

Diplomarbeit

**Liegen ähnliche Mechanismen
Kategorisierungsprozessen mit und ohne
Rückmeldung zugrunde?**

Jan Zwickel

Erstbetreuer: Prof. Joachim Funke

Zweitbetreuer: Dr. Klaus-Eckart Rogge

Jan Zwickel
Körnerstraße 11
74912 Kirchartt
Jan.Zwickel@urz.uni-hd.de
06221/653665

Erklärung

Hiermit versichere ich, dass ich die vorliegende Diplomarbeit selbständig verfasst und außer den aufgeführten keine anderen Quellen und Hilfsmittel verwendet habe. Soweit aus den im Literaturverzeichnis angegebenen Werken einzelne Stellen im Wortlaut oder Sinn nach entnommen wurden, ist dies in jedem Fall unter Angabe der Entlehnung kenntlich gemacht. Die Arbeit wurde noch nicht anderweitig für Prüfungszwecke vorgelegt.

Heidelberg, den 22. Oktober 2003

Jan Zwickel

Ich bin nicht ihr Stiller
Stiller in Max Frisch „Stiller“, S. 49

Danksagung

Mein Dank gilt Wolfgang Bösche, Timm Lochmann, Katrin Scharpf und Christoph Witzel für konzeptuelle und syntaktische Hinweise. Für Ideenanstöße bin ich Andy Wills dankbar. Weiteren Dank gebührt Matthias Jöst und Claudia Röhrich für das Korrekturlesen der Diplomarbeit. Schließlich geht mein Dank an Professor Joachim Funke und Dr. Klaus-Eckart Rogge für ihre unkomplizierte Annahme und Betreuung meines Diplomarbeitthemas.

Abstrakt

Es werden drei Experimente berichtet, die der Frage nachgehen, ob ein als „Blocking“ bezeichnetes Phänomen, die Vernachlässigung redundanter Information, auch in Klassifikationsparadigmen ohne Rückmeldung auftritt. Der Blocking-Effekt wird als typisches Merkmal fehlergetriebenen Lernens angesehen und daher gewöhnlich nur mit Kategorisierungsparadigmen mit Rückmeldung in Verbindung gebracht. Während sich die Ergebnisse der ersten Studie mittels zwei post-hoc Annahmen gut mit der Hypothese eines Blocking-Effektes auch in Klassifizierungsparadigmen ohne Rückmeldung vereinbaren lassen, kann die zweite Studie vermutlich eine Grenze dieses Effektes durch deliberate kognitive Prozesse aufzeigen. Die ersten zwei Experimente untersuchen den Blocking-Effekt bei drei Kategorien. Im dritten Experiment erhalten die Teilnehmer eine Entscheidungsmöglichkeit lediglich zwischen zwei Kategorien. Es wird versucht die empirischen Daten mittels Netzwerksimulationen zu deuten. Ein Hauptaugenmerk ruht dabei auf der Annahme eines Rückmeldeprozesses über die bisherige Konsistenz der Zuordnung. Um die empirischen Daten zu erklären, müssen keine spezifischen Annahmen hinsichtlich der genauen Verarbeitungsprozesse der Rückmeldung getroffen werden. Damit lässt sich die hier diskutierte Modellvorstellung gut in allgemeine Kategorisierungsmodelle mit Rückmeldung (z.B. Pearce & Hall, 1980, Pearce, 1994) einbauen. Auch die Modellvorstellungen von Kruschke und Johansen (1999) lassen sich mit den hier beschriebenen Modellvorschlägen vereinbaren. Es wird versucht eine Brücke zwischen Kategorisierungsmodellen mit und ohne Rückmeldung zu schlagen.

1	HINTERGRUND	1
1.1	Phänomenbeschreibung	4
1.1.1	Der Blocking-Effekt bei Tieren	4
1.1.2	Überraschungseffekte beim Menschen	5
1.2	Mögliche Erklärungsansätze	7
1.2.1	Das Rescorla-Wagner Modell	7
1.2.2	Delta-Netze als mögliche Implementationsform der Rescorla-Wagner Regel	24
1.2.3	Konfigurale Modelle	37
1.2.4	Kompetitives Lernen	39
1.3	Kategorisierungsmodelle	47
1.4	Klassifikationsmodelle	48
1.5	Die Delta Regel und Klassifikation ohne Rückmeldung	52
2	EXPERIMENTE	56
2.1	Pilotstudie	57
2.1.1	Methode	58
2.1.2	Ergebnisse und Diskussion	63
2.2	Experiment 1	65
2.2.1	Methode	65
2.2.2	Ergebnisse	65
2.2.3	Diskussion Experiment 1	74
2.2.4	Zusammenfassung Experiment 1	97
2.3	Experiment 2	97
2.3.1	Ergebnisse	98
2.3.2	Diskussion Experiment 2	104
2.3.3	Zusammenfassung Experiment 2	111
2.4	Experiment 3	112
2.4.1	Ergebnisse	113
2.4.2	Diskussion Experiment 3	121
2.5	Allgemeine Diskussion	127
3	ZUSAMMENFASSUNG	132

1 Hintergrund

Könnte der Mensch jedes Ereignis nur isoliert betrachten, so wäre er sehr schnell von der Fülle an möglichen Ereignissen überwältigt. Schafft er es hingegen Einzelereignisse in Kategorien einzuordnen, so kann er Erfahrungen von früheren Ereignissen auf aktuelle Situationen übertragen. Voraussetzung dafür ist allerdings, dass die Ereignisse in die richtigen Kategorien eingeordnet werden, um geeignete Handlungsauswahlen zu treffen. Diese grundlegende Eigenschaft durchzieht unser kognitives Handeln. Ereignisse, Handlungen und sogar Menschen werden in Kategorien mehr oder minder gepresst. Die Informationsreduktion erlaubt, wenn sie sinnvoll erfolgt, die geeignete Aktion zu vollziehen, ohne darauf angewiesen zu sein, den individuellen Reiz gesehen zu haben.

Doch nicht nur der Laie verwendet Kategorien um seinen Informationsbereich zu strukturieren und zu vereinfachen - auch Wissenschaftler greifen auf Kategorien zurück, um Ereignisse einzuordnen und Reaktionen vorherzusagen. Ein augenfälliges Beispiel stellt der Chemiker und sein Versuch, über Stoffklassen Reaktionseigenschaften vorherzusehen, dar. Biologen klassifizieren Tiere um über analoge Strukturen Funktionen zu erschließen. Selbst Physiker, welche den Ruf besitzen ihren Gegenstandsbereich hoch formalisiert zu bearbeiten, müssen auf abstrakte Weise festlegen, unter welchem Aspekt bzw. welcher Kategorie ein Objekt betrachtet werden soll. Festgelegt werden muss, ob das Objekt auf der mikroskopischen oder makroskopischen Ebene beschrieben werden soll, ob das Objekt als starr und kreisförmig oder als viskos und aus Einzelteilen bestehend angesehen wird. Ebenso spielt das Bilden von Kategorien eine wichtige Rolle in der psychologischen Forschung. Inwieweit Forschungsergebnisse verallgemeinerbar sind, steht und fällt letztendlich mit der Annahme, aus welcher Kategorie eine Beobachtung entstammt. Das Entwerfen eines Experimentes setzt voraus, dass für die allgemeine Aussage, die angestrebt werden soll, eine „Interpretation“ in Kategorien gefunden werden kann, welche im Experiment verwendet werden können. Diese Operationalisierung in eine Kategorie soll dann möglichst glaubhaft die „Idee“ der Aussage repräsentieren. Teilbereiche der Psychologie haben sich darauf spezialisiert, Kategorienbildung über Fremdpersonen bei Laien zu erforschen. Wenn man bedenkt, dass die richtige Einordnung in Freund und Feind über Krieg und Frieden entscheiden kann, so erkennt man die Relevanz. Bei der Behandlung von psychischen und körperlichen Krankheiten wird versucht durch Wahl geeigneter Kategorien, in diesem Fall Krankheitsbilder, wenn nicht über eine gemeinsame Ätiologie so doch wenigstens über eine gemeinsame Behandlungsweise zu abstrahieren.

Kant (1995) postulierte, dass Menschen nur in bestimmten Kategorien wie Raum und Zeit denken können. Auch Poppers Idee (1935), der Natur einzelne falsifizierbare Fragen zu stellen, welche im Experiment beantwortet werden sollen, impliziert, soll es nicht in ein solipsistisches, idiosynkratisches Forschen abgleiten, dass relevante Kategorien gebildet werden können. Erst dann kann eine Einzelbeobachtung, welche eine Hypothese widerlegt, eine ganze Klasse, aus der die Einzelbeobachtung zu stammen scheint, als unwahrscheinlich erscheinen lassen. Erst durch diese Einordnung in Kategorien kann also von der Einzelbeobachtung auf eine Klasse generalisiert werden. Aufgrund dieser Relevanz verwundert es nicht, dass eine reichhaltige Literatur über den Erwerb von Kategorien in der Psychologie vorhanden ist. Allerdings wird dabei meistens davon ausgegangen, dass der Kategorienerwerb mit regelmäßiger Rückmeldung erfolgt (z.B. Gluck & Bower, 1988a; Chapman & Robbins, 1990). Das Paradigma, in dem der Erlernende Rückmeldung über seine Umwelt erhält, stellt aber eine augenscheinlich nicht gerechtfertigte Annahme über die Form des Informationsgehaltes der Umwelt dar.

In dieser Studie soll deshalb untersucht werden, inwieweit sich ein gemeinsames Modell für Lernen mit und ohne Rückmeldung empirisch rechtfertigen lässt. Ein gemeinsames Modell hätte neben der allgemeinen Vereinfachung auch die praktische Konsequenz, dass Lernende nicht im Voraus darüber entscheiden müssten, ob die Lernaufgabe ein Lernen mit oder ohne Rückmeldung darstellt. Dies ist deshalb wichtig, weil in realistischen Situationen nicht immer bereits am Anfang des Lernvorganges entschieden werden kann, ob für erfolgreiche Handlungen Rückmeldungen gegeben werden. Kann ein gemeinsames Modell für Kategorienlernen mit und ohne Rückmeldung etabliert werden, so muss nicht explizit entschieden werden welche Lernprozesse aktiviert werden müssen, da bei Auftreten von Rückmeldungen diese einfach integriert werden können.

Der Testfall zwischen unterschiedlichen Modellvorstellungen stellt in dieser Studie eine Situation mit redundanter Information dar, in welcher ein selektiver Lernprozess bei einer Klassifikationsaufgabe *ohne* Rückmeldung beobachtet werden soll. Das Auftreten eines selektiven Lerneffektes würde eine weite Klasse an Modellen für Klassifikationslernen als unplausibel erscheinen lassen. Kategorienlernmodelle¹ könnten hingegen diesen selektiven Lerneffekt erklären und somit den Wert einer gemeinsamen Modellannahme bekräftigen.

¹ In Anlehnung an Chater (2001) wird im Folgenden der Ausdruck „Klassifikation“ für Kategorisierungsprozesse *ohne* Rückmeldung und „Kategorisierung“ für Kategorisierungsprozesse *mit* Rückmeldung verwendet. Aus Gründen der Deutlichkeit wird aber manchmal dennoch der Zusatz „mit“ oder „ohne Rückmeldung“ verwendet.

Diese gegeneinander zu testenden Modellklassen sind dabei weniger idiosynkratisch gewählt als es auf den ersten Blick erscheinen mag. Vielmehr dürfte sich die grundlegende Differenz zwischen einem Algorithmus, der den Neuigkeitswert einer Information integriert, und einem Algorithmus, der unabhängig von der Überraschung, bzw. Wissen über kovariierende Reize eine neue Beobachtung sozusagen „automatisch“ lernt, auch leicht auf andere konnektionistische und symbolische Ansätze übertragen lassen. In diesen beiden Modellvorstellungen, zwischen denen durch die unten geschilderten Experimente entschieden werden soll, spiegelt sich eine alte Auseinandersetzung zwischen zwei psychologischen Erklärungsansätzen wider. Der simplen Häufigkeitsanalyse oder Kontiguität wie sie Hebb vertrat (Hebb, 1949) und die von kompetitiven Netzwerken berechnet wird, steht die Kontingenz oder Zusammenhangsanalyse wie sie vom Rescorla-Wagner Modell (Rescorla & Wagner, 1972) bzw. der Delta-Regel (Stone, 1986) beachtet wird gegenüber.

Diese Unterscheidung wird weniger auf der Implementations- als auf der Algorithmenebene untersucht werden (Marr, 1982). Es soll also zwischen zwei Klassen (oder auch Kategorien) von Modellen entschieden werden, indem die Unvereinbarkeit der zwingenden Eigenschaften eines dieser Modelle mit empirischen Daten gezeigt wird. Über diese Algorithmenebene der Modellunterscheidung sollte auch nicht die Tatsache hinwegtäuschen, dass ein Implementationsversuch in ein konnektionistisches Modell vorgenommen wird. Dies erfolgt mehr aus Gründen der Anschaulichkeit, sowie aus der Überzeugung, dass konnektionistische Modelle mit ihren strengen Aussagen schnell Umsetzungsprobleme bei fehlerhaften theoretischen Ansätzen offenbaren. Meines Erachtens stellen konnektionistische Systeme das geeignete Implementationsmittel dar, da bei diesen Systemen die beobachteten Eigenschaften aus grundlegenden Prinzipien entstehen. Durch diese theoretische Stringenz verhindern konnektionistische Systeme post-hoc Modellanpassungen weitestgehend. Wo prinzipielle Offenheit der Veränderung besteht, wie z.B. in der Wahl eines Lernparameters, wird darauf geachtet, dass diese Wahl in weiten und plausiblen Bereichen keine qualitativen Unterschiede hervorruft. Aufgrund der empirischen Daten muss die Modellvorstellung später leicht modifiziert werden². Dafür erlauben die Daten zwischen unterschiedlichen Modellen aus der „Klasse“ der durch Fehler (Rückmeldung) getriebenen Lernsysteme zu unterscheiden.

² Dies scheint im Widerspruch zu der oben erwähnten „Immunität“ des konnektionistischen Ansatzes gegen post-hoc Anpassungen zu stehen. Allerdings handelt es sich bei der hier erwähnten Anpassung um eine durch eine geänderte Modellvorstellung *inhaltlich* relevante Veränderung des konnektionistischen Modells. Gemeint im Text war, dass die assoziative Grundvorstellungen die Modellveränderungen *beschränken* und nicht jede Modellveränderung mit einer konnektionistischen Vorstellung vereinbar ist.

1.1 Phänomenbeschreibung

„Blocking“, auch „Effekt redundanter Vorhersagen“ oder „Überraschungseffekt“ genannt, stellt ein ideales Phänomen dar, um zwischen Modellen, welche auf einer Form von fehlergetriebenem Lernen beruhen und Modellen, welche unabhängig von Fehlerrückmeldungen lernen, zu unterscheiden. Die ersten Studien zu diesem Phänomen fanden im Tierreich statt.

1.1.1 Der Blocking-Effekt bei Tieren

In einem mittlerweile klassischen Experiment untersuchte Kamin (1969), ob Ratten sich in ihrem Verhalten durch einen potentiell konditionierenden Stimulus beeinflussen lassen, falls dieser nur redundante Information enthält. Die Redundanz eines Lichtreizes wurde dadurch erreicht, dass ein Elektroschock gleichzeitig durch einen Ton angekündigt wurde. Die abhängige Variable stellte die Unterlassung eines zuvor positiv verstärkten Verhaltens bei Vorliegen der Warnreize dar. Die Ratten mussten also lernen die Situationen in mindestens zwei Kategorien zu unterteilen. In eine, in der sie ungestört ihrem Verlangen nach Futter nachgehen konnten und eine, in der ein Elektroschock erfolgen würde und in der deshalb eine Verhaltensunterdrückung gezeigt wurde.

In einer Vorphase wurde eine Kontingenz zwischen dem Drücken eines Hebels und darauf folgender Futtergabe etabliert. Nachdem ein stabiles Reaktionsschema erlernt worden war, wurde in einer ersten Phase zu verschiedenen Zeitpunkten ein Ton dargeboten. Gekoppelt an die Darbietung des Tones war ein Schock unabhängig davon, ob in dieser Zeit der Futterhebel gedrückt wurde oder nicht. Aufgrund der aversiven Konditionierung durch einen Elektroschock zeigten die Ratten während der Tonpräsentation Vermeidungsverhalten und drückten seltener den Hebel als an Zeitpunkten zu denen kein Ton dargeboten wurde. In einer zweiten Phase wurde anstelle eines einzelnen Tones, gleichzeitig ein Ton mit einem Lichtreiz dargeboten. Mit den gleichen aversiven Folgen, die der Ton ohne Lichtreiz in der ersten Phase hatte. Schließlich wurde in der dritten Phase überprüft, ob der Zusammenhang zwischen Licht und Elektroschock gelernt worden war. Die Kontrollgruppe der Studie unterschied sich nur darin von der Experimentalgruppe, dass sie die erste Phase übersprang. Während in der Experimentalgruppe die Präsentation des Lichtes allein zu keiner bedeutenden Suppression der Hebelbenützung führte, konnte in der Kontrollgruppe eine signifikant seltenere Betätigung des Hebels während der Lichtzeiten nachgewiesen werden (vgl. Abbildung 1.1).

	Phase 1	Phase 2	Phase 3	Ergebnis
Experimental-Gruppe :	T→Schock	TL→Schock	L→Schock	keine Suppression
Kontroll-Gruppe :		TL→Schock	L→Schock	Suppression

Abb. 1.1: Blocking-Experiment (Kamin, 1969). T symbolisiert einen Ton, L einen Lichtreiz. Die Wirkung des darauffolgenden elektrischen Schocks wurde durch Verhaltensvermeidung gemessen.

Kamin interpretierte diesen Befund dahingehend, dass die Ratten vorwiegend dann die Verbindungen zwischen einem Stimulus und einem Ereignis erlernten, wenn das Ereignis *überraschend* war. Für die Experimentalgruppe war der Elektroschock in der ersten Phase noch überraschend. Deshalb wurde die Verbindung zwischen Ton und Elektroschock gelernt. In der zweiten Phase hingegen war der Schock nicht mehr überraschend, da die Bestrafung aufgrund des Tonreizes erwartet wurde. Als Konsequenz wurde der Zusammenhang zwischen dem Lichtreiz und der Bestrafung nicht gelernt, der Lichtreiz wurde *geblockt*. In der Kontrollgruppe hingegen war der Schock in der zweiten Phase noch überraschend, weshalb sowohl die Verbindung zwischen Ton und Bestrafung als auch zwischen Licht und Bestrafung erlernt wurde. Diese Erklärung für das selektive Lernphänomen wurde im Wesentlichen von Autoren anderer Tierstudien übernommen (z.B. Mackintosh, 1975; Pearce & Hall, 1980; Rescorla & Wagner, 1972). Für eine Übersicht siehe Miller, Barnet und Grahame (1995) oder Walkenbach und Haddad (1980). Dieser Effekt ebnete den Weg von der Vorstellung einer starren Stimulus-Response Verbindung (Skinner, 1957) zu einer flexibleren Vorstellung über den Zusammenhang zwischen Reiz und Reaktion, die weitere nicht im Stimulus enthaltene Bewertungskriterien wie z.B. in diesem Fall Überraschung enthält. Diese Vorstellung wurde 1972 von Rescorla und Wagner formalisiert und wird in Abschnitt 1.2.1 diskutiert.

1.1.2 Überraschungseffekte beim Menschen

Diese Form des Lernens in Abhängigkeit vom Überraschungswert der Information zeigte sich auch in vielen Studien, in denen Menschen die Untersuchungspopulation bildeten. So demonstrierten z.B. Lachnit, Kimmel, Bevill, Martin, Levey und Hamm (1990) Blocking bei Menschen mittels Konditionierung eines Lidschlagreflexes durch einen Luftstoß, der von Lichtreizen oder Tönen vorhergesagt wurde. Diese Studie ist vor allem deshalb bemerkenswert, weil sie eine der wenigen darstellt, die Blocking bei Menschen in einer Aufgabe mit geringen kognitiven Verarbeitungsanteilen zeigen konnte. Während für Blockingphänomene bei kognitiven Aufgaben auch die bewusste Entscheidung, nur

bestimmte Informationen zu nutzen, als Erklärung in Frage kommt, können hier Metakognitionen nahezu ausgeschlossen werden. In „kognitiveren“ Aufgaben zeigten z.B. Dickinson, Shanks und Evenden (1984) den Blocking-Effekt mittels eines Computerspiels. Die Teilnehmer konnten die Zerstörung eines Panzers auf versteckte Minen, welche die Panzer überqueren mussten, oder den eigenen Schuss zurückführen. In der Experimentalgruppe beobachteten die Versuchspersonen, zuerst ohne zu schießen, wie die Panzer aufgrund von versteckten Minen zerstört wurden. In der daran anschließenden Phase sollten die Versuchspersonen auf die Panzer schießen. Die Kontrollgruppe übersprang die Beobachtungsphase. In einer dritten Phase sollten die Versuchspersonen auf einer Ratingskala einschätzen, wie effektiv die eigenen Schüsse waren. Es zeigte sich, dass die Experimentalgruppe dadurch, dass sie in der Beobachtungsphase die Wirksamkeit der Minen erkennen konnte, von dem Resultat der eigenen Schüssen weniger überrascht wurde und deshalb den Schüssen eine geringere Effektivität beimaß.

Chapman und Robbins (1990) zeigten, dass der prädiktive Wert von Aktien bei einer Börsensimulation niedriger eingeschätzt wurde, wenn diese Aktien gleichzeitig mit anderen Aktien, welche bereits einen prädiktiven Wert erhalten hatten, einen Marktanstieg vorhersagten. Verglichen wurde die Einschätzung der Vorhersagekraft mit dem Vorhersagewert von Aktien, welche zusammen mit Aktien auftraten die sich als nicht prädiktiv für den Marktgewinn erwiesen hatten. Gluck und Bower (1988a) benutzten ein Paradigma in dem verschiedene Symptome mit unterschiedlicher Wahrscheinlichkeit eine von zwei Krankheiten vorhersagten. Wenn die Wahrscheinlichkeit einer Krankheit bei Vorliegen eines Symptoms erhöht wurde, so erniedrigte sich der Vorhersagewert eines anderen mit dieser Krankheit assoziierten Symptoms. Erfasst wurde diese Veränderung durch die Einschätzung der Wahrscheinlichkeit einzelner Krankheiten bei Vorliegen einzeln präsentierter Symptome. Die Erhöhung des prädiktiven Wertes einer Informationsquelle und somit die Verminderung der Überraschung hat also die Abwertung des Informationswertes einer anderen Informationsquelle für die selbe Zielvariable zur Folge. Kruschke (eingereicht) extendierte das Auftreten von Blocking-Phänomenen auch für Paradigmen, die eine lineare Extrapolation zwischen Ausprägungen bestimmter Hinweisreize und Kategorienzugehörigkeit erfordern, da die Ausprägungen der Testitems nicht in der Trainingsphase auftraten.

Auf das beobachtbare Phänomen bezogen scheinen die Blocking-Experimente mit Menschen zwar denen bei Tieren zu entsprechen, allerdings erscheint mir die implizite Annahme, dass auch die gleichen Erklärungsmodelle bei Tieren und Menschen greifen, in diesem Fall etwas kurz gegriffen. Während ich es für nachvollziehbar halte, dass Tiere sich

nicht für eine bestimmte Informationsquelle aus metakognitiven Gründen entscheiden, denke ich, dass die kognitive Einbettung der Aufgaben bei Menschen reflektiertere Reaktionsweisen hervorrufen. Beispielsweise dürfte das beobachtbare Verhalten, für die Entscheidung eine Informationsquelle zu benutzen und diese nur dann zu ergänzen, wenn die Strategie nicht zum Ziel führt, gut durch Fehlerminimierungsmodelle beschreibbar sein. Ob dies allerdings dem zugrundeliegenden Phänomen gerecht wird, erscheint mir zumindest fragwürdig. Eventuell sollte für diese Form der Aufgabenlösung durch Verwendung abstrakterer Theorieelemente eine weniger „systemnahe“ Interpretation, als die in den folgenden Abschnitten vorgestellte, gefunden werden. Ein generelles Anliegen dieser Diplomarbeit bestand darin, so wenig kognitive Strategien wie möglich zu evozieren bzw. kognitive Strategien als Mediatorvariable für Blockierung aufzufassen. Ein erster Hinweis auf die Dissoziierbarkeit auch anhand beobachtbarer Phänomene zwischen simplem, eher stimulusgetriebenem Blocking, und metakognitiv begründetem Blocking, tritt in der Interpretation des zweiten Experimentes auf.

1.2 Mögliche Erklärungsansätze

Frühe psychologische Theorien gingen davon aus, dass allein die zeitliche Paarung eines konditionierten Stimulus (CS) mit einem unkonditionierten Stimulus (US) zu einer assoziativen Verbindung führt. Diese Verbindung sollte laut Modellvorstellung unabhängig vom Kontext anderer Stimuli erfolgen (Hull, 1943; chapter XIII, Hebb, 1949 p. 62). Wie aber in den 60er Jahren vor allem Kamin's Blocking-Experiment zeigte, war Kontiguität der Stimuli nicht ausreichend, um Lernvorgänge zu beschreiben. Als wichtige Variable für das Verständnis von einfachen Lernvorgängen wurde deshalb die Kontingenz untersucht. Diese bewertet nicht nur das gemeinsame Eintreten von CS und US sondern ebenso die Häufigkeit mit der nur einer der beiden Reize auftritt. Mit dem Wechsel von einfachen Häufigkeitsanalysen des Zusammentreffens zweier Reize zur mathematischen Korrelation bzw. Kontingenz wurden theoretische Parameter wie die Verlässlichkeit eines CS als Prädiktor für einen US sowie die Redundanz eines Stimulus in das Modell aufgenommen.

1.2.1 Das Rescorla-Wagner Modell

Diese Idee der Kontingenz implementierte das Rescorla-Wagner Modell (Rescorla & Wagner, 1972). Im Modell stellt das Ausmaß der Lernveränderung eine lineare Funktion der Differenz zwischen einer Erwartung und einer Beobachtung dar. Es benutzt ein kompetitives Prinzip, mit welchem es sich gegen einfaches Hebbsches Lernen abgrenzt, da alle eingehenden Verbindungen zu einer Ausgabeinheit um eine begrenzte Gewichtsstärke konkurrieren. Lernen zwischen einem Reiz und einer Konsequenz wird deshalb nicht nur von diesem Reiz

und dessen Konsequenz bestimmt, sondern ebenfalls dadurch, welche Verbindungsstärke bereits von anderen Reizen zu der Konsequenz besteht. Durch das Kapazitätslimit der Gesamtverbindungsstärken entsteht ein Wettstreit zwischen den einzelnen Reizen um die Verbindungsstärke zur Konsequenz. Das Modell ermöglicht die Erklärung historischer³ Effekte zwischen einzelnen Reizen wie z.B. Blocking, was eine Erweiterung simpler behavioraler Stimulus-Response Modelle darstellt. Exzitatorische und inhibitorische Verbindungsstärken zwischen einem Reiz und einem US werden dabei als sich gegenseitig ausschließende Zustände betrachtet, welche durch eine entweder positive oder negative Verbindungsstärke erreicht werden. Das Rescorla-Wagner Modell versucht bei mehreren Informationsquellen die Informationen nach Reliabilität der Vorhersage sowie nach Informationsgehalt zu gewichten. Abbildung 1.2 stellt das Rescorla-Wagner Modell schematisch dar.

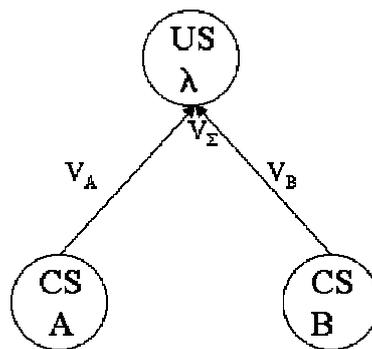


Abb. 1.2. : Schematische Darstellung des Rescorla-Wagner Modells. Die gesamte einlaufende Verbindungsstärke (V_{Σ}), wird mit der maximal zu erreichenden Verbindungsstärke λ verglichen.

Formale Eigenschaften des Rescorla-Wagner Modells

Sei V_A die Stärke der Verbindung eines Stimulus A zu einem unconditionierten Stimulus. So ist im einfachen Fall eines einzigen Reizes die Veränderung ΔV_A nach einer Präsentation des Reizes A gegeben durch:

$$\Delta V_A = (\lambda - V_A) \quad \text{Gleichung 1}$$

λ repräsentiert dabei das Ziel, welches aufgrund des Reizes als Ausgabe geliefert werden soll

³ Historische Erklärungen sind im Rescorla-Wagner Modell nur in soweit möglich als sie zu unterschiedlichen Verbindungsstärken einzelner Reize führen. Wie diese Stärken allerdings erreicht wurden spielt dabei keine Rolle. Die Abstraktionsebene verlagerte sich weg von den Stimuli auf die Verbindungsstärken.

bzw. die maximale Verbindungsstärke, die durch den jeweiligen US erreicht werden kann, V_A die bereits gelernte Verbindung zwischen Reiz A und dem US. Gewöhnlich wird $\lambda = 1$ gesetzt wenn ein Reiz vorliegt und $\lambda = 0$ wenn kein Reiz vorliegt. λ repräsentiert somit die Aktivationsstärke, welche asymptotisch (nach unendlich vielen Lerndurchgängen) gezeigt werden soll. Der gewählte asymptotische Wert von λ hat dabei keinen Einfluss auf die qualitative Aussage des Modells, da die Gewichte bis auf lineare Transformationen mittels λ eindeutig sind.

Am Anfang einer Paarung eines Reizes A mit einem US besteht noch keine Verbindung zwischen den Ereignissen und V_A ist 0. Da aber der US vorliegt wird $\lambda = 1$ gesetzt und eine relativ große Differenz zwischen λ und V_A tritt auf (vgl. Abbildung 1.2). Dieser „mathematische“ Fehler zwischen Vorhersage V_A und tatsächlichem Ereignis λ wird, in Anlehnung an das Lernparadigma von Kamin (1969), als *Überraschung* angesehen. Kamin's Begriff der Überraschung unterscheidet sich jedoch von dem der Überraschung im Rescorla-Wagner Modell: In der Modellvorstellung von Kamin führt jegliche Form der Überraschung, unabhängig davon, ob der US zu stark oder zu schwach vorhergesagt wurde, zu einer Verstärkung der Verbindung. Im Rescorla-Wagner Modell hingegen erfolgt Lernen abhängig vom *Vorzeichen* der Differenz durch Verstärkung *oder* Verminderung der Assoziationen. Da die Überraschung am Anfang einer Paarung noch sehr groß ist, wird ΔV_A ebenfalls groß und V_A verändert sich schnell in Richtung von λ . Diese Veränderung wird mit zunehmender Annäherung an λ verlangsamt, bis schließlich asymptotisch λ erreicht wird. Tritt nun nach Lernen des Zusammenhangs zwischen Reiz A und dem US bei Vorliegen von A der US nicht mehr auf, entsteht erneut eine Differenz welche V_A nun gegen 0 gehen lässt.

Wird in die Berechnung der Veränderung nur der Reiz A und der US einbezogen, so werden andere Reize und ihr eventuell prädiktiver Wert vernachlässigt. Um die Konditionierung sensitiv für Kontextreize zu machen, wurde das Modell so erweitert, dass eine feste Kapazität an Verbindungsstärken zwischen den Reizen aufgeteilt werden muss. Diese Kapazitätslimitierung lässt sich symbolisch leicht durch $V_A + V_B + \dots + V_i = \lambda$ für den asymptotischen Fall nach unendlicher Zeit beschreiben. Umgangssprachlich kann dieses Phänomen durch Konkurrenz zwischen den einzelnen Reizen um die Verbindungsstärke beschrieben werden. Formal ergibt sich:

$$\Delta V_A = (\lambda - V_{AB})$$

Gleichung 2

für Reiz A bei gleichzeitiger Präsentation mit Reiz B und

$$\Delta V_B = (\lambda - V_{AB})$$

Gleichung 3

für die Verbindungsstärke des Reizes B zum US.

Eine wesentliche Annahme des Modells besteht nun in der Verknüpfungsfunktion zwischen V_A und V_B zu der Gesamtvorhersage V_{AB} . Es wird angenommen, dass sich V_{AB} linear aus V_A und V_B ergibt.

Für dieses Modell kann gezeigt werden, dass es nach genügend Lerndurchgängen diejenige „Gewichtung“ der Verbindungsstärken der einzelnen Reize V_A und V_B findet, welche erlaubt die vier möglichen Kombinationen aus Reiz A und B so in zwei Klassen einzuteilen, dass das Auftreten oder Ausbleiben des US richtig vorhergesagt werden kann. Diese Aussage gilt aber nur für den Fall, dass die Aktivitäts-Muster der Stimuli (Kombinationen aus den Reizen A und B) hinsichtlich ihrer Vorhersage linear separierbar sind. Das bedeutet, dass die Vorhersage sich aus der linearen Kombination einzelner Musterelemente (Reize) ergibt. Sind die Muster nicht linear separierbar, so führt der Algorithmus asymptotisch zu Gewichten, welche den mittleren quadratischen Fehler zwischen den Rückmeldungen λ und $V_{\text{EingabeMuster}}$ minimieren. Dies gilt nicht nur für zwei Vorhersagereize sondern für beliebig viele (vgl. Gluck & Bower, 1988a).

Nicht-linear separierbare Probleme können dennoch unter Beibehaltung wesentlicher Annahmen des Rescorla-Wagner Modells gelöst werden. So kann zum einen durch eine geeignetere Kodierung und zum anderen durch zwischengeschaltete Elemente die Lösbarkeit durch den Rescorla-Wagner Algorithmus erreicht werden. Diese zwischengeschalteten Elemente werden in der Domäne von konnektionistischen Modellen auch mit „hidden units“ bezeichnet. Im Folgenden wird auf die erste Lösungsmöglichkeit, die Lösung durch orthogonale Kodierung, am Beispiel des nicht-linear separierbaren XOR-Problems eingegangen (vgl. Rumelhart, Hinton, Williams, 1986). Die Lösung durch „hidden units“ wird im Rahmen des Delta-Modells in Abschnitt 1.2.4 diskutiert.

Im XOR bzw. exklusiven Oder Problem geht es darum, dass eine Gesamtaussage als Wahr angesehen werden soll, wenn Aussage A oder Aussage B richtig ist, nicht aber wenn beide richtig sind oder beide falsch. Im Folgenden wird die Eigenschaft „ist wahr“ durch eine 1 und die Eigenschaft „ist nicht wahr“ durch eine 0 repräsentiert. Im XOR Problem soll nun auf (10) und (01) aus Klasse 1 sowie auf (00) und (11) aus Klasse 2 gleichreagiert werden. Gleichzeitig soll aber auf Stimuli (10) und (00), auf Stimuli aus verschiedenen Klassen also, unterschiedlich reagiert werden. Damit ist dieses Problem nicht-linear lösbar. Das Problem kann durch das Rescorla-Wagner Modell folgendermaßen beschrieben werden: Es werden

Verbindungsgewichte gesucht, welche, multipliziert mit den Eingangsaktivitäten für Muster aus der einen Klasse, einen Schwellenwert überschreiten, nicht aber für Muster aus der anderen Klasse. Ein US also für die eine Musterklasse vorhergesagt wird, aber nicht für die andere. Formal:

$$E_A * V_A + E_B * V_B > S \text{ für alle Muster aus Klasse 1}$$

$$E_A * V_A + E_B * V_B \leq S \text{ für alle Muster aus Klasse 2.}$$

In diesen Gleichungen nimmt E_A den Wert 1 an, falls das betreffende Muster das Merkmal A besitzt und 0, falls dieses Merkmal bei dem aktuellen Muster nicht vorliegt. Für E_B gilt Entsprechendes bezüglich Merkmal B. V_A bzw. V_B sind die jeweiligen Verbindungsstärken zwischen den Merkmalskodierungen A und B und dem Ausgabeknoten, welcher aufgrund der Schwelle S entscheidet.

Dadurch, dass das Rescorla-Wagner Modell die Eingabevektoren der Muster mit den Verbindungsgewichten multipliziert, legt es ein lineares Ähnlichkeitsmaß an. Dies bedeutet, einzelne Merkmale wie z.B. „1 an erster Stelle“ werden addiert, um herauszufinden in welche Klasse ein Muster gehört. Daher sind Stimuli in Klasse (00) / (11) untereinander unähnlicher als zu Stimuli aus der Klasse (10) / (01). Die Information „1 an Stelle 1“ allein hat somit noch keine Bedeutung für die Zuordnung, solange der Kontext, d.h. die zweite Information, unbekannt ist. Damit die Klassen unter einem linearen Ähnlichkeitsmaß ähnlicher werden, muss eine „höhere“ Repräsentation gewählt werden, welche über beide Indikatoren gleichzeitig eine Aussage macht wie z.B. „beide gleich“. Das heißt, die Kodierung eines Merkmals soll nicht unabhängig von den anderen Merkmalen erfolgen sondern gleichzeitig mehrere Merkmale berücksichtigen. Dies kann dadurch geschehen, dass eine weitere Einheit eingeführt wird, welche diesen Zustand (beide Indikatoren gleich) repräsentiert, oder dass jedes Muster als für sich stehend und unabhängig kodiert wird. Im ersten Fall wird also ein Zusammenhang geschaffen, welcher die Beziehung der Muster untereinander bei Betrachtung von mehr als einem einzelnen Indikator zeigt, im anderen Fall werden unter der Annahme maximaler Unabhängigkeit vier neue Eingabemuster geschaffen. Diese eher abstrakte Diskussion soll durch eine grafische Darstellung des Problems verdeutlicht werden. In Abbildung 1.3 ist auf Dimension 1 der Indikator A und auf Dimension 2 Indikator B aufgetragen. Falls Gewichtungsfaktoren existieren, welche multipliziert mit den

Merkmalskodierungen durch eine lineare Entscheidungsfunktion⁴ die Kategorisierung erlauben, so lässt sich eine Gerade finden welche die beiden Musterklassen trennt⁵.

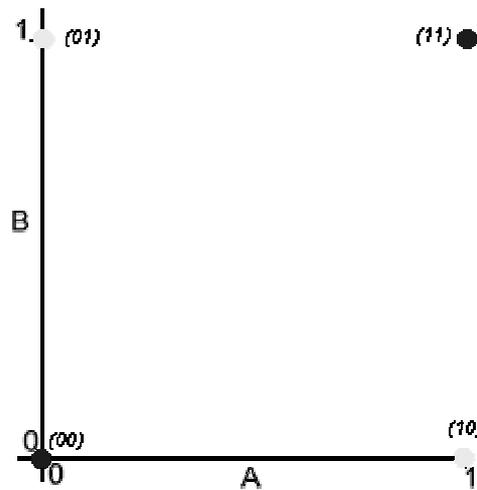


Abb. 1.3: Die 4 Eingaberepräsentationen lassen sich nicht durch eine Gerade in die zwei Gruppen, bestehend aus den Mustern (00)/(11) sowie (10)/(01), trennen.

Diese Trennung ist für das XOR Problem, wie aus Abbildung 1.3 ersichtlich, nicht mit einer Geraden möglich. Der Lösung, jedes Muster von A und B Kombinationen als eigenständiges Muster zu repräsentieren, würde einer vierdimensionalen Darstellung entsprechen, in der jede Dimension ein vollständiges Muster kodiert. Hier lässt sich zwar leicht eine Hyperebene finden (wenn auch nicht anschaulich so doch mathematisch), aber die Muster werden orthogonal kodiert und Lernen über die Folgen beispielsweise des Musters (10) beeinflusst nicht die Reaktion auf Muster (11).

Formaler kann das Problem folgendermaßen beschrieben werden: Gefunden werden sollen Gewichte, so dass für den Fall, in dem nur ein Merkmal vorliegt ((10) bzw. (01)), $V_A * E_A = \lambda = 1$ bzw. $V_B * E_B = \lambda = 1$ gilt. Für den Fall, dass beide Reize vorliegen soll $V_{AB} = 0$ gelten. Gleichzeitig lautet die Randbedingung $V_{AB} = V_A + V_B$. Dies macht die Lösung unmöglich. Werden nun alle Muster orthogonal kodiert, entsteht ein neues Merkmal für die beiden Merkmale A und B zusammen. Es ergeben sich folgende Gleichungen: $V_A = 1 = V_B = 1; V_{A \text{ und } B \text{ zusammen}} = 0$. Da die Randbedingung nicht mehr beschränkt, ist dies lösbar.

Für das Rescorla-Wagner Modell wurden bei geeigneter Kodierung verschiedene Äquivalenzen bewiesen (vgl. Abbildung 1.4), was das Modell mathematisch gut handhabbar macht. Sutton und Barto (1981) bewiesen, dass das Rescorla-Wagner Modell äquivalent zu

⁴ Wichtig ist dabei die Betonung der linearen Entscheidungsfunktion. Diese Aussage gilt genaugenommen nur für monotone Entscheidungsfunktionen. So ließen sich z.B. mit einer Cosinus-Funktion beide „Extremwerte“ (00) und (11) auf 1 Abbilden und alle anderen Werte auf Zahlen darunter.

⁵ Die Begründung dafür findet sich im Rahmen des Delta-Modells in Abschnitt 1.2.4.

der Widrow-Hoff Regel bzw. dem Delta-Modell ist. Stone (1986) bewies die Äquivalenz des weiter unten diskutierten Delta-Modells mit der linearen Regression. Bei der linearen Regression soll aufgrund von Eingabemustern die Vorhersage der Ausgabemuster unter Minimierung des mittleren quadratischen Fehlers erfolgen. Aus Eingabemustern (Reize x_i) sollen also Ausgabemuster (y) durch lineare Gewichtung (β_i) gefunden werden: $y = \sum_i \beta_i x_i$.

Bei mehreren abhängigen Variablen (mehreren USs) müssen verschiedene lineare Regressionen berechnet werden. Im Rescorla-Wagner Modell sollen Gewichtungen V_i gefunden werden, welche unterschiedliche Summen für verschiedene Musterklassen von Ereignissen produzieren. Diese Äquivalenz gilt natürlich nur für asymptotisches Lernen im Delta-Netzwerk bzw. bei unendlich häufiger Paarung des CS mit dem US im Rescorla-Wagner Modell. Für diesen Fall gleichen die Gewichte der Regression dann denen des unten diskutierten Deltamodells bzw. den Verbindungsstärken V_x des Rescorla-Wagner Modells.

Chapman und Robbins (1990) zeigten die asymptotische Äquivalenz zur Δp Regel. Die Δp Regel stellt ein „objektives“ Maß für den Zusammenhang zwischen einzelnen Reizen und Ereignissen dar. Für den Fall nur eines Prädiktors berechnet Δp die bedingte Wahrscheinlichkeit des Auftretens eines Ereignisses (O) bei Vorliegen eines Reizes (A) minus der bedingten Wahrscheinlichkeit für das Auftreten des Ereignisses bei Nichtvorliegen des Reizes (\bar{A}): $\Delta p = P(O|A) - P(O|\bar{A})$ (vgl. Shanks 1995).

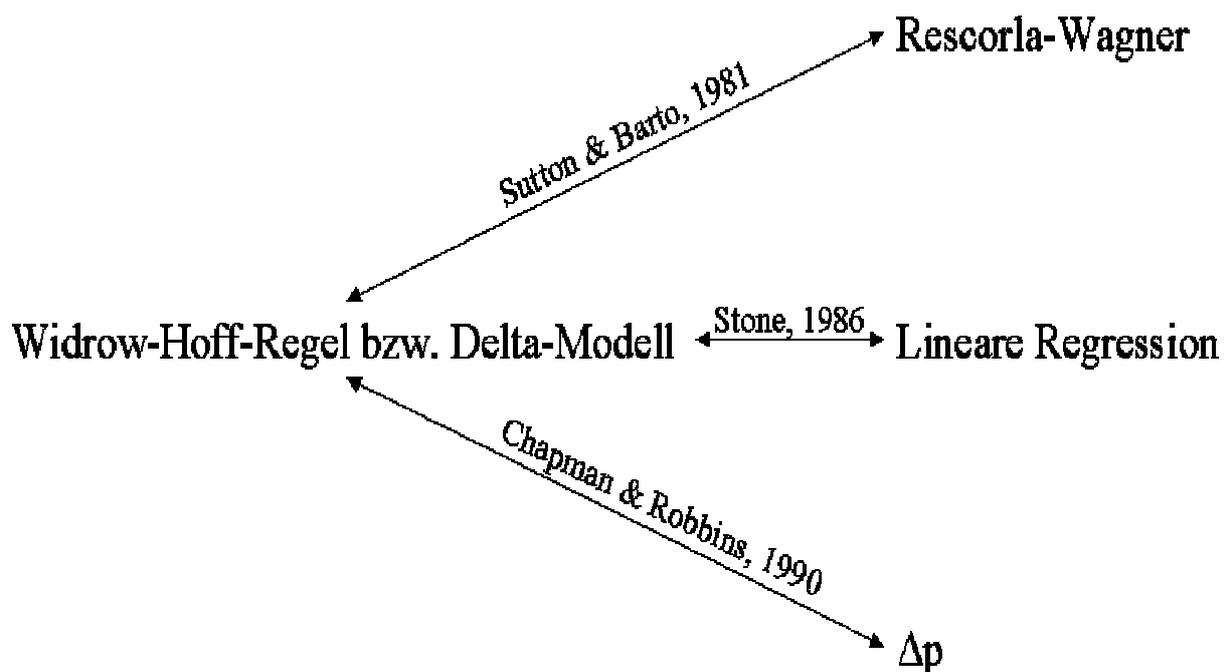


Abb. 1.4: Zusammenhang zwischen verschiedenen Modellen. Nach Bösche (2002).

Rescorla und Wagner (1972) nahmen zwei weitere Parameter, α und β , in ihre Gleichung auf. α ist ein vom jeweiligen Einzelreiz abhängiger Lernparameter, welcher von den Autoren als Salienz gedeutet wurde: „The value of alpha roughly represents stimulus salience and indicates our assumption that different stimuli may require associative strength at different rates despite equal reinforcement.“ (Rescorla & Wagner, 1972, p.76). β wurde als Lernparameter eingeführt, der vom US abhängig ist. Als Wertebereich für α und β wurde festgelegt:

$$0 \leq \alpha, \beta \leq 1.$$

Damit ergibt sich die Formel zur Veränderung der Assoziationsstärke des Reizes A, der gleichzeitig mit dem Reiz B dargeboten wird zu:

$$\Delta V_A = \alpha_A \beta_{US} (\lambda_{US} - V_{AB}) \quad \text{Gleichung 4}$$

bzw. Für die Verbindungsstärke B mit US :

$$\Delta V_B = \alpha_B \beta_{US} (\lambda_{US} - V_{AB}) \quad \text{Gleichung 5}$$

Dass hier λ ebenfalls durch den US indiziert wird, hat den Grund, dass sich der erwartete Wert von λ bei Nichtvorliegen oder Betonung eines US (Schock doppelter Stärke) ändert. In dieser ursprünglichen Formulierung des Modells wurde leider kein formales Vorgehen zur Bestimmung der α und β Werte vorgeschlagen, lediglich der Hinweis gegeben, dass β bei Vorliegen eines Reizes viel höher sein sollte als bei Fehlen des Reizes. Damit soll gewährleistet sein, dass die Extinktion langsamer als das Erlernen eines positiven Zusammenhangs erfolgt. Verschiedene formale Diskussionen des Salienzparameters erfolgten später durch z.B. Mackintosh (1975), Pearce & Hall (1980) aber auch Wagner (1978).

Formalisierungen des Salienzbegriffes in Erweiterungen des Rescorla-Wagner Modells

Der Salienzbegriff wird in verschiedenen Erweiterungen des Rescorla-Wagner Modells auf unterschiedliche Weise formalisiert. Diese Modelle erklären also die Lerneigenschaften nicht durch eine gleichbleibende Repräsentation des CS sondern durch eine variierende Repräsentationsstärke bzw. Assoziierbarkeit des CS. Bereits Wagner und Rescorla (1972) erkannten an, dass der Parameter α sich bei wiederholter Präsentation eines CS verändern kann. Allerdings beließen es die Autoren bei der allgemeinen Feststellung, dass wiederholte

Exposition eines CS die Salienz desselben vermindern sollte. Mackintosh (1975) hingegen sieht die Salienz und damit Assoziierbarkeit eines Reizes mit einer Reaktion durch zwei Faktoren bestimmt:

Zum einen wird die Salienz durch die physikalischen Eigenschaften des Reizes und der psychophysischen Ausstattung des sensorischen Apparates bestimmt. Auf diesen Faktor geht Mackintosh (1975) nicht weiter ein. Zum anderen können die Reize aber auch im Laufe der Lernerfahrungen aufgrund ihrer unterschiedlichen Vorhersagekraft hinsichtlich des Ereignisses eine Salienzänderung durchlaufen. Die Salienz eines Reizes A soll in dem Fall erhöht werden, in dem seine Vorhersage wesentlich die Differenz zwischen Erwartung und Realität vermindert. Die Erwartung ergibt sich dabei aus der Vorhersage der restlichen Reize:

$$\Delta\alpha_A \text{ sei positiv wenn gilt } |\lambda - V_A| < |\lambda - V_X| \quad \text{Gleichung 6}$$

$$\Delta\alpha_A \text{ sei negativ wenn gilt } |\lambda - V_A| \geq |\lambda - V_X| \quad \text{Gleichung 7}$$

$$\text{Wobei gilt: } V_X = \sum_{i, i \neq A} V_i,$$

α_A die Salienz des Reizes A, λ die Zielvorhersage, V_A die Verbindungsstärke des Reizes A zu dem Ausgabeknoten, bzw. der Vorhersage des US und V_X die Verbindungsstärke aller Reize außer A. Lernen erfolgt nach Gleichung 4.

Die Änderungen von α für einen Stimulus bestimmt damit nicht direkt die Salienz für einen anderen Stimulus. Obwohl die Salienz als nicht kapazitätslimitiert aufgefasst wird, erfolgt dennoch eine schnelle Abnahme, wenn durch einen anderen Stimulus bereits eine gute Vorhersage erreicht wird. Begründet werden diese Annahmen durch Beobachtungen wie z.B., dass in Blocking-Experimenten in der zweiten Phase, die ersten Paarungen zwischen dem theoretisch geblockten Reiz und dem US noch gelernt werden. Dies widerspricht zwar den Vorhersagen des Rescorla-Wagner Modells, kann aber mit dem oben postulierten Salienzmodell erklärt werden: Aufgrund der relativ hohen Salienz des neuen Stimulus B im ersten Durchgang der zweiten Phase kann eine Verbindung zwischen B und dem US aufgebaut werden. Allerdings verliert B aufgrund Formel 6 schnell seine Salienz, da V_B nahe 0 und V_A damit näher an λ ist⁶. Dies führt zur empirisch beobachteten, schnellen Abnahme des Lernens der Verbindung zwischen B und dem US. Um Daten von Wagner, Logan, Haberlandt und Price (1968, nach Mackintosh, 1975) zu erklären, diskutiert Mackintosh die

⁶ Diese Erklärung benötigt genaugenommen zwei Annahmen, die Mackintosh nicht erwähnt. Erstens darf das Lernen der Verbindung A-US nicht asymptotisch erfolgt sein, da sonst B überhaupt keine Verbindungsstärke mehr aufbauen könnte und zweitens, dass neu eingeführte Stimuli eine hohe Salienz besitzen.

Möglichkeit, dass auch die Performanz durch die Salienz beeinflusst wird und somit eine Reaktionswahrscheinlichkeit durch die Abnahme der Salienz geringer wird. Wagner und Kollegen zeigten, dass ein weniger valider Stimulus nicht nur keine weitere Verbindungsstärke mehr aufbauen kann, sondern sogar anscheinend, gemessen über die Performanz, an Verbindungsstärke verliert. Dies versucht Mackintosh auf die geringeren Performanzwahrscheinlichkeit aufgrund der kleineren Salienz und nicht auf eine Abnahme der Verbindungsstärke zurückzuführen.

Die Erweiterung des Rescorla-Wagner Modells durch Wagner (1978) versucht eine stärkere psychologische Interpretation der mathematischen Gleichungen zu ermöglichen, indem die angenommenen Prozesse in strukturelle Einheiten angesiedelt werden. Wenn auch Unklarheiten darüber bestehen, inwieweit dieses Modell angemessen für Verarbeitungsvorgänge von Tieren angesehen werden kann, so stellt es jedenfalls einen Versuch dar, mit einem Modell den Zusammenhang zwischen der mathematischen Modellierung und dem Gegenstandsbereich zu knüpfen. Ungeachtet von der Plausibilität der „Implementation“ wird im Folgenden der formale Algorithmus fokussiert: Die Veränderungsstärke der Gewichte ergibt sich aus drei Faktoren. Der erste Faktor stellt ein stimulusabhängiger Salienzparameter (α) dar. Der zweite Faktor bezieht sich auch auf die Assozierbarkeit des CS. Er ergibt sich aus der Differenz zwischen der maximalen Konditionierungsstärke des CS und der Vorhersage des CS aufgrund von Kontextreizen sowie früheren, im Kurzzeitgedächtnis überdauernden, Präsentationen des CS. Der dritte Faktor schließlich bezieht die Differenz zwischen der maximalen Assoziationsstärke des US und seiner Vorhersage durch frühere, im Kurzzeitgedächtnis überdauernder Aktivierungsstärke, und der Vorhersage durch den CS ein:

$$\Delta V_A = \alpha(1-r)\beta(\lambda-R)$$

Gleichung 8

λ gibt die maximal unterstützte Verbindungsstärke des CS an, r die Repräsentation des CS aufgrund von Kontextreizen und früheren Präsentationen des CS, sowie R die Repräsentation des US aufgrund früherer US Präsentationen und dem CS.

Nach Wagners (1978) Modellvorstellung erfahren unerwartete Ereignisse häufiger als erwartete Ereignisse einen Rehearsal-Prozess und üben deshalb einen stärkeren Einfluss auf das Antwortverhalten aus. Vermittelt durch diesen Rehearsal-Prozess beeinflusst die Diskrepanz zwischen Erwartung im Kurzzeitgedächtnis und eingetretener Verstärkung das

Antwortverhalten. Durch diese Annahmen lassen sich Phänomene wie z.B. "latent inhibition" erklären, in denen die unverstärkte Präsentation eines Reizes A zu einer Verminderung der Lernstärke in später verstärkten Durchgängen führt: Durch die Verknüpfung des CS mit dem Kontext des Versuches auch bei unverstärkter Darbietung wird der CS durch den Kontext vorhersagbarer. Aufgrund des zweiten Faktors in Gleichung 8 wird deshalb die Assozierbarkeit der CS vermindert. Allerdings erfolgt diese Erklärung um den Preis eines weiteren Parameters.

Die formale Abhandlung des Salienzparameters bei Pearce und Hall (1980) und Mackintosh (1975) widersprechen sich diametral. Während Mackintosh annimmt, dass die Salienz proportional zur Vorhersagekraft eines Reizes wächst, nimmt die Salienz in der Formulierung von Pearce und Hall (1980) dann ab, wenn der Stimulus den US gut vorhersagen kann. Das Mackintosh Modell dürfte sich daher gut bei gleichbleibender Verstärkungskontingenz eignen, während das Pearce und Hall Modell in einer sich ändernden Umwelt jeweils das mögliche Potential von noch nicht genutzten Reizen bewahrt. Die Salienz des Reizes A im Durchgang n bestimmt sich durch die Diskrepanz im vorausgehenden Durchgang n-1 zwischen der Vorhersage des US durch aufgetretene Merkmale und seiner tatsächlichen Ausprägung:

$$\alpha_A = |\lambda^{n-1} - \sum V_i^{n-1}| \quad \text{Gleichung 9}$$

die Veränderung der Verbindungsstärke folgt aus:

$$\Delta V_A = S_A \alpha_A \lambda^n = S_A |\lambda^{n-1} - \sum V_i^{n-1}| \lambda^n \quad \text{Gleichung 10}$$

S_A symbolisiert Intensität des CS A. Aus Gründen der Übersichtlichkeit wurde bei α_A und ΔV_A auf die Indizierung durch "n" verzichtet.

Die Stärke des Verbindungszuwachses hängt in diesem Modell also auch von der Intensität des gerade gezeigten US (λ^n) ab. Die bessere Repräsentation eines US durch die Vorhersage aufgrund eines CS führt in dieser Modellvorstellung zu einer schwächeren Veränderung der Verbindungsstärken. Eine Motivationsquelle für Pearce und Hall das Mackintosh (1975) Modell abzulehnen war ein experimenteller Befund von Hall und Pearce. Hall und Pearce (1979, nach Pearce & Hall, 1980) zeigten, dass ein nur leicht positiv verstärkter Reiz, trotz anwachsender Verbindungsstärke zur Reaktion, eine Abnahme der Salienz erfährt. Dieser Befund widerspricht den Vorhersagen des Mackintosh Modells. Im Experiment wurde ein Stimulus mit einem schwachen elektrischen Schock gepaart. Das Erlernen der Verbindung

zwischen diesem Stimulus und einem stärkeren Schock in einer zweiten Phase verlief langsamer als in der Kontrollgruppe, welche einen anderen Stimulus in der ersten Phase mit dem leichten Schock erhalten hatte.

Problematisch für die Modellvorstellung von Pearce und Hall (1980) ist vor allem der Rückgriff auf die CS und US Repräsentationsstärken aus dem vorhergehenden Lerndurchgang um die Salienz im laufenden Durchgang zu berechnen. Dadurch wird ein extra Speicher für alle CS erforderlich. Dies macht das Modell nicht nur aufgrund der eventuell großen Anzahl benötigter Speicherelemente sondern auch aufgrund der „Auslagerung“ des Wissens aus den assoziativen Verbindungen wesentlich komplizierter. Außerdem erfordert die Aktualisierung der Speicher einen äußeren Kontrollprozess, der meiner Meinung nach nur schwierig mit assoziativen Modellen vereinbar ist. Weiterhin erlebt das Modell durch diese „Übernahme“ aus dem vorhergehenden Durchgang eine undefiniertheit im Startzustand $n=1$ oder wie Pearce und Hall (1980) es formulieren: „[...] the starting value of α for a supposedly novel stimulus may well be set by experience that the subjects have had with other similar stimuli.“ (p. 538) Oder: „The associability of A will be at its *normal* starting value[...]“ (p. 540; Hervorhebung hinzugefügt).

Ein generelles Problem dieser Modelle, welche Lernen als primär abhängig von Aufmerksamkeitsprozessen (Salienz) auffassen, besteht darin, eine Abnahme von Verbindungsstärken zu erklären, die aus dem Zusammenspiel mehrerer CS resultiert. Werden z.B. zwei CS zuerst einzeln mit einem US gepaart und treten danach die zwei CS gemeinsam mit dem US auf, so führt dies zu einer Abnahme der Verbindungsstärken zwischen den CS und dem US. Das Rescorla-Wagner Modell kann dies leicht dadurch erklären, dass in der ersten Phase jeder CS die maximale Verbindungsstärke mit dem US erhalten hatte. Bei gemeinsamen Auftreten führt dies nun zu einer „Übererwartung“ und die Verbindungsstärke muss abnehmen (vgl. Rescorla & Wagner, 1972; Gluck & Bower, 1988b). Das Mackintosh-Modell hingegen integriert den zweiten Stimulus nicht in seine Bestimmung der erforderlichen Veränderung und reagiert damit nicht auf die „Überdeterminierung“. Das Pearce-Modell kann ohne Erweiterungen und Zusatzannahmen keine Abnahme der Verbindungsstärken erklären, da der Lernschritt wesentlich durch den Salienzparameter bestimmt wird und dieser nur Werte größer oder gleich Null annehmen kann. Eine Übersicht der Modelle gibt Tabelle 1.

Tab. 1: Modellvorstellungen des Rescorla-Wagner Modells sowie verschiedener Weiterentwicklungen. α =Salienzparameter, β =Lernratenparameter, V_i = Verbindungstärke zwischen Reiz i und einem US, λ =asymptotisch zu erreichende Assoziierungs- bzw. Repräsentationsstärke des US, l =asymptotisch zu erreichende Assoziierungsstärke des CS zu seiner Repräsentation im Kurzzeitgedächtnis, r =Repräsentation des CS im Kurzzeitgedächtnis. R =Repräsentation des US im Kurzzeitgedächtnis, S_A = Intensität des CS.

	Lernvorstellung	Salienz	Performanz
Rescorla & Wagner, 1972	$\Delta V_A = \alpha_A \beta (\lambda - \Sigma V_i)$ Die Verbindungstärke wird proportional zur Differenz zwischen der Vorhersage durch alle Merkmale und der Beobachtung verändert.	Keine genaueren Angaben, abhängig vom CS.	Unterscheidung zwischen Lernmodell und einem noch zu spezifizierenden Antwortmodell; monotone Relation zwischen Verbindungsstärke und Reaktion wird angenommen.
Mackintosh, 1975	$\Delta V_A = \alpha_A \beta (\lambda - V_A)$ Die Verbindungsstärke eines Reizes A mit einem US wird proportional zur Differenz der Vorhersage durch den Reiz A und dem US verändert. Die Salienz moduliert die Stärke der Veränderung.	$\Delta \alpha_A$ positiv wenn gilt $ \lambda - V_A < \lambda - V_X $ $\Delta \alpha$ negativ wenn gilt $ \lambda - V_A \geq \lambda - V_X $	Monotoner Zusammenhang zwischen Verbindungsstärke und Reaktion. Eventuell erhöht Salienz die Reaktionsstärke.
Wagner, 1978	$\Delta V_A = \alpha_A (1-r) \beta (\lambda - R)$ Die Stärke der Änderung der Verbindung hängt sowohl von der Vorhersehbarkeit des CS als auch des US ab.	α_A Stimulusabhängig. Die Assoziationsfähigkeit des CS ist proportional zu seiner Unvorhersagbarkeit aus den Kontextreizen: $(1-r)$.	Nichterwartete Ereignisse führen zu einem häufigeren Rehearsal-Prozess und üben dadurch einen stärkeren Einfluss auf das Antwortverhalten aus.
Pearce & Hall, 1980	$\Delta V_A = S_A \alpha_A \lambda$ Die Verbindungsstärke wird durch die Intensität des CS, die Salienz des CS und die Stärke des US determiniert.	$\alpha_A = \lambda^{n-1} - \Sigma V_i^{n-1} $ Die Salienz verhält sich antiproportional zur Vorhersagekraft der Merkmale im vorausgehenden Durchgang .	Durch Aktivierung von Gedächtnisrepräsentationen des US wird Verhalten ausgelöst.

Modellierung des Blockingexperimentes von Kamin (1969)

Wie können nun das Rescorla-Wagner Modell und seine Erweiterungen auf das Blocking-Experiment von Kamin (1969) angewendet werden?

Die Studie von Kamin (1969) aus Sicht des Rescorla-Wagner Modells

Aus Gründen der Einfachheit soll ein aus drei Einheiten bestehendes Rescorla-Wagner Modell angenommen werden (vgl. Abbildung 1.2). A repräsentiert dabei den Ton, B das Licht und λ die Erfahrung des Schocks. Bei Vorliegen von Ton, Licht oder Schock erhalten die betreffenden Einheiten die Aktivierungsstärke 1 ansonsten wird deren Aktivität auf 0 gesetzt. Am Anfang wird der Schock noch nicht erwartet. Deshalb ist die Verbindung zwischen Ton und Schock sowie Licht und Schock gering. Während der ersten Phase des Experimentes, in welcher der Ton gekoppelt mit einem Schock auftritt, nimmt die Verbindungsstärke zwischen der Toneinheit und der Schockeinheit zu und erreicht nach genügend Assoziationsdurchgängen schließlich die Stärke $\lambda=1$. In der zweiten Phase, wenn sowohl der Ton, das Licht als auch der Schock gleichzeitig auftreten, ist die Verbindung zwischen Ton und Schock bereits so stark, dass die richtige Vorhersage des Schocks erfolgt. Aus diesem Grund bewirkt Gleichung 1 nur eine kleine Veränderung der Verbindungsstärke zwischen Licht und Schock. Da in der Kontrollbedingung in der zweiten Phase noch keine Verbindungsstärken etabliert sind, wird sowohl der Ton als auch das Licht mit dem Schock assoziiert. Ein Lernalgorithmus, der unabhängig vom Überraschungswert lernt (vgl. kompetitives Lernen weiter unten), sagt hingegen in beiden Gruppen die gleiche Assoziationsstärke voraus.

Die Studie von Kamin (1969) aus Sicht des Mackintosh Modells (1975)

Da zu Beginn der zweiten Phase bereits der Ton als verlässlicher Prädiktor des US etabliert wurde, nimmt die Salienz des Lichtes, welches immer zusammen mit dem Ton auftritt, ab (Gleichung 7). Daraus folgt, dass die Verbindung zwischen dem Licht und dem Schock schwächer gelernt wird als zwischen dem Ton und dem Schock. Allerdings wird in den *ersten* Durchgängen der zweiten Phase die Verbindung zwischen dem Licht und dem Schock noch leicht verstärkt, da hier die Salienz noch nicht aufgrund der Erfahrung abgenommen hat. In der Kontrollgruppe wurde in der zweiten Phase noch kein verlässlicher Prädiktor etabliert und sowohl der Ton als auch das Licht werden zu gleichen Teilen assoziiert.

Die Studie von Kamin (1969) aus Sicht des Wagner Modells (1978)

Dadurch, dass der Ton in der ersten Phase mit dem US gepaart war, ruft der Ton in der zweiten Phase eine gute Repräsentanz des Elektroschocks im Kurzzeitgedächtnis hervor. Diese Repräsentation des Elektroschocks schwächt sich schnell wieder ab, und Rehearsal tritt, aufgrund der guten Vorhersage, nur vermindert auf (Wagner, 1978, p. 180). Deshalb wird die Verbindung zwischen Licht und Schock nicht mehr so gut wie zwischen Ton und Schock in der ersten Phase gelernt. Diese Vorstellung entspricht im Wesentlichen der des Rescorla-Wagner Modells.

Die Studie von Kamin (1969) aus Sicht des Pearce & Hall Modells (1980)

Durch das Erlernen der Verbindung zwischen dem Ton und dem Schock in der ersten Phase entsteht in der zweiten Phase keine Differenz mehr zwischen der maximalen Stärke und der bereits erreichten Stärke. Aus diesem Grund wird in allen auf den ersten Durchgang folgenden Durchgängen der Salienzparameter gleich 0 sein (Gleichung 9) und somit das Licht nicht mehr mit dem Schock assoziiert. Im Gegensatz zur Modellvorhersage von Mackintosh (1975) besitzt hier auch der Ton eine Salienz von 0.

Wie durch diese Auflistung deutlich geworden sein sollte, spielen die spezifischen Annahmen unterschiedlicher, auf fehlerminimierendem Lernen beruhender Modelle, keine wichtige Rolle für die Vorhersage von Blocking-Effekten. Der Blocking-Effekt erweist sich damit also als typisches Merkmal fehlerminimierenden Lernens, unabhängig von der genauen Implementation. Im ursprünglichen Rescorla-Wagner Modell entsteht Blocking allein durch die verminderte Diskrepanz zwischen Vorhersage und Ereignis. Dagegen beruht Blocking nach den Modellvorstellungen von Mackintosh (1975) und Pearce und Hall (1980) auf einer verminderten Assoziierfähigkeit des blockierten Reizes aufgrund der verminderten Salienz. Diese Modelle wurden hier so ausführlich vorgestellt, um die Einordnung des weiter unten vorgeschlagenen Modells für Klassifizierungsaufgaben zu erleichtern.

Bewertung des Rescorla-Wagner Modells

Das Rescorla-Wagner Modell erklärt eine Fülle verschiedener Phänomene aus dem Bereich der Pavlovschen Konditionierung in einem einheitlichen und explizit formulierten Ansatz. Es erlaubt durch zusätzliche Annahmen kontingentes Verstärkungslernen durch ein formales Modell zu beschreiben. Kontingentes Lernen bedeutet, dass nicht nur die Anzahl an Paarungen zwischen US und CS, sondern auch die Häufigkeit mit welcher der US oder der CS alleine auftreten, die Gewichtsstärken beeinflusst. Der problematische Fall, der zusätzliche Annahmen für das Rescorla-Wagner Modell nötig macht, stellt dabei das alleinige Auftreten des US ohne CS dar. In diesem Fall sollte die Gewichtsstärke zwischen dem CS und dem US abnehmen, ohne dass der CS überhaupt auftritt. Das Rescorla-Wagner Modell erwartet aber nur für Gewichtsverbindungen von *aktiven* Reizen eine Veränderung, da andernfalls alle Reize die gleiche Verbindungsstärke erhalten würden. Um doch kontingentes Lernen zu ermöglichen, muss das Modell durch Kontextreize erweitert werden: Kontextreize erfahren, da sie immer vorliegen, auch dann eine Verstärkung durch den US, wenn der US ohne vorausgehenden CS auftritt. Die daraus resultierende positive Verbindung zum US blockt schließlich den Erwerb der CS-US Assoziation⁷.

Aufgrund dieser genauen Spezifikation der Vorhersagen kommt es auch zwangsläufig zu verschiedenen invalidierenden Ergebnissen. Allerdings widersprechen sich diese invalidierenden Ergebnisse manchmal auch selbst. So wird von Mackintosh (1975) das Mackintosh-Modell der Rescorla-Wagner Regel unter anderem deshalb vorgezogen, weil empirische Ergebnisse nahe legen, dass Blocking bei der ersten Präsentation der zusammengesetzten Stimuli in Phase zwei noch nicht auftritt⁸ (Mackintosh, 1975, pp. 283, vgl. auch Mackintosh, 1978). Mackintosh beruft sich dazu auf die Studie von Kamin (1969). Miller, Barnet und Grahame (1995) wiederum sehen nach einer Literaturdurchsicht diese Vorhersage des Mackintosh Modells als empirisch nicht bestätigt und damit als Nachteil an (p. 367). Miller, Barnet und Grahame (1995) kommen weiter zu dem Schluss, dass das Rescorla-Wagner Modell 18 Phänomene gut beschreibt, darunter Phänomene wie

⁷ Genaugenommen funktioniert dieser Mechanismus nur, wenn von einer sehr viel größeren Lernrate als Extinktionsrate ausgegangen wird. Da der Kontextreiz per Definition immer vorliegt, liegt er auch häufig vor, ohne mit einem US gepaart zu sein. Dies müsste zur Extinktion des Kontextreizes und somit geringerer Konkurrenz für den CS führen - es sei denn, spezielle Annahmen über den Zeitpunkt, zu dem Lernen erfolgen kann, werden getroffen. Rescorla und Wagner (1972) fordern eine größere Lern- als Extinktionsrate. Die Größenordnung der dafür nötigen Extinktionsrate dürfte aber inkonsistent zu der Extinktionsgeschwindigkeit gewöhnlicher Extinktionsversuche sein.

⁸ Das Modell von Pearce & Hall (1995) scheint gerade dazu entworfen worden zu sein, diesen Effekt zu zeigen. Ansonsten würde keine Notwendigkeit bestehen, die Salienz vom vorherigen Durchgang abhängig zu machen, sondern es könnte die Salienz zuerst berechnet werden und dann die Gewichtsveränderung durchgeführt werden.

Lernkurven, Löschungskurven, Stimulusgeneralisierung, Diskriminierung, Blocking und sogar instrumentelles Konditionieren, während grundlegende Modellvorstellungen des Rescorla-Wagner Modells nur schwer mit 23 anderen Beobachtungen, darunter Phänomene wie spontane Erholung, Veränderung bei erneutem Lernen, gleichzeitige exzitatorische und inhibitorische Verbindungen zwischen einem CS und einem US, vereinbar sind. In ihrem Resümee schließen Miller, Barnet und Grahame (1995), dass das Rescorla-Wagner Modell zwar nicht alle Beobachtung erklären kann, dass ihm aber keines der alternativen Modelle darin überlegen ist. Als wesentliche Probleme führen die Autoren an, dass

- (a) das Rescorla-Wagner Modell von einer Limitierung der assoziativen Stärke durch den US, wie durch Parameter λ bestimmt, ausgeht,
- (b) Exzitation und Inhibition als sich gegenseitig ausschließende Mechanismen angesehen werden,
- (c) eine fixe Salienz angenommen wird,
- (d) über die Erwerbgeschichte hinweg abstrahiert wird – für die Vorhersage nur die Verbindungsstärken benutzt werden nicht aber die Akquisition derselben und
- (e) die Performanz alleine aus assoziativen Verbindungen heraus begründet wird.

Bei all diesen verschiedenen Resultaten erscheint es mir angemessen, das Rescorla-Wagner Modell als ein heuristisches Modell zu nutzen, welches durch seine hohe Formalisierung und seine einfache Implementierung leicht die Testung fehlerminimierender gegen nichtfehlerminimierender Modelle ermöglicht. Das wesentliche zu testende Element in dieser Diplomarbeit, die Fehlerkorrektur des Rescorla-Wagner bzw. des Deltamodells, wird von Miller, Barnet und Grahame (1995) als keine problematische Annahme angesehen. Zweck der Diplomarbeit war nicht, zwischen speziellen Ausformungen dieses Fehlerkorrektur-Modells zu unterscheiden. Vielmehr sollte eine allgemeine Algorithmenklasse als plausibel für Kategorisierung ohne Rückmeldung etabliert werden. Da alle eben diskutierten Modelle in diese Algorithmenklasse fallen, wird der folgenden theoretischen Besprechung nur das Rescorla-Wagner Modell aufgrund seiner klaren mathematischen Eigenschaften zugrunde gelegt. Dies soll aber nicht als Verwerfung der anderen Modelle verstanden werden. Später werden, wenn es für die Erklärung der empirischen Daten nötig wird, konkretere Annahmen bezüglich der Salienz getroffen, als dies in dem ursprünglichen Rescorla-Wagner Modell der Fall ist. Damit wird gleichzeitig einem Kritikpunkt von Miller, Barnet und Grahame (1995) begegnet, welche im Fehlen einer Formalisierung des Salienzbegriffes eine Schwäche des Rescorla-Wagner Modells sehen.

Während das Rescorla-Wagner Modell häufig als Lernmodell für Tierexperimente Verwendung fand, wurde für Modelle, welche menschliche Lernphänomene beschreiben sollten, das äquivalente Delta-Modell verwendet. Wie oben angedeutet wurde, kann das Rescorla-Wagner Modell in einem zweischichtigen Netz realisiert werden, welches als Lernalgorithmus die Delta-Regel benutzt. Da diese Implementationsstrategie häufig in Erklärungsmodellen für Klassifizierungsphänomene Anwendung fand (z.B. Gluck & Bower, 1988a, Gluck, 1991, McClelland & Rumelhart, 1985, Bösche, 2002, Zwickel & Wills, 2002), und die konkrete Implementierung die theoretische Stringenz der Folgerungen aus den Annahmen deutlicher macht, soll im Folgenden auf Delta-Netze als Implementationsform⁹ der Delta-Regel eingegangen werden.

1.2.2 Delta-Netze als mögliche Implementationsform der Rescorla-Wagner Regel

Die Delta-Regel wurde bereits häufig in konnektionistischen Netzen implementiert (z.B. Gluck & Bower, 1988a, Sutton & Barto, 1981). Konnektionistische Netze haben die gemeinsame Eigenschaft, dass sie aus einzelnen simplen Elementen bestehen, die miteinander über variable Gewichte, auch Kanten genannt, verbunden sind. Das „Wissen“ des Netzes ist dabei in den Gewichtsstärken gespeichert, so dass die Aktivierung eines Elementes einen Einfluss auf den Zustand der daran angeschlossenen Elemente ausüben kann. Ob dieser Einfluss exzitatorisch oder inhibitorisch verläuft, hängt vom Vorzeichen des Gewichtes ab. Vorteile der Netze bestehen in ihrer massiven Parallelverarbeitung der einzelnen Stimuli, was die Wechselwirkung zwischen den einzelnen Stimuli erleichtert. Durch die gleichzeitige Verarbeitung mehrerer Stimuli kann das Verarbeitungsergebnis eines Stimulus einfach auf die Verarbeitung eines anderen Stimulus Einfluss nehmen. Dies wäre zwar auch in sequentiellen Modellen möglich, allerdings müssten dafür komplizierte Speicher oder Zustandsmodelle eingeführt werden. Diese Wechselwirkung stellt hingegen eine „natürliche“ Eigenschaft konnektionistischer Netze dar. Als weiterer Vorteil der Netze wird häufig „graceful degradation“ genannt, die Fähigkeit auch bei Wegfall einzelner Elemente noch eine annähernd gleiche Leistung zu zeigen (z.B. Plaut, McClelland, Seidenberg & Patterson, 1996). Auch die Fähigkeit, Ähnlichkeiten aus Einzelexemplaren zu abstrahieren, wird in

⁹ Damit soll nicht gesagt werden, dass konnektionistische Netze per se nur auf der Implementationsebene von Marr (1982) anzusiedeln sind (vgl. Fodor & Pylyshyn 1988). In diesem konkreten Fall liegt aber vermutlich tatsächlich kein Mehrwert in der Implementationsform des konnektionistischen Netzes in Sinne von weiterer daraus folgender Modellvorausagen für die Algorithmenebene vor. Andererseits beschränkt die Implementierung, welche Modellerweiterungen später als plausibel angesehen werden. Höhere Kontrollprozesse wären vermutlich nicht mit der konnektionistischen Vorstellung kompatibel.

vielen Ansätzen konnektionistischer Netze genutzt (z.B. Rumelhart & McClelland, 1986, Elman, 1990). Konnektionistische Netze unterscheiden sich voneinander in der Art, welche Verknüpfungen zwischen den einzelnen Elementen erlaubt sind, bezüglich der Schwellenwerte und eingesetzter Entscheidungsfunktionen sowie eventuell bezüglich der Lernalgorithmen (Rumelhart, Hinton & McClelland, 1986). Ein weiterer Vorteil konnektionistischer Netze besteht darin, dass es mit ihnen, im Gegensatz zu regelbasierten Systemen (z.B. Anderson & Betz, 2001), leicht möglich ist, durch *teilweise* Aktivierung der Elemente, Unsicherheitsfaktoren zu modellieren. Dadurch können gleichzeitig mehrere Strategien verfolgt werden, bis sich im Laufe der Weiterentwicklung eine erfolgreiche Wahl herauskristallisiert. Damit muss nicht schon am Anfang einer Verarbeitung festgelegt werden, welche Regel ausgeführt werden soll, bzw. was das Ziel auf einer niedrigeren Ebene ist. Ebenso muss nicht explizit festgelegt werden, was geschehen soll, wenn die Umweltbedingungen keine der Antezedenzen der Regeln genau erfüllen. So würde z.B. ein konnektionistisches Netz, welches die Kategorie „Großmütter“ erlernt hat, über die Ähnlichkeit der repräsentierenden Merkmalen zu denen eines Großvaters, den Großvater automatisch ähnlich wie die Großmutter behandeln. Ein regelgeleitetes System hingegen müsste zuerst noch definieren, woran das System die Ähnlichkeit erkennen kann und wie es darauf reagieren sollte. Natürlich kann auch ein regelgeleitetes System wie z.B. ACT-R (Anderson & Lebiere, 1998) auf Ähnlichkeiten reagieren und auch Produktionsregeln ausführen, die nur *ähnliche* Antezedenzbedingungen besitzen. Gewöhnlich wird dann die Ähnlichkeit über die Anzahl gemeinsamer Elemente bestimmt und ein gewisser Fehler bei der Auswahl der Regeln angenommen (z.B. Anderson & Betz, 2001). Diese zufällige Aktivierung ähnlicher Produktionsregeln scheint aber ein willkürlicherer Zusatz als die gleichzeitige Mitaktivierung ähnlicher Repräsentanten durch den üblichen Aktivierungsprozess in konnektionistischen Netzen zu sein. Vermutlich wurden diese „konnektionistischen“ Zusatzannahmen nötig, um empirische Daten erklären zu können. Es lassen sich jedoch unterschiedliche empirische Vorhersagen ableiten: Während in regelgeleiteten Systemen Lernen nur für die aktivierte Regel erfolgen sollte, lernen, durch die Mitaktivierung ähnlicher Exemplare in konnektionistischen Netzen, auch die nicht „selektierten“ Exemplare direkt¹⁰ mit. Diese Exemplare könnten zumindest theoretisch auch einzelne Regeln darstellen. Damit

¹⁰ Ein Verfechter von regelbasierten Systemen könnte einwenden, dass auch dort andere Exemplare durch die zufällige Fehlaktivierung mitlernen. Dies sollte aber empirisch nur zu sehr schwachen Effekten und zu keiner Anpassung der „mitaktivierten“ Exemplare an das Ergebnis des präsentierten Exemplare führen, da in regelbasierten Systemen die einzelnen Exemplare typischerweise unabhängig von den anderen verarbeitet werden.

würden sich auch die konnektionistischen Systeme an die regelgeleiteten Systeme, trotz unterschiedlichen empirischen Vorhersagen im Detail, annähern.

Da Kategoriensysteme per Definition ähnliche Elemente zusammenfassen, stellt dies eine gute Anwendungsumgebung für konnektionistische Netze dar. Konnektionistische Netze wären weniger in Bereichen erfolgreich, in denen eine Ähnlichkeit der Eingabe nicht zu ähnlichen Ausgaben führt. Allerdings kann selbst diese Aufgabe, wie weiter unten ausgeführt, prinzipiell von Delta-Netzwerken erledigt werden. Konnektionistische Netze, welche der Delta-Lernregel folgen, werden häufig auch Delta-Modelle genannt. Das Delta-Modell (z.B. McClelland & Rumelhart, 1985, Gluck & Bower, 1988a) erlaubt es, als typischer Vertreter konnektionistischer Netze, komplizierte Entscheidungen auf so viele „lokale“, einfache Entscheidungen zu reduzieren, dass es möglich ist, das gesamte Modell aus denselben Einheiten, aufgebaut anzusehen. Durch diese Vereinfachung werden Erklärungsansätze möglich, die eine Homunkulus-Erklärung, und damit eine Verschiebung des Problems, ersetzen können. Gleichzeitig weisen diese Erklärungsansätze Ähnlichkeiten zu den heutigen Vorstellungen über die Verarbeitung in biologischen Neuronen auf. Die einzelnen Einheiten, werden deshalb auch häufig „Neurone“ genannt. Informationen werden in konnektionistischen Netzen stets nur lokal benötigt. Damit ist eine zentrale Einheit, welche die Verarbeitung regelt, überflüssig. Jedes Neuron entscheidet lokal aufgrund der bei ihm über Kanten einlaufenden Signale, ob es Aktivierung an die angeschlossenen Neurone sendet. Das einfachste Delta-Modell besteht aus zwei Schichten von Knoten (Neuronen) und deren Verknüpfungen. Die Eingabeschicht besteht aus Neuronen, welche nur von Außen Informationen erhalten. Wenn dabei ihr Schwellwert überschritten wird, leiten sie eine Aktivierung an die Zielschicht weiter. Neurone der Zielschicht summieren die einlaufende Aktivität auf und ändern bei Überschreitung ihres Schwellwertes den Aktivitätszustand von 0 auf 1. Jedes Neuron der Eingabeschicht ist dabei mittels Kanten mit jedem Neuron der Ausgabeschicht verbunden. Wissen wird damit im Deltanetz, wie auch in anderen konnektionistischen Netzen, durch die Kantengewichte repräsentiert. Ein hohes Kantengewicht zwischen zwei Neuronen bedeutet, dass ein Zusammenhang zwischen beiden Einheiten gelernt wurde. Daher führt die Aktivität des einen Neurons zur Aktivierung des anderen. Um diese Gewichte zu finden wird ein Lernalgorithmus eingesetzt, welcher der Rescorla-Wagner Regel entspricht. Formal lässt sich das Delta-Modell vollständig durch die Angabe der Anzahl der Neuronen in der Eingabe- und Ausgabeschicht, den Eingabe- und Ausgabemustern, Schwellenwerten und zugehörigen Entscheidungsfunktionen sowie der

genauen Spezifikation der Lernregel beschreiben. Die Aktivität der Ausgabeschicht berechnet sich durch die Summe der mit den Kanten gewichteten Aktivität der Eingabeneuronen:

$$o_j = \sum_i w_{ij} a_i \tag{Gleichung 11}$$

dabei notiert o_j für die eintreffende Aktivierung bei Ausgabeknoten j , w_{ij} für die Gewichte von Eingabeneuron i zu Ausgabeneuron j und a_i für die Aktivität des Eingabeneurons i , gewöhnlich im Bereich 0-1.

Die Standard Delta-Lernregel lautet:

$$\Delta w_{ij} = \beta (\lambda - o_j) a_i \tag{Gleichung 12}$$

wobei β ein Lernratenparameter und λ das Klassifizierungsziel darstellen.

In Abbildung 1.5 ist ein aus 4 Eingabeknoten und 2 Ausgabeknoten bestehendes Deltamodell visualisiert. In den Ausgabeknoten ist jeweils der Schwellwert, der bei der Aufsummierung überschritten werden muss um das Neuron zu aktivieren, eingeschrieben. Die Kanten sind mit den Gewichtsstärken beschriftet. Die Delta-Lernregel ist rechts oben aufgeführt. Nur bei Anliegen des Reizes (0000) würde der linke Ausgabeknoten aktiviert, da nur in diesem Fall keine negative Aktivierung (Hemmung) über die negativen Kanten aufsummiert würde. Bei allen anderen Mustern würde nur der rechte Ausgabeknoten aktiviert.

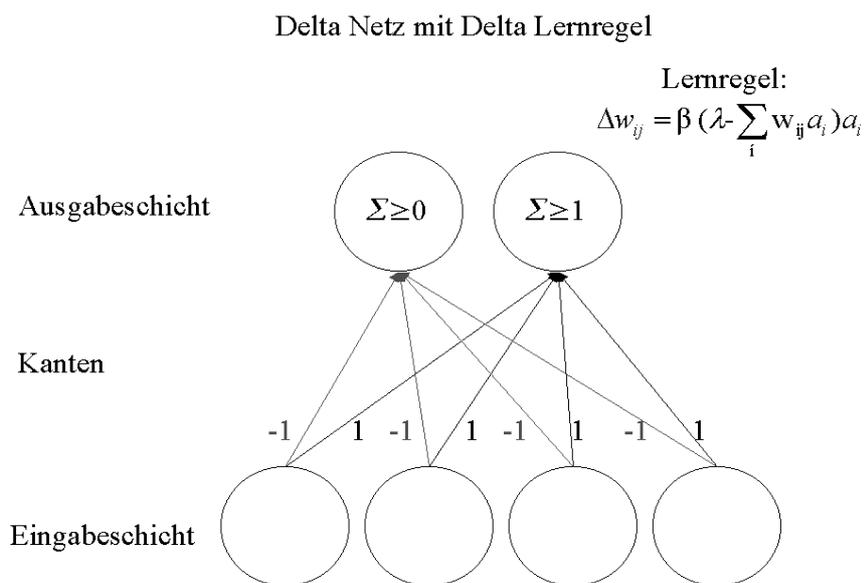


Abb. 1.5: Beispiel für ein Delta-Netzwerk.

Nach Gleichung 12 verändern die Gewichte ihre Stärke in Abhängigkeit von der Differenz der am Ausgabeknoten resultierenden Aktivierung und der erwarteten Aktivierung. Wie bereits oben im Rahmen des Rescorla-Wagner Modells erwähnt, entsprechen aufgrund dieses Fehleralgorithmus nach einer unendlichen Anzahl von Beispielaktivierungsmustern die Gewichte zwischen den einzelnen Eingabeinformationen und den Ausgabeknoten den Gewichten, die sich aus einer linearen Regression mit diesen Beispielen ergeben würden. Darin liegt auch ein gewisses Problem dieser Modellvorstellung. Unterscheidungen von nicht-linear separierbaren Kategorien sind damit nicht möglich. Bei nicht-linear separierbaren Kategorien handelt es sich, wie oben erwähnt, um eine Aufteilungsstruktur, welche die Vorhersage der einzelnen Kategorien nicht durch eine lineare Kombination der einzelnen Merkmale erlaubt. Nicht-linear separierbare Kategorienstrukturen können aber mittels eines Delta-Netzwerkes durch eine geänderte Repräsentationsform an der Eingabeschicht, z.B. im Sinne der oben diskutierten orthogonalen Kodierung, bzw. durch „hidden units“ gelöst werden. Die „hidden units“ stellen eine Zwischenschicht zwischen den Eingabe- und Ausgabeneuronen dar. Ihre Funktion besteht in diesem Fall darin, die oben diskutierte Lösung durch orthogonale Kodierung der Eingabeschicht in einer Zwischenschicht nachzubilden (Gluck & Bower, 1988b). Nach dieser eher informellen Darstellung des Delta Modells sollen im Folgenden kurz seine Eigenschaften formaler diskutiert werden. Schließlich wird noch versucht die Lösung von nicht-linear trennbaren Problemen durch Delta-Netzwerke zu visualisieren.

Das Ziel eines konnektionistischen Netzes bei Klassifikationsaufgaben besteht darin, einen Gewichtsvektor zu finden, der multipliziert mit dem Eingabevektor erlaubt die Eingabemuster in die gewünschten Kategorien aufzuteilen. Dazu muss die Multiplikation des Gewichtsvektors mit dem Eingabevektor für die eine Klasse von Mustern einen Wert unterhalb einer Schwelle S ergeben und für die andere Klasse von Mustern oberhalb der Schwelle:

$$\mathbf{a} \cdot \mathbf{w} > S \text{ für alle } \mathbf{a} \in \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n\} \text{ und}$$

$$\mathbf{a} \cdot \mathbf{w} < S \text{ für alle } \mathbf{a} \in \{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_m\}$$

wobei \mathbf{a} der Eingabevektor, \mathbf{w} der Gewichtsvektor, \mathbf{X}_i Eingabemuster der Klasse 1 und \mathbf{Y}_i Eingabemuster der Klasse 2 darstellen. „ \cdot “ symbolisiert das innere Produkt, bzw. bei Normierung der Vektoren auf die Länge 1, die Korrelation. Aus Symmetriegründen wurde auf die Behandlung von $\mathbf{a} \cdot \mathbf{w} = S$ verzichtet. Für diesen Fall kann das Gleichheitszeichen dadurch

eliminiert werden, dass S minimal verändert wird. Dies ist falls S eine *reelle* Zahl repräsentiert, immer möglich.

Damit mittels einer fixen Schwelle die Muster getrennt werden können, muss eine lineare Trennfunktion existieren. Der Gewichtsvektor muss orthogonal zu dieser Funktion stehen. Dies kann man sich grafisch recht leicht verdeutlichen. In Abbildung 1.6 repräsentieren dunkle Punkte Eingabevektoren der Klasse 1 und helle Punkte Eingabevektoren der Klasse 2. Weiterhin ist der Gewichtsvektor sowie eine Trennfunktion eingezeichnet. Dem inneren Produkt $\mathbf{a} \cdot \mathbf{w}$ entspricht in der grafischen Darstellung ein Projektion des Eingabevektors auf den Gewichtsvektor. Damit gilt, dass $\mathbf{a} \cdot \mathbf{w}$ genau dann auf den Schwellenwert S abgebildet wird, falls \mathbf{a} auf der Trennfunktion liegt. Weiter gilt: $\mathbf{a} \cdot \mathbf{w} < S$ für Punkte unterhalb der Trennfunktion und $\mathbf{a} \cdot \mathbf{w} > S$ für Punkte oberhalb der Trennfunktion.

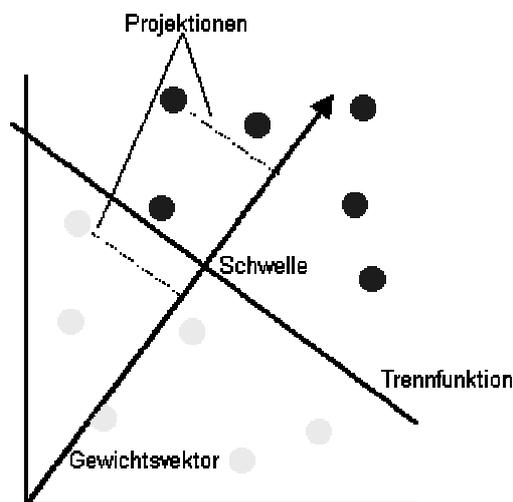


Abb. 1.6: Dunkle und helle Muster werden durch eine Trennfunktion separiert. Der Gewichtsvektor steht senkrecht auf der Trennfunktion. Deshalb bleiben alle Projektionen unter der Trennfunktion unterