1.2521	0.00011162

4 Diskussion

Wie gut eignet sich das URM als ein Modell menschlicher Kategorisierung und somit als ein Verfahren für am Menschen orientierte kognitive Roboter? Das war die zentrale Fragestellung dieser Arbeit. Hierzu wurde das bisher kaum untersuchte Modell an sieben prominenten Experimenten aus der Kategorisierungsforschung evaluiert, welche derzeit besondere Herausforderungen für aktuelle Modelle darstellen. Dabei wurden Modellverbesserungen vorgestellt, welche die generelle und erfolgreiche Modellierung eines Teils dieser Experimente erst ermöglichten. Dies waren die Integration eines Aufmerksamkeitsmechanismus über ein Gradientenverfahren sowie die Erweiterung des Beta-Bernoulli-Priors von Griffiths et al. (2007) zu einem Dirichlet-Categorical-Prior.

Erstere Verbesserung ermöglichte es Experiment 1 erfolgreich zu modellieren, in welchem die charakteristischen Lernkurven bei sechs Kategorisierungsregeln unterschiedlichen Schwierigkeitsgrades vorhergesagt werden sollten. Bereits Nosofsky et al. (1994) vermuteten beim zum URM verwandten RMC, dass der fehlende Aufmerksamkeitsmechanismus eine mögliche Ursache der schlechten Vorhersagekraft des RMC in diesem Experiment ist, allerdings präsentierten sie keinen Implementationsvorschlag. In dieser Arbeit wurde dies indirekt nachgeholt.

Die zweite Verbesserung betraf die Eingeschränktheit des bisherigen Priors von Griffiths et al. (2007), welcher nur Modellierungen von Stimulusdimensionen mit zwei möglichen Ausprägungen erlaubt. Durch die Generalisierung des alten Priors zu einem Dirichlet-Categorical-Prior sind stattdessen Stimulusdimensionen mit beliebig vielen Ausprägungen möglich, wodurch die Experimente 2, 5 und 6 modelliert werden konnten.

Die durchgeführte Evaluation des URM stellt eine erste ausführlichere Fundierung des Modells innerhalb der Kategorisierungsforschung dar und somit die erste Analyse dessen Eignung als psychologisch orientiertes Verfahren für am Menschen orientierte Roboter. Im Folgenden sollen die Modellierungsergebnisse näher beleuchtet, sowie die theoretischen und technischen Aspekte insbesondere auch hinsichtlich der Anwendung im Robotikkontext besprochen werden.

4.1 Bewertung der Modellierungsergebnisse

Die Modellierungsgüte des URM in den sieben ausgewählten Experimenten lässt sich als überwiegend positiv bewerten. So zeigte sich, nach Integration der Verbesserungen, bei vier von sieben Experimenten eine qualitativ korrekte Vorhersage der Befunde. Im Vergleich zu SUSTAIN, welches die Befunde laut Love et al. (2004) sehr gut prognostizieren könne, zeigte sich zudem nahezu durchgehend eine niedrigere bis deutlich niedrigere und somit günstigere Modellierungsflexibilität bei einem ähnlich guten Fit. Tatsächlich konnten die Ergebnisse aus Love et al. (2004) für SUSTAIN nicht repliziert werden. Die Ursache hierfür bleibt unklar, da bis auf den Modellierungscode für Experiment 1 von Gureckis (2014) alle anderen Codes nicht mehr existieren (persönliche Kommunikation mit Todd Gureckis, 14. April 2016). Da im Code des ersten Experiments für SUSTAIN eine leicht Abweichung von der originalen Experimentalprozedur von Nosofsky et al. (1994) gefunden wurde, ist zumindest in Betracht zu ziehen, dass auch in den übrigen nicht mehr existenten Codes Prozedurabweichungen eine Ursache für die unterschiedlichen Ergebnisse darstellen könnten.

Unabhängig davon deuten die Ergebnisse in dieser Arbeit darauf hin, dass SUSTAIN vor allem durch seine Flexibilität Befunde gut erklären kann. Das ist hingegen keine gewünschte Eigenschaft wenn auf potentielle Kategorisierungsmechanismen beim Menschen geschlossen werden soll. So zeichnet ein gutes Modell aus, dass auch bei Variation der freien Parameter nicht jedes beliebige Verhalten prognostiziert werden kann, sondern eine generelle Tendenz vorliegt nur eine Art von Verhalten zu prognostizieren. Bis auf Experiment 7 zeigt sich in allen anderen Experimenten eine teilweise deutlich größere Flexibilität (bis Faktor 10) bei SUSTAIN. Es liegt daher nahe die Ursache der niedrigeren SSE dieses Modells in Experimenten, in welchen beide Modelle sehr gut abschneiden, vorrangig der Flexibilität von SUSTAIN zuzuschreiben. So ist beispielsweise die SSE dieses Modells von 0.00057 in den Experimenten 5 und 6 beachtlich und nur ein Siebtel der SSE des URM mit bereits niedrigen 0.004. Jedoch lässt eine derart niedrige SSE vielmehr eine Tendenz zum Noise-Fitting vermuten.

Erstaunlich ist hingegen, dass in Experiment 2 diese Anpassungsfähigkeit von SUSTAIN zu keiner besseren Modellierung der Befunde führte. Sowohl in Bezug auf die SSE als auch auf die um Faktor $\frac{9}{10}$ geringere Modellflexibilität in diesem Experiment konnte sich das URM klar durchsetzen.

Neben diesen Erfolgen konnten jedoch auch für beide Modelle nicht bzw. schwer modellierbare Experimente identifiziert werden. Hierbei handelt es sich um die Experimente 3 und 4, in welchen die Berücksichtigung des Nutzungskontexts bei der Kodierung der Kategorienrepräsentation

relevant ist und Experiment 7, in welchem eine eindimensionale Sortierung von Stimuli in zwei gleichgroße Gruppen vorherzusagen ist. In den ersten beiden der genannten Experimente prognostizieren sowohl das URM als auch SUSTAIN eine größere Lernschwierigkeit in der Inferenz- gegenüber der Klassifikationsbedingung, welche nach den Befunden von Yamauchi und Kollegen jedoch nur bei nicht linear separablen Kategorien auftritt. Auch beim Random-Search zur Berechnung der Konstante ϕ zeigte sich bei beiden Modellen bei den ermittelten einzigartigen Prognosen eine höhere durchschnittliche Anzahl der benötigten Trainingsblöcke in der Inferenz- gegenüber der Klassifikationsbedingung. Das traf sowohl auf die Situation mit linear separablen als auch die mit nicht linear separablen Kategorien zu.

Wenn auch diese generelle Tendenz bei den einzigartigen Prognosen der Random-Search Prozedur bei beiden Modellen ersichtlich war, konnten jedoch bei SUSTAIN vereinzelte Prognosen identifiziert werden, welche zumindest ansatzweise das qualitative Muster der Befunde der Experimente 3 und 4 aufwiesen. Das wiederum war beim URM nicht der Fall. So ist letzteres Modell zwar in der Lage eine geringere Schwierigkeit in der Inferenz- gegenüber der Klassifikationsbedingung vorherzusagen, dann aber immer nur für beide Experimente und nicht Experiment 3 allein. Die Experimente von Yamauchi und Kollegen stellen also eine klare Herausforderung für das URM und zumindest teilweise auch für SUSTAIN dar, welche in weiterführenden Analysen nähere Betrachtung erfordert.

Ähnliches gilt für Experiment 7, in welchem beide Modelle eine Sortierung nach Familienähnlichkeit, anstelle einer eindimensionalen Sortierung prognostizieren. Love et al. (2004), welche angeben dieses Experiment mit SUSTAIN erfolgreich modelliert zu haben, sehen einen hierfür wesentlichen Mechanismus in der Aufmerksamkeitslenkung von SUSTAIN. Dieser Ansatz erscheint vielversprechend. Angelehnt an den Mechanismus von SUSTAIN könnte daher beispielsweise analog zu (58) eine Anpassung der Aufmerksamkeitsgewichte bei unsupervidiertem Training stattfinden, sodass die relative Generierungswahrscheinlichkeit eines Stimulus nach jedem Trial durch den derzeit eigenen Cluster erhöht wird. Ob dieses Prinzip von SUSTAIN tatsächlich auch im URM umsetzbar ist und ob diese Vermischung von mechanistischen und rationalen Ansätzen (wie bereits durch das im URM implementierte Gradientenverfahren geschehen) auch theoretisch begründbar ist, bleibt Gegenstand zukünftiger Forschung.

4.2 Theoretische Aspekte des Verfahrens

In Bezug auf die derzeitige theoretische Fundierung des URM sind ebenfalls zwei Modellaspekte zu hinterfragen. Hierbei ist der erste und offensichtliche die Frage nach der Adäquatheit der bisher verwendeten Basisverteilung und Likelihood. Zukünftige Modellierungen von Kategorisierungsexperimenten auf Basis eines HDPMM sollten stärker auf diesen Aspekt fokussieren und insbesondere theoretisch fundiertere Verteilungen inspizieren. Dabei sind bezüglich der (hierarchischen) Architektur der Verteilung prinzipiell keine Schranken gesetzt. Zwar kann der zeitliche Aufwand einer manuellen Implementation einer neuen Basisverteilung in npBayes2.1 recht hoch ausfallen, jedoch lässt sich beispielsweise JAGS (Plummer, 2003) oder eine anderer Monte Carlo Simulationssoftware in die Samplingprozedur des HDPMM integrieren um (21) anschließend ohne bedeutenden Mehraufwand Modelle mit beliebig komplexe Priorverteilungen testen zu können.

Ein bedeutsamerer Aspekt des URM, welcher einer näheren Betrachtung bedarf, ist allerdings die Natur des Clusteringprozess des HDP, das sogenannte Rich-Gets-Richer-Prinzip. Es ist bekannt, dass ein solches Phänomen in komplexen Netzwerken auftritt, sodass Knoten mit vielen Verbindungen häufiger neue Verknüpfungen erhalten als Knoten mit wenigen Verbindungen (Barabási, 2009). So lässt sich beispielsweise bei der Reichtumsverteilung, bei Stadtgrößen, bei Zitationen von wissenschaftlichen Artikeln oder bei Buchverkäufen ein Rich-Gets-Richer-Prozess beobachten (Angle, 1993; Newman, 2005). Es ist hingegen ungeklärt, ob ein solches Phänomen auch in den grundlegendsten kognitiven Prozessen wie der Kategorisierung anzutreffen ist. So gibt zwar Anderson für sein ebenfalls Rich-Gets-Richer-basiertes Kategorisierungsmodell RMC an, dass die rationale Basis für ein solches Prinzip offensichtlich sei, jedoch ist es sonst in keinem anderen einflussreichen Kategorisierungsmodell anzutreffen. Tatsächlich sehen selbst die Autoren der Artikel, in welchen DPMMs oder HDPMMs als Kategorisierungsverfahren vorgeschlagen werden, die Stärken des DP bzw. HDP ausschließlich in der dynamischen Anpassung der Anzahl der Cluster an die Daten (Griffiths et al., 2007; Griffiths, Sanborn, Canini, & Navarro, 2008; Kemp, Perfors, & Tenenbaum, 2007; Sanborn et al., 2006) und damit einhergehend im Komfort keine komplexen Modelselektionsverfahren anwenden zu müssen (Navarro, Griffiths, Steyvers, & Lee, 2006), beziehungsweise in der Möglichkeit, das Prinzip der Einfachheit (siehe auch Chater & Vitányi, 2003) auf Partitionen anwenden zu können (Kemp, Tenenbaum, Niyogi, & Griffiths, 2010). Das Rich-Gets-Richer-Prinzip im DPMM bzw. HDPMM scheint dabei nur ein Nebeneffekt zu sein, welcher entweder nicht bemerkt oder als irrelevant betrachtet wurde. Tatsächlich aber ist es ein

bedeutsamer Mechanismus des jeweiligen Modells, welcher zudem einen charakteristischen Bias auf die Verteilung der Clustergrößen im DPMM bzw. HDPMM ausübt und somit nur eine Art von Rich-Gets-Richer-Prozess unter vielen Alternativen darstellt. Im Kontext der Textanalyse haben sich beispielsweise anstelle von DPs sogenannte Pitman-Yor Prozesse als brauchbarer erwiesen, da Letztere eine Verteilung der Clustergrößen mit einer sogenannten Power-Law Form garantieren und damit Worthäufigkeiten in Texten besser abbilden (Sato & Nakagawa, 2010; Teh, 2006, 2010).

Auch in anderen Domänen existieren spezialisierte Verteilungen, welche das Resultat eines Rich-Gets-Richer-Prozesses beschreiben und dennoch von der durch den CRP bzw. des CRFs induzierten Verteilung verschieden sind. In der Domäne der Städteentwicklung gilt beispielsweise die Paretoverteilung als gute Approximation der Verteilung der Städtegrößen (Córdoba, 2008). Für die Modellierung von Rich-Gets-Richer-Prozessen bei Webseitenzugriffe, Größe von Unternehmen und wissenschaftlichen Zitationen werden hingegen die Zipf und die Yule Verteilung diskutiert (Kochen, Crickman, & Blaivas, 1982; Li, 2002). Wenn also das Rich-Gets-Richer-Prinzip auch in einem basalen kognitiven Mechanismus wie der Kategorisierung anzutreffen ist, bleibt die Frage ob die charakteristische Clustegrößenverteilung, induziert durch den CRP bzw. CRF, die Realität adäquat abbildet oder nicht vielmehr eine alternative Verteilung eine bessere Beschreibung darstellt.

Neben diesem URM-typischen Mechanismus ist auch offen, ob die vorgestellte Modellverbesserung für Experiment 1, das Erlernen von Dimensionsgewichten über Gradient-Descent ein realistischer Mechanismus darstellt. Es ist zwar ohne Frage, dass dem URM bisher ebenso wie dem RMC eine Methode zur Bestimmung der Dimensionssalienz gefehlt hat (siehe auch Nosofsky et al., 1994), doch gibt es mittlerweile Zweifel an der Realitätsnähe des Gradientenverfahrens, welches unter anderem auch in ALCOVE und SUSTAIN verwendet wird. Der Grund hierfür ist, dass Menschen z.B. beim Erlernen von Kategorisierungsregeln keinen durchgängigen graduellen Lernfortschritt zeigen, sondern einen sprunghaften (J. Smith & Ell, 2015). Smith und Ell (2015) konnten beispielhaft belegen, dass weder aktuelle Einzelsystemmodelle mit Gradient-Descent noch Multisystemmodelle mit einem gesonderten System zur Identifizierung von Regeln diese sprunghaften Übergänge prognostizieren können. Wenn also ein Einzelsystemansatz gewählt wird, wie es bei SUSTAIN oder dem URM der Fall ist, muss die bisherige gängige Implementierung des Aufmerksamkeitsmechanismus hinterfragt werden. So ist beispielsweise zu prüfen ob den sprunghaften Leistungsverbesserungen zumindest mit einer adaptiven theoriegeleiteten statt einer wie bisher üblichen fixen Lernrate begegnet werden kann.

Für die Anwendung des Kategorisierungsverfahrens in kognitiven Robotern bedarf es darüber hinaus einer Methode zur Bestimmung der Wichtigkeit von Dimensionen, bei welcher die für die Kategorisierung relevanten Dimensionen nicht, wie in der Kategorisierungsforschung üblich, im

Code vorgegeben sind. Beispielsweise wurden auch in den Modellierungen dieser Arbeit Stimuli über Ausprägungen in relevanten abstrakten Dimensionen wie Größe, Farbe oder Form beschrieben, während ein künstliches kognitives System zunächst nur einfache Sensordaten vorliegen hat. Heller, Sanborn und Chater (2009) beschreiben für eine solche Situation ein rationales Verfahren, welches es ermöglicht nicht nur Dimensionsbiases sondern auch die relevanten Dimensionen bzw. ihre Basisvektoren aus den Daten zu lernen, sodass die Daten im Raum beliebig rotiert sein können.
Heller et al. (2009) zeigen, dass über dieses Verfahren das voneinander abweichende Kategorisierungsverhalten von Kleinkindern unterschiedlicher Altersstufen und Erwachsen in einer perzeptuellen Klassifikationsaufgabe durch sich entwickelnde Dimensionsbiases prognostiziert werden kann, allerdings fehlen, wie auch ursprünglich beim URM, weitere Evaluationen. Für die Entwicklung eines psychologisch orientierten Kategorisierungsverfahrens für kognitive Roboter bietet es sich an, diesen Ansatz zunächst weiter zu untersuchen.

4.3 Technische Aspekte des Verfahrens

Um das URM in weiteren Untersuchungen in der Kategorisierungsforschung oder auch im Robotikkontext möglichst effizient anwenden zu können, bedarf es einer grundlegenden Überarbeitung oder Ersetzung der derzeitigen Softwareimplementation. npBayes2.1G besteht derzeit zum Großteil aus C-Routinen, diese sind jedoch in Form von .mex Dateien in einem zentralen MATLAB Code eingebunden, welcher sich als durchaus träge erwiesen hat. Die Parameteroptimierung für Experiment 1 benötigte beispielsweise trotz Parallelisierung auf zwei Xeon X5650, zwei Xeon E5-2620 v4 und einem i7 3630QM ca. 2 Tage Rechenzeit. Eine experimentelle Implementation eines DPMM, zusammen mit der im Methodikteil vorgestellten Mean-Partition-Berechnung in Java erzeugte dabei den Eindruck, dass allein durch einen Sprachenwechsel bedeutsame Geschwindigkeitszuwächse möglich sind. Dies mag vermutlich auch der Grund sein, warum neuere Implementationen des HDP komplett in C++ (Wang & Blei, 2017), Python (Wang, 2015) bzw. Java (Bleier, 2011) umgesetzt wurden. Darüber hinaus sind seit dem Release von npBayes2.1 (Teh, 2004) verbesserte Inferenzverfahren vorgestellt worden, welche demnach in npBayes2.1G nicht implementiert sind. Hierzu gehören Split-Merge-MCMC-Sampling-Algorithmen (Rana, Phung, & Venkatesh, 2013; Wang & Blei, 2012), ein Verfahren zur Parallelisierung und Auslagerung der Samplingberechnungen auf Grafikkarten über CUDA (Suchard et al., 2010) oder die besonders für den Robotikkontext interessante Online Variational Inferenz von Wang, Paisley und Blei (2011), welche für die Verarbeitung von große Mengen an live

gestreamten Daten entwickelt wurde. Letzteres könnte den HDPMM von Nakamura, Nagai & Iwahasi (2011) bzw. Aoki, Nishihara, Nakamura und Nagai (2016), in welcher viele Sensordaten live verarbeitet werden, zumindest technisch verbessern.
Sofern die Berechnung der Mean-Partition für ein Szenario von Interesse ist, bieten sich auch hier Möglichkeiten der Optimierung. Beispielsweise könnte für den Robotikkontext eine Ersetzung der ungarischen Methode durch den Deep-Greedy-Switching-Algorithmus von Naiem und El-Beltagy (2013) lohnend sein. Letzterer stellt eine Approximation der Ungarischen Methode zur beschleunigten Verarbeitung von großen Mengen an Streamingdaten dar und ist auch als parallelisierte Variante verfügbar.
Ist hingegen eine optimale Lösung des Zuordungsproblems zwingend, so könnte der Suchalgorithmus von Andolfatto et al. (2007) modifiziert werden, sodass lediglich die Kostenmatrizen jeder Partition nach einer Veränderung der vorgeschlagenen Mean-Partition aktualisiert werden, um anschließend deren mittlere quadrierte Distanz zu allen Partitionen über die dynamische ungarische Methode von Mills-Tettey et al. (2007) zu berechnen (entwickelt in persönlichem Gespräch mit Timo von Oertzen, 15. Januar 2017).

4.4 Konklusion

Trotz der noch bestehenden Mängel bei der Prognosegüte des Verfahrens, stellt das URM einen vielversprechenden Ansatz für die weitere Forschung dar. So ist bisher offen und auch weitestgehend unbeachtet, ob das im URM zentrale Rich-Gets-Richer-Prinzip, welches sich in vielen anderen Bereichen des alltäglichen Lebens wiederfindet, tatsächlich auch in der Kognition vorliegt.
Unabhängig davon kann das URM als ein allgemeines Framework der rationalen Kategorisierung nach Anderson betrachtet werden, da es alle anderen rationalen Kategorisierungsmodelle in sich vereint. In Kombination mit der in der Diskussion eingangs beschriebenen Erweiterung durch JAGS oder einer anderen Monte Carlo Simulationssoftware sind dabei beliebig komplexe hierarchische Basisverteilungen unkompliziert verwend- und testbar. Hierdurch lässt sich mit dem Modell nicht nur ähnlichkeitsbasierte Kategorisierung modellieren, sondern auch die bisher noch verhältnismäßig wenig beforschte theoriegeleitete Kategorisierung aus Perspektive der Bayes‘schen Kognition untersuchen, welche wiederum für die kognitive Robotik und die Frage nach den Mechanismen der menschlichen Wissensakquisition unmittelbar nützlich sein kann.

Literaturverzeichnis

Abbott, J., Nagy, Z., Beyeler, F., & Nelson, B. (2007). Robotics in the Small, Part I: Microbotics. *IEEE Robotics & Automation Magazine*, *14*(2), 92–103. https://doi.org/10.1109/MRA.2007.380641

Almudevar, A., & Field, C. (1999). Estimation of Single-Generation Sibling Relationships Based on DNA Markers. *Journal of Agricultural, Biological, and Environmental Statistics*, *4*(2), 136–165. https://doi.org/10.2307/1400594

Amari, S., & Misra, R. (1997). Closed-form expressions for distribution of sum of exponential random variables. *IEEE Transactions on Reliability*, *46*(4), 519–522. https://doi.org/10.1109/24.693785

Anderson, J. (1990). *The adaptive character of thought*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Anderson, J. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*(3), 409–429. https://doi.org/10.1037/0033-295X.98.3.409

Anderson, J. (2015). *Cognitive Psychology and its Implications* (8th ed.). New York: Worth Publishers.

Anderson, J., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, *96*(4), 703–719. https://doi.org/10.1037/0033-295X.96.4.703

Anderson, M. (2003). Embodied Cognition: A field guide. *Artificial Intelligence*, *149*(1), 91–130. https://doi.org/10.1016/S0004-3702(03)00054-7

Angeles, J. (2014). *Fundamentals of Robotic Mechanical Systems* (4th ed.). Swizterland: Springer International Publishing.

Angle, J. (1993). Deriving the size distribution of personal wealth from “the rich get richer, the poor get poorer.” *The Journal of Mathematical Sociology*, *18*(1), 27–46. https://doi.org/10.1080/0022250X.1993.9990114

Aoki, T., Nishihara, J., Nakamura, T., & Nagai, T. (2016). Online Joint Learning of Object Concepts and Language Model using Multimodal Hierarchical Dirichlet Process. In I. Suh (Ed.), *2016*

IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 2636–2642). New York, NY: IEEE. https://doi.org/10.1109/IROS.2016.7759410

Apostolopoulos, D. (2001). *Analytic Configuration of Wheeled Robotic Locomotion*. (Carnegie Mellon University, Ed.). Retrieved from http://repository.cmu.edu/robotics/29/

Argall, B., Chernova, S., Veloso, M., & Browning, B. (2009). A survey of robot learning from demonstration. *Robotics and Autonomous Systems*, *57*(5), 469–483. https://doi.org/10.1016/j.robot.2008.10.024

Asada, M., MacDorman, K., Ishiguro, H., & Kuniyoshi, Y. (2001). Cognitive developmental robotics as a new paradigm for the design of humanoid robots. *Humanoid Robots*, *37*(2–3), 185–193. https://doi.org/10.1016/S0921-8890(01)00157-9

Ashby, F. (2013). Human Category Learning, Neural Basis. In H. Pashler (Ed.), *Encyclopedia of the Mind* (pp. 130–134). SAGE Publications.

Ashby, F., Alfonso-Reese, L., Turken, A., & Waldron, E. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442–481. https://doi.org/10.1037/0033-295X.105.3.442

Ashby, F., & Maddox, W. (2004). Human Category Learning. *Annu. Rev. Psychol.*, *56*(1), 149–178. https://doi.org/10.1146/annurev.psych.56.091103.070217

Aulinas, J., Petillot, Y., Salvi, J., & Lladó, X. (2008). The SLAM problem: a survey. In T. Alsinet, J. Puyol-Gruart, & C. Torras (Eds.), *Proceedings of the 2008 conference on Artificial Intelligence Research and Development* (pp. 363–371). Amsterdam: IOS Press.

Bakker, P., & Kuniyoshi, Y. (1996). Robot See, Robot Do : An Overview of Robot Imitation. In N. Sharkey (Ed.), *AISB96 Workshop on Learning in Robots and Animals* (pp. 3–11).

Barabási, A. (2009). Scale-Free Networks: A Decade and Beyond. *Science*, *325*(5939), 412. https://doi.org/10.1126/science.1173299

Barsalou, L. (2007). Grounded Cognition. *Annu. Rev. Psychol.*, *59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barsalou, L. (2010). Grounded Cognition: Past, Present, and Future. *Topics in Cognitive Science*, *2*(4), 716–724. https://doi.org/10.1111/j.1756-8765.2010.01115.x

Barsalou, L., & Billman, D. (1989). Systematicity and Semantic Ambiguity. In D. Gorfein (Ed.), *Resolving Semantic Ambiguity* (pp. 146–203). New York: Springer-Verlag.

Bather, J. (1996). A Conversation with Herman Chernoff. *Statistical Science*, *11*(4), 335–350.

Bayes, & Price. (1763). An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philosophical Transactions*, *53*, 370–418. https://doi.org/10.1098/rstl.1763.0053

Beer, R. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, *4*(3), 91–99. https://doi.org/10.1016/S1364-6613(99)01440-0

Beer, R., Quinn, R., Chiel, H., & Ritzmann, R. (1997). Biologically Inspired Approaches to Robotics: What Can We Learn from Insects? *Commun. ACM*, *40*(3), 30–38. https://doi.org/10.1145/245108.245118

Berger-Wolf, T., Sheikh, S., DasGupta, B., Ashley, M., Caballero, I., Chaovalitwongse, W., & Putrevu, S. (2007). Reconstructing sibling relationships in wild populations. *Bioinformatics*, *23*(13), i49–i56. https://doi.org/10.1093/bioinformatics/btm219

Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, *13*(2), 281–305.

Best, C. A., Yim, H., & Sloutsky, V. M. (2013). The cost of selective attention in category learning: Developmental differences between adults and infants. *Journal of Experimental Child Psychology*, *116*(2), 105–119. https://doi.org/10.1016/j.jecp.2013.05.002

Bhatti, J., Plummer, A. R., Iravani, P., & Ding, B. (2015). A survey of dynamic robot legged locomotion. In J. Han (Ed.), *2015 International Conference on Fluid Power and Mechatronics* (pp. 770–775). IEEE. https://doi.org/10.1109/FPM.2015.7337218

Billman, D., & Knutson, J. (1996). Unsupervised concept learning and value systematicitiy: A complex whole aids learning the parts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(2), 458–475. https://doi.org/10.1037/0278-7393.22.2.458

Birk, A. (2011). What is Robotics? An Interdisciplinary Field Is Getting Even More Diverse [Education]. *IEEE Robotics & Automation Magazine*, *18*(4), 94–95. https://doi.org/10.1109/MRA.2011.943235

Blair, M., & Homa, D. (2001). Expanding the search for a linear separability constraint on category learning. *Memory & Cognition*, *29*(8), 1153–1164. https://doi.org/10.3758/BF03206385

Blair, M., & Homa, D. (2003). As easy to memorize as they are to classify: The 5–4 categories and the category advantage. *Memory & Cognition*, *31*(8), 1293–1301. https://doi.org/10.3758/BF03195812

Blei, D. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Bleier, A. (2011). *Java Gibbs sampler for the Hierarchical Dirichlet Processes*. Retrieved from https://github.com/arnim/HDP

Bobrowski, L., & Łukaszuk, T. (2009). Feature selection based on relaxed linear separability. *Biocybernetics and Biomedical Engineering*, *29*(2), 43–58.

Breazeal, C. (2004). Social interactions in HRI: the robot view. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *34*(2), 181–186. https://doi.org/10.1109/TSMCC.2004.826268

Brogårdh, T. (2007). Present and future robot control development—An industrial perspective. *Annual Reviews in Control*, *31*(1), 69–79. https://doi.org/10.1016/j.arcontrol.2007.01.002

Brunner, E., & Munzel, U. (2000). The Nonparametric Behrens-Fisher Problem: Asymptotic Theory and a Small-Sample Approximation. *Biometrical Journal*, *42*(1), 17–25. https://doi.org/10.1002/(SICI)1521-4036(200001)42:1<17::AID-BIMJ17>3.0.CO;2-U

Buehren, M. (2014). *Functions for the rectangular assignment problem: Requires MATLAB 6.5.1 (R13SP1)*. Retrieved from http://www.mathworks.com/matlabcentral/fileexchange/6543-functions-for-the-rectangular-assignment-problem

Burkard, R., Dell'Amico, M., & Martello, S. (2012). *Assignment Problems*. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611972238

Canini, K., Shashkov, M., & Griffiths, T. (2010). Modeling Transfer Learning in Human Categorization with the Hierarchical Dirichlet Process. In J. Fürnkranz & T. Joachims (Eds.), *Proceedings of the 27th International Conference on Machine Learning (ICML-10)* (pp. 151–158). Haifa, Israel: Omnipress.

Capek, K., & Shenef, Y. (2016). *Karel Capeks R.U.R. - Rossum Universal Robots: ins Deutsche übersetzt und aktualisiert von Yehuda Shenef*. Books on Demand.

Chater, N., & Oaksford, M. (1999). Ten years of the rational analysis of cognition. *Trends in Cognitive Sciences*, *3*(2), 57–65. https://doi.org/10.1016/S1364-6613(98)01273-X

Chater, N., & Oaksford, M. (Eds.). (2008). *The Probabilistic Mind: Prospects for Bayesian Cognitive Science*. Oxford: Oxford University Press.

Chater, N., Tenenbaum, J., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Special Issue: Probabilistic Models of Cognition*, *10*(7), 287–291. https://doi.org/10.1016/j.tics.2006.05.007

Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences*, *7*(1), 19–22. https://doi.org/10.1016/S1364-6613(02)00005-0

Chatterjee, A. (2010). Disembodying cognition. *Language and Cognition*, *2*(1), 79–116. https://doi.org/10.1515/LANGCOG.2010.004

Chen, S., Li, Y., & Kwok, N. M. (2011). Active vision in robotic systems: A survey of recent developments. *The International Journal of Robotics Research*, *30*(11), 1343–1377. https://doi.org/10.1177/0278364911410755

Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, *49*(4), 327–335. https://doi.org/10.1080/00031305.1995.10476177

Christaller, T. (1999). Cognitive robotics: a new approach to artificial intelligence. *Artificial Life and Robotics*, *3*(4), 221–224. https://doi.org/10.1007/BF02481184

Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Sciences*, *3*(9), 345–351. https://doi.org/10.1016/S1364-6613(99)01361-3

Clark, A., & Grush, R. (1999). Towards a Cognitive Robotics. *Adaptive Behavior*, *7*(1), 5–16. https://doi.org/10.1177/105971239900700101

Collier, M. (2005). Hume and cognitive science: The current status of the controversy over abstract ideas. *Phenomenology and the Cognitive Sciences*, *4*(2), 197–207. https://doi.org/10.1007/s11097-005-0139-5

Connell, J. H., & Mahadevan, S. (Eds.). (1993). *Robot Learning*. Boston, MA: Springer US.

Córdoba, J.-C. (2008). On the distribution of city sizes. *Journal of Urban Economics*, *63*(1), 177–197. https://doi.org/10.1016/j.jue.2007.01.005

Dahiya, R., Metta, G., Valle, M., & Sandini, G. (2010). Tactile Sensing—From Humans to Humanoids. *IEEE Transactions on Robotics*, *26*(1), 1–20. https://doi.org/10.1109/TRO.2009.2033627

Dean, T., & Kambhampati, S. (1996). Planning and Scheduling. In A. Tucker & H. Abelson (Eds.), *CRC Handbook of Computer Science and Engineering* (pp. 1–40). Boca Raton: CRC Press.

Desouza, G., & Kak, A. (2002). Vision for mobile robot navigation: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(2), 237–267. https://doi.org/10.1109/34.982903

D'Mello, S., & Franklin, S. (2011). Computational modeling/cognitive robotics complements functional modeling/experimental psychology. *Special Issue: Cognitive Robotics and Reevaluation of Piaget Concept of Egocentrism*, *29*(3), 217–227. https://doi.org/10.1016/j.newideapsych.2009.07.003

Dong, L., & Nelson, B. (2007). Tutorial - Robotics in the small Part II: Nanorobotics. *IEEE Robotics & Automation Magazine*, *14*(3), 111–121. https://doi.org/10.1109/MRA.2007.905335

Dorigo, M., & Colombetti, M. (1994). Robot shaping: developing autonomous agents through learning. *Artificial Intelligence*, *71*(2), 321–370. https://doi.org/10.1016/0004-3702(94)90047-7

Dove, G. (2010). On the need for Embodied and Dis-Embodied Cognition. *Frontiers in Psychology*, *1*, 242. https://doi.org/10.3389/fpsyg.2010.00242

Dudewicz, E. J., Ma, Y., Mai, E., & Su, H. (2007). Exact solutions to the Behrens–Fisher Problem: Asymptotically optimal and finite sample efficient choice among. *Journal of Statistical Planning and Inference*, *137*(5), 1584–1605. https://doi.org/10.1016/j.jspi.2006.09.007

Featherstone, R., & Orin, D. (2000). Robot Dynamics: Equations and Algorithms. In *Proceedings of the 2000 IEEE lntemational Conference on Robotics and Automation* (pp. 826–834). IEEE. https://doi.org/10.1109/ROBOT.2000.844153

Fink, J. (2012). Anthropomorphism and Human Likeness in the Design of Robots and Human-Robot Interaction. In S. S. Ge, O. Khatib, J.-J. Cabibihan, R. Simmons, & M.-A. Williams (Eds.), *Social Robotics: 4th International Conference, ICSR 2012, Chengdu, China, October 29-31, 2012. Proceedings* (pp. 199–208). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-34103-8

Folstein, J. R., Palmeri, T. J., & Gauthier, I. (2013). Category Learning Increases Discriminability of Relevant Object Dimensions in Visual Cortex. *Cerebral Cortex*, *23*(4), 814–823. https://doi.org/10.1093/cercor/bhs067

Frigyik, B., Kapila, A., & Gupta, M. (2010). *Introduction to the Dirichlet Distribution and Related Processes: UWEE Technical Report Number UWEETR-2010-0006*. (University of Washington, Ed.). Retrieved from http://mayagupta.org/publications/FrigyikKapilaGuptaIntroToDirichlet.pdf

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, D. (2014). *Bayesian Data Analysis*. Boca Raton, FL, USA: Chapman and Hall/CRC.

Geman, S., & Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *Pattern Analysis and Machine Intelligence, IEEE Transactions On*, *PAMI-6*(6), 721–741. https://doi.org/10.1109/TPAMI.1984.4767596

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition*, *65*(2–3), 263–297. https://doi.org/10.1016/S0010-0277(98)00002-X

Gershman, S., & Blei, D. (2012). A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, *56*(1), 1–12. https://doi.org/10.1016/j.jmp.2011.08.004

Ghallab, M., Nau, D., & Traverso, P. (2016). *Automated Planning and Acting*. Cambridge: Cambridge University Press.

Glassen, T., & Nitsch, V. (2015). Regel-basierte Kategorisierung mit dem Hierarchischen Dirichlet Prozess. *Kognitive Systeme*, *2*(2). https://doi.org/10.17185/duepublico/40721

Glassen, T., & Nitsch, V. (2016). Hierarchical Bayesian models of cognitive development. *Biological Cybernetics*, *110*(2), 217–227. https://doi.org/10.1007/s00422-016-0686-6

Glenberg, A. M., Witt, J. K., & Metcalfe, J. (2013). From the Revolution to Embodiment: 25 Years of Cognitive Psychology. *Perspectives on Psychological Science*, *8*(5), 573–585. https://doi.org/10.1177/1745691613498098

Goldstone, R. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition*, *52*(2), 125–157. https://doi.org/10.1016/0010-0277(94)90065-5

Goldstone, R., & Kersten, A. (2003). Concepts and Categorization. In I. Weiner (Ed.), *Handbook of Psychology* (pp. 599–622). John Wiley & Sons, Inc. https://doi.org/10.1002/0471264385.wei0422

Goodrich, M. A., & Schultz, A. C. (2008). Human–Robot Interaction: A Survey. *Foundations and Trends\circledR in Human–Computer Interaction*, *1*(3), 203–275. https://doi.org/10.1561/1100000005

Griffiths, T. ., Canini, K., Sanborn, A., & Navarro, D. (2007). Unifying rational models of categorization via the hierarchical Dirichlet process. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual conference of the Cognitive Science Society* (pp. 323–328). Hillsdale and NJ: Erlbaum.

Griffiths, T., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. (2010). Probabilistic models of cognition: exploring representations and inductive biases. *Trends in Cognitive Sciences, 14*(8), 357–364. https://doi.org/10.1016/j.tics.2010.05.004

Griffiths, T., Sanborn, A., Canini, K., & Navarro, D. (2008). Categorization as nonparametric density estimation. In N. Chater & M. Oaksford (Eds.), *The Probabilistic Mind: Prospects for Bayesian Cognitive Science* (pp. 303–328). Oxford: Oxford University Press.

Gupta, A., & Nadarajah, S. (2004). *Handbook of Beta Distribution and Its Applications*. Boca Raton, Florida: CRC Press.

Gurdan, D., Stumpf, J., Achtelik, M., Doth, K., Hirzinger, G., & Rus, D. (2007). Energy-efficient Autonomous Four-rotor Flying Robot Controlled at 1 kHz. In P. Dario & A. D. Luca (Eds.), *Proceedings 2007 IEEE International Conference on Robotics and Automation* (pp. 361–366). IEEE. https://doi.org/10.1109/ROBOT.2007.363813

Gureckis, T. (2014). *sustain_python*. Retrieved from https://github.com/NYUCCL/sustain_python

Gureckis, T., & Love, B. (2003). Towards a unified account of supervised and unsupervised category learning. *Journal of Experimental & Theoretical Artificial Intelligence*, *15*(1), 1–24. https://doi.org/10.1080/09528130210166097

Gusfield, D. (2002). Partition-distance: A problem and class of perfect graphs arising in clustering. *Information Processing Letters*, *82*(3), 159–164. https://doi.org/10.1016/S0020-0190(01)00263-0

Hastings, W. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, *57*(1), 97–109. https://doi.org/10.2307/2334940

Heller, K., Sanborn, A., & Chater, N. (2009). Hierarchical Learning of Dimensional Biases in Human Categorization. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Advances in Neural Information Processing Systems 22* (pp. 727–735). Curran Associates, Inc. Retrieved from https://warwick.ac.uk/fac/sci/psych/people/asanborn/asanborn/hierarchical_biases.pdf

Hill, C., Amodeo, A., Joseph, J. V., & Patel, H. R. H. (2008). Nano- and microrobotics: how far is the reality? *Expert Review of Anticancer Therapy*, *8*(12), 1891–1897. https://doi.org/10.1586/14737140.8.12.1891

Holmes, P., Full, R., Koditschek, D., & Guckenheimer, J. (2006). The Dynamics of Legged Locomotion: Models, Analyses, and Challenges. *SIAM Review*, *48*(2), 207–304. https://doi.org/10.1137/S0036144504445133

Holyoak, K. J., & Cheng, P. W. (2010). Causal Learning and Inference as a Rational Process: The New Synthesis. *Annu. Rev. Psychol.*, *62*(1), 135–163. https://doi.org/10.1146/annurev.psych.121208.131634

Huelsenbeck, J. P., & Andolfatto, P. (2007). Inference of Population Structure Under a Dirichlet Process Model. *Genetics*, *175*(4), 1787. https://doi.org/10.1534/genetics.106.061317

Hull, C. (1920). Quantitative aspects of evolution of concepts: An experimental study. *Psychological Monographs: General and Applied*, *28*(1), i-86. https://doi.org/10.1037/h0093130

Hwang, Y. K., & Ahuja, N. (1992). Gross Motion Planning - A Survey. *ACM Computing Surveys*, *24*(3), 219–291. https://doi.org/10.1145/136035.136037

Jabin, S. (Ed.). (2010). *Robot Learning*. Rijeka: Sciyo. Retrieved from http://www.intechopen.com/books/robot-learning

Jonker, R., & Volgenant, T. (1986). Improving the Hungarian assignment algorithm. *Operations Research Letters*, *5*(4), 171–175. https://doi.org/10.1016/0167-6377(86)90073-8

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, *4*, 237–285. https://doi.org/10.1613/jair.301

Kaneko, K., Harada, K., & Kanehiro, F. (2008). Humanoid Robot HRP-3. In R. Chatila, Merlet J., & C. Laugier (Eds.), *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 2471–2478). IEEE. https://doi.org/10.1109/IROS.2008.4650604

Katz, D., Kenney, J., & Brock, O. (2008). How Can Robots Succeed in Unstructured Environments? In R. Platt, S. Haddadin, C. Kemp, L. Natale, & N. E. Sian (Eds.), *Proceedings of the Robotics: Science & Systems 2008 Manipulation Workshop - Intelligence in Human Environments*.

Kemp, C., Edsinger, A., & Torres-Jara, E. (2007). Challenges for robot manipulation in human environments [Grand Challenges of Robotics]. *IEEE Robotics & Automation Magazine*, *14*(1), 20–29. https://doi.org/10.1109/MRA.2007.339604

Kemp, C., Perfors, A., & Tenenbaum, J. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*(3), 307–321. https://doi.org/10.1111/j.1467-7687.2007.00585.x

Kemp, C., Tenenbaum, J., Niyogi, S., & Griffiths, T. (2010). A probabilistic model of theory formation. *Cognition*, *114*(2), 165–196. https://doi.org/10.1016/j.cognition.2009.09.003

Khaleghi, B., Khamis, A., Karray, F. O., & Razavi, S. N. (2013). Multisensor data fusion: A review of the state-of-the-art. *Information Fusion*, *14*(1), 28–44. https://doi.org/10.1016/j.inffus.2011.08.001

Kochen, M., Crickman, R., & Blaivas, A. (1982). Distribution of scientific experts as recognized by peer consensus. *Scientometrics*, *4*(1), 45–56. https://doi.org/10.1007/BF02098005

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*(1), 59–69. https://doi.org/10.1007/BF00337288

Komatsu, L. K. (1992). Recent views of conceptual structure. *Psychological Bulletin*, *112*(3), 500–526. https://doi.org/10.1037/0033-2909.112.3.500

Konietschke, F., Placzek, M., Schaarschmidt, F., & Hothorn, L. A. (2015). nparcomp: An R Software Package for Nonparametric Multiple Comparisons and Simultaneous Confidence Intervals. *Journal of Statistical Software; Vol 1, Issue 9 (2015)*.

Konovalov, D. A., Litow, B., & Bajema, N. (2005). Partition-distance via the assignment problem. *Bioinformatics*, *21*(10), 2463–2468. https://doi.org/10.1093/bioinformatics/bti373

Kruschke, J. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*(1), 22–44. https://doi.org/10.1037/0033-295X.99.1.22

Kruschke, J. (2010). *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. Burlington, MA: Academic Press.

Kruse, R., Borgelt, C., Klawonn, F., Moewes, C., Steinbrecher, M., & Held, P. (2013). *Computational Intelligence: A Methodological Introduction*. London: Springer-Verlag London. https://doi.org/10.1007/978-1-4471-5013-8

Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, *2*(1–2), 83–97. https://doi.org/10.1002/nav.3800020109

Kurtz, K. J., Levering, K. R., Stanton, R. D., Romero, J., & Morris, S. N. (2013). Human learning of elemental category structures: Revising the classic result of Shepard, Hovland, and Jenkins (1961). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 552–572. https://doi.org/10.1037/a0029178

Lafaye, J., Gouaillier, D., & Wieber, P. (2014). Linear Model Predictive Control of the locomotion of Pepper, a humanoid robot with omnidirectional wheels. In C. Balaguer (Ed.), *2014 IEEE-RAS International Conference on Humanoid Robots* (pp. 336–341). IEEE. https://doi.org/10.1109/HUMANOIDS.2014.7041381

Lake, B., Zaremba, W., Fergus, R., & Gureckis, T. (2015). Deep neural networks predict category typicality ratings for images. In R. Dale, C. Jennings, P. Maglio, T. Matlock, D. Noelle, A. Warlaumont, & J. Yoshimi (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society*. Cognitive Science Society.

Lassaline, M., & Murphy, G. (1996). Induction and category coherence. *Psychonomic Bulletin & Review, 3*(1), 95–99. https://doi.org/10.3758/BF03210747

Lattimore, T., & Hutter, M. (2013). No Free Lunch versus Occam's Razor in Supervised Learning. In D. L. Dowe (Ed.), *Algorithmic Probability and Friends. Bayesian Prediction and Artificial Intelligence* (pp. 223–235). Heidelberg: Springer.

Lee, M., & Wagenmakers, E. (2014). *Bayesian Cognitive Modeling: A Practical Course*. Cambridge: Cambridge University Press.

Lenarcic, J., & Husty, M. (Eds.). (2012). *Latest Advances in Robot Kinematics*. Dordrecht: Springer Science+Business Media Dordrecht. https://doi.org/10.1007/978-94-007-4620-6

Leonard, J. J., & Durrant-Whyte, H. F. (1992). *Directed Sonar Sensing for Mobile Robot Navigation*. New York: Springer Science+Business Media New York. https://doi.org/10.1007/978-1-4615-3652-9

Li, W. (2002). Zipf's Law Everywhere. *Glottometrics*, *5*, 14–21.

Link, W. A., & Eaton, M. J. (2012). On thinning of chains in MCMC. *Methods in Ecology and Evolution*, *3*(1), 112–115. https://doi.org/10.1111/j.2041-210X.2011.00131.x

Little, J. N. (1993). Advances in laboratory robotics for automated sample preparation. *Chemometrics and Intelligent Laboratory Systems*, *21*(2), 199–205. https://doi.org/10.1016/0169-7439(93)89010-8

Lorencik, D., & Sincak, P. (2013). Cloud robotics: Current trends and possible use as a service. In Fodor J., L. Hluchý, & L. Vokorokos (Eds.), *IEEE 11th International Symposium on Applied Machine Intelligence and Informatics* (pp. 85–88). IEEE. https://doi.org/10.1109/SAMI.2013.6480950

Love, B. (2002). Comparing supervised and unsupervised category learning. *Psychonomic Bulletin & Review, 9*(4), 829–835. https://doi.org/10.3758/BF03196342

Love, B., & Medin, D. (1998). SUSTAIN: A Model of Human Category Learning. In J. Mostow & C. Rich (Eds.), *Proceedings of the Fifteenth National Conference on Artificial Intelligence* (pp. 671–676). Menlo Park, California: The AAAI Press.

Love, B., Medin, D., & Gureckis, T. (2004). SUSTAIN: A Network Model of Category Learning. *Psychological Review, 111*(2), 309–332. https://doi.org/10.1037/0033-295X.111.2.309

Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. New York: John Wiley & Sons, Inc.

Lungarella, M., Metta, G., Pfeifer, R., & Sandini, G. (2003). Developmental robotics: a survey. *Connection Science*, *15*(4), 151–190. https://doi.org/10.1080/09540090310001655110

Lynch, S. (2007). *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. New York: Springer-Verlag New York.

Mahon, B. Z. (2015). What is embodied about cognition? *Language, Cognition and Neuroscience*, *30*(4), 420–429. https://doi.org/10.1080/23273798.2014.987791

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Links and Interactions Between Language and Motor Systems in the Brain*, *102*(1–3), 59–70. https://doi.org/10.1016/j.jphysparis.2008.03.004

Masehian, E., & Sedighizadeh, D. (2007). Classic and Heuristic Approaches in Robot Motion Planning – A Chronological Review. *International Journal of Mechanical, Aerospace, Industrial, Mechatronic and Manufacturing Engineering*, *1*(5), 228–233.

Mason, M. T. (2001). *Mechanics of Robotic Manipulation*. Cambridge, MA: MIT Press.

Mattar, E. (2013). A survey of bio-inspired robotics hands implementation: New directions in dexterous manipulation. *Robotics and Autonomous Systems*, *61*(5), 517–544. https://doi.org/10.1016/j.robot.2012.12.005

McColeman, C., Barnes, J., Chen, L., Meier, K. M., Walshe, R. C., & Blair, M. R. (2014). Learning-induced changes in attentional allocation during categorization: a sizable catalog of attention change as measured by eye movements. *PloS One*, *9*(1), e83302. https://doi.org/10.1371/journal.pone.0083302

McDermott, D. (1992). Robot planning. *AI Magazine*, *13*(2). https://doi.org/10.1609/aimag.v13i2.992

McDonnell, J., & Gureckis, T. (2011). Adaptive clustering models of categorization. In E. Pothos & A. Wills (Eds.), *Formal Approaches in Categorization* (pp. 220–252). Cambridge: Cambridge University Press.

McGeer, T. (1990). Passive Dynamic Walking. *The International Journal of Robotics Research*, *9*(2), 62–82. https://doi.org/10.1177/027836499000900206

McKee, G. T. (2006). The Maturing Discipline of Robotics. *International Journal of Engineering Education*, *22*(4), 692–701.

Medin, D. (1989). Concepts and conceptual structure. *American Psychologist*, *44*(12), 1469–1481. https://doi.org/10.1037/0003-066X.44.12.1469

Medin, D., Dewey, G., & Murphy, T. (1983). Relationships between item and category learning: Evidence that abstraction is not automatic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(4), 607–625. https://doi.org/10.1037/0278-7393.9.4.607

Medin, D., Wattenmaker, W., & Hampson, S. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology*, *19*(2), 242–279. https://doi.org/10.1016/0010-0285(87)90012-0

Mills-Tettey, G. A., Stentz, A., & Dias, M. B. (2007). *The Dynamic Hungarian Algorithm for the Assignment Problem with Changing Costs*. Pittsburgh, Pennsylvania: Carnegie Mellon University. Retrieved from http://repository.cmu.edu/cgi/viewcontent.cgi?article=1148&context=robotics

Morse, A. F., Herrera, C., Clowes, R., Montebelli, A., & Ziemke, T. (2011). The role of robotic modelling in cognitive science. *Special Issue: Cognitive Robotics and Reevaluation of Piaget Concept of Egocentrism*, *29*(3), 312–324. https://doi.org/10.1016/j.newideapsych.2011.02.001

Müller, P., & Quintana, F. (2004). Nonparametric Bayesian Data Analysis. *Statistical Science*, *19*(1), 95–110.

Munkres, J. (1957). Algorithms for the Assignment and Transportation Problems. *Journal of the Society for Industrial and Applied Mathematics*, *5*(1), 32–38. https://doi.org/10.1137/0105003

Murphy, G., & Medin, D. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*(3), 289–316. https://doi.org/10.1037/0033-295X.92.3.289

Murray, R. M., Li, Z., & Sastry, S. S. (1994). *A Mathematical Introduction to Robotic Manipulation*. Boca Raton: CRC Press.

Naiem, A., & El-Beltagy, M. (2013). On the Optimality and Speed of the Deep Greedy On the Optimality and Speed of the Deep Greedy Switching Algorithm for Linear Assignment Problems. In V. Prasanna & G. Westrom (Eds.), *2013 IEEE 27th International Symposium on Parallel & Distributed Processing Workshops and PhD Forum* (pp. 1829–1837). New York, NY: IEEE. https://doi.org/10.1109/IPDPSW.2013.54

Nakamura, T., Nagai, T., & Iwahashi, N. (2011). Multimodal Categorization by Hierarchical Dirichlet Process. In N. Amato (Ed.), *IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 1520–1525). New York, NY, USA: IEEE. https://doi.org/10.1109/IROS.2011.6094763

Navarro, D. J., Griffiths, T. L., Steyvers, M., & Lee, M. D. (2006). Modeling individual differences using Dirichlet processes. *Special Issue on Model Selection: Theoretical Developments and Applications Special Issue on Model Selection: Theoretical Developments and Applications*, *50*(2), 101–122. https://doi.org/10.1016/j.jmp.2005.11.006

Neal, R. M. (2000). Markov Chain Sampling Methods for Dirichlet Process Mixture Models. *Journal of Computational and Graphical Statistics*, *9*(2), 249–265. https://doi.org/10.1080/10618600.2000.10474879

Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, *46*(5), 323–351. https://doi.org/10.1080/00107510500052444

Nicholls, H. R., & Lee, M. H. (1989). A Survey of Robot Tactile Sensing Technology. *The International Journal of Robotics Research*, *8*(3), 3–30. https://doi.org/10.1177/027836498900800301

Nitsch, V., & Glassen, T. (2015). Investigating the effects of robot behavior and attitude towards technology on social human-robot interactions. In Y. Nakauchi (Ed.), *Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 535–540). IEEE. https://doi.org/10.1109/ROMAN.2015.7333560

Nosofsky, R. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57. https://doi.org/10.1037/0096-3445.115.1.39

Nosofsky, R., Gluck, M., Palmeri, T., Mckinley, S., & Glauthier, P. (1994). Comparing modes of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition*, *22*(3), 352–369. https://doi.org/10.3758/BF03200862

Nosofsky, R., & Johansen, M. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review*, *7*(3), 375–402.

Oaksford, M., & Chater, N. (Eds.). (1998). *Rational Models of Cognition*. Oxford: Oxford University Press.

Okamura, A. M., Smaby, N., & Cutkosky, M. R. (2000). An Overview of Dexterous Manipulation. In *Proceedings of the 2000 IEEE International Conference on Robotics and Automation* (pp. 255–262). IEEE.

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K., & Spivey, M. (2013). Computational Grounded Cognition: A New Alliance between Grounded Cognition and Computational Modeling. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00612

Pfeifer, R., Lungarella, M., & Iida, F. (2007). Self-Organization, Embodiment, and Biologically Inspired Robotics. *Science*, *318*(5853), 1088–1093. https://doi.org/10.1126/science.1145803

Pfeifer, R., Lungarella, M., & Iida, F. (2012). The Challenges Ahead for Bio-inspired "Soft" Robotics. *Commun. ACM*, *55*(11), 76–87. https://doi.org/10.1145/2366316.2366335

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In K. Hornik, F. Leisch, & A. Zeileis (Eds.), *Proceedings of the 3rd international workshop on distributed statistical computing* (pp. 125–134).

Poole, H. H. (1989). *Fundamentals of Robotics Engineering*. Springer Netherlands.

Pransky, J. (2014). The Pransky interview: Dr mark W. Tilden, robotics physicist. *Industrial Robot: An International Journal*, *41*(2), 113–118. https://doi.org/10.1108/ir-01-2014-0305

Rana, S., Phung, D., & Venkatesh, S. (2013). Split-Merge Augmented Gibbs Sampling for Hierarchical Dirichlet Processes. In J. Pei, V. S. Tseng, L. Cao, H. Motoda, & G. Xu (Eds.), *Advances in Knowledge Discovery and Data Mining: 17th Pacific-Asia Conference, PAKDD 2013, Gold Coast, Australia, April 14-17, 2013, Proceedings, Part II* (pp. 546–

557). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-37456-2

Reed, S. (1978). Category vs. item learning: Implications for categorization models. *Memory & Cognition*, *6*(6), 612–621. https://doi.org/10.3758/BF03198251

Riesen, K., & Bunke, H. (2009). Approximate graph edit distance computation by means of bipartite graph matching. *7th IAPR-TC15 Workshop on Graph-Based Representations (GbR 2007)*, *27*(7), 950–959. https://doi.org/10.1016/j.imavis.2008.04.004

Ross, B., & Makin, V. (1999). Prototype versus Exemplar Models in Cognition. In R. Sternberg (Ed.), *The Nature of Cognition* (pp. 205–241). Cambridge, MA: MIT Press.

Sanborn, A., Griffiths, T., & Navarro, D. (2006). A more rational model of categorization. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th annual conference of the Cognitive Science Society* (pp. 726–731). Mahwah, N.J.: Lawrence Erlbaum.

Sanborn, A., Griffiths, T., & Navarro, D. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review*, *117*(4), 1144–1167. https://doi.org/10.1037/a0020511

Sato, I., & Nakagawa, H. (2010). Topic models with power-law using Pitman-Yor process. In B. Rao, B. Krishnapuram, A. Tomkins, & Q. Yang (Eds.), *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 673–682).

Schaal, S. (1999). Is imitation learning the route to humanoid robots? *Trends in Cognitive Sciences*, *3*(6), 233–242. https://doi.org/10.1016/S1364-6613(99)01327-3

Schaffer, C. (1994). A Conservation Law for Generalization Performance. In W. W. Cohen & H. Hirsh (Eds.), *Machine Learning, Proceedings of the Eleventh International Conference* (pp. 259–265). San Francisco and CA: Morgan Kaufmann.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied*, *75*(13), 1–42. https://doi.org/10.1037/h0093825

Shimoga, K. B. (1996). Robot Grasp Synthesis Algorithms: A Survey. *The International Journal of Robotics Research*, *15*(3), 230–266. https://doi.org/10.1177/027836499601500302

Silvén, M. (2002). Origins of knowledge: learning and communication in infancy. *Learning and Instruction*, *12*(3), 345–374. https://doi.org/10.1016/S0959-4752(01)00026-3

Smith, J., & Ell, S. (2015). One Giant Leap for Categorizers: One Small Step for Categorization Theory. *PloS One*, *10*(9), e0137334. https://doi.org/10.1371/journal.pone.0137334

Smith, J., & Minda, J. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1411–1436. https://doi.org/10.1037/0278-7393.24.6.1411

Smith, J., & Minda, J. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 3–27. https://doi.org/10.1037/0278-7393.26.1.3

Smith, L., & Gasser, M. (2005). The Development of Embodied Cognition: Six Lessons from Babies. *Artificial Life*, *11*(1–2), 13–29. https://doi.org/10.1162/1064546053278973

Spong, M. W., & Fujita, M. (2011). Control in Robotics. In T. Samad & A. Annaswamy (Eds.), *The Impact of Control Technology*. IEEE Control Systems Society. Retrieved from http://www.ieeecss.org/sites/ieeecss.org/files/documents/IoCT-Part1-04Robotics.pdf

Srinivas, M., & Patnaik, L. (1994). Genetic algorithms: a survey. *Computer*, *27*(6), 17–26. https://doi.org/10.1109/2.294849

Sternberg, R., & Sternberg, K. (2012). *Cognitive Psychology* (6th ed.). Belmont: Wadsworth.

Suchard, M. A., Wang, Q., Chan, C., Frelinger, J., Cron, A., & West, M. (2010). Understanding GPU Programming for Statistical Computation: Studies in Massively Parallel Massive Mixtures. *Journal of Computational and Graphical Statistics*, *19*(2), 419–438. https://doi.org/10.1198/jcgs.2010.10016

Suri, S. (2006). *Bipartite Matching & the Hungarian Method*. Retrieved from http://athena.nitc.ac.in/kmurali/Courses/CombAlg2013/suri.pdf

Suzumori, K., Endo, S., Kato, N., & Suzuki, H. (2007). A Bending Pneumatic Rubber Actuator Realizing Soft-bodied Manta Swimming Robot. In P. Dario & A. D. Luca (Eds.), *Proceedings 2007 IEEE International Conference on Robotics and Automation* (pp. 4975–4980). IEEE. https://doi.org/10.1109/ROBOT.2007.364246

Teh, Y. (2004). *Nonparametric Bayesian Mixture Models - release 2.1.* Retrieved from https://www.stats.ox.ac.uk/~teh/research/npbayes/npbayes-r21.tgz

Teh, Y. (2006). A Hierarchical Bayesian Language Model based on Pitman-Yor Processes. In M. Carpuat & K. Duh (Eds.), *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics* (pp. 985–992).

Teh, Y. (2010). Dirichlet Process. In C. Sammut & G. Webb (Eds.), *Encyclopedia of Machine Learning* (pp. 280–287). Springer US. https://doi.org/10.1007/978-0-387-30164-8

Teh, Y., & Jordan, M. (2010). Hierarchical Bayesian nonparametric models with applications. In N. Hjort, C. Holmes, P. Müller, & S. Walker (Eds.), *Bayesian Nonparametrics* (pp. 158–206). Cambridge and UK: Cambridge University Press.

Teh, Y., Jordan, M., Beal, M., & Blei, D. (2006). Hierarchical Dirichlet Processes. *Journal of the American Statistical Association*, *101*(476), 1566–1581. https://doi.org/10.1198/016214506000000302

Tenenbaum, J., Kemp, C., Griffiths, T., & Goodman, N. (2011). How to Grow a Mind: Statistics, Structure, and Abstraction. *Science*, *331*(6022), 1279–1285. https://doi.org/10.1126/science.1192788

The Mathworks. (2016). *How the Genetic Algorithm Works*. Retrieved from http://de.mathworks.com/help/gads/how-the-genetic-algorithm-works.html

Thrun, S. (2002). Robotic Mapping: A Survey. In G. Lakemeyer & B. Nebel (Eds.), *Exploring Artificial Intelligence in the New Millennium* (pp. 1–36). San Francisco: Morgan Kaufmann.

Thrun, S., & Leonard, J. (2008). Simultaneous Localization and Mapping. In B. Siciliano & O. Khatib (Eds.), *Springer Handbook of Robotics* (pp. 871–889). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-30301-5

Thrun, S., & Mitchell, T. (1995). Lifelong Robot Learning. In L. Steels (Ed.), *The Biology and Technology of Intelligent Autonomous Agents* (pp. 165–196). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-79629-6

Transeth, A. A., Pettersen, K. Y., & Liljebäck, P. al. (2009). A survey on snake robot modeling and locomotion. *Robotica*, *27*(07), 999–1015. https://doi.org/10.1017/S0263574709005414

Vanpaemel, W., & Lee, M. (2012). Using priors to formalize theory: Optimal attention and the generalized context model. *Psychonomic Bulletin & Review, 19*(6), 1047–1056. https://doi.org/10.3758/s13423-012-0300-4

Veksler, V. D., Myers, C. W., & Gluck, K. A. (2015). Model flexibility analysis. *Psychological Review, 122*(4), 755–769. https://doi.org/10.1037/a0039657

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction*, *4*(1), 45–69. http://dx.doi.org/10.1016/0959-4752(94)90018-3

Waldmann, M. (2017). Kategorisierung und Wissenserwerb. In J. Müsseler & M. Rieger (Eds.), *Allgemeine Psychologie* (pp. 357–399). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-53898-8

Wang, C. (2015). *online-hdp*. Retrieved from https://github.com/blei-lab/online-hdp

Wang, C., & Blei, D. (2012). *A Split-Merge MCMC Algorithm for the Hierarchical Dirichlet Process*. Retrieved from https://arxiv.org/abs/1201.1657

Wang, C., & Blei, D. (2017). *Hierarchical Dirichlet Process (with Split-Merge Operations)*. Retrieved from https://github.com/blei-lab/hdp

Wang, C., Paisley, J., & Blei, D. (2011). Online Variational Inference for the Hierarchical Dirichlet Process. In G. Gordon, D. Dunson, & M. Dudík (Eds.), *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics* (pp. 752–760). Brookline, MA: Microtome Publishing.

Welch, B. L. (1947). The Generalization of `Student's' Problem when Several Different Population Variances are Involved. *Biometrika*, *34*(1/2), 28–35. https://doi.org/10.2307/2332510

Welding, S. (2013). *Sind Wittgensteins Bemerkungen über Sprachspiele, Spiele und Familienähnlichkeiten philosophisch instruktiv?* (Braunschweig : Seminar für Philosophie, 2013, Ed.). Braunschweig. Retrieved from http://nbn-resolving.de/urn:nbn:de:gbv:084-13102311349

Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. *1960 WESCON Convention Record*, 96–104.

Wills, A. J., & Pothos, E. M. (2012). On the adequacy of current empirical evaluations of formal models of categorization. *Psychological Bulletin*, *138*(1), 102–125. https://doi.org/10.1037/a0025715

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636. https://doi.org/10.3758/BF03196322

Wilson, R., & Clark, A. (2009). How to situate cognition: Letting nature take its course. In M. Aydede & P. Robbins (Eds.), *The Cambridge Handbook of Situated Cognition* (pp. 55–77). Cambridge: Cambridge University Press.

Yamauchi, T., Love, B., & Markman, A. (2002). Learning nonlinearly separable categories by inference and classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 585–593. https://doi.org/10.1037/0278-7393.28.3.585

Yamauchi, T., & Markman, A. (1998). Category learning by inference and classification. *Journal of Memory and Language*, *39*(1), 124–148. https://doi.org/10.1006/jmla.1998.2566

Yousef, H., Boukallel, M., & Althoefer, K. (2011). Tactile sensing for dexterous in-hand manipulation in robotics—A review. *Solid-State Sensors, Actuators and Microsystems Workshop*, *167*(2), 171–187. https://doi.org/10.1016/j.sna.2011.02.038

Yuh, J. (2000). Design and Control of Autonomous Underwater Robots: A Survey. *Autonomous Robots*, *8*(1), 7–24. https://doi.org/10.1023/A:1008984701078

Zabinsky, Z. B. (2003). *Stochastic Adaptive Search for Global Optimization*. New York: Springer US.

Zaki, S., Nosofsky, R., Jessup, N., & Unverzagt, F. (2003). Categorization and recognition performance of a memory-impaired group: Evidence for single-system models. *Journal of the International Neuropsychological Society*, *9*(3), 394–406. https://doi.org/10.1017/S1355617703930050

Zhang, D., Hu, D., Shen, L., & Xie, H. (2006). A Bionic Neural Network for Fish-Robot Locomotion. *Journal of Bionic Engineering*, *3*(4), 187–194. https://doi.org/10.1016/S1672-6529(07)60002-X

Zimmerman, D. (1987). Comparative Power of Student T Test and Mann-Whitney U Test for Unequal Sample Sizes and Variances. *The Journal of Experimental Education*, *55*(3), 171–174. https://doi.org/10.1080/00220973.1987.10806451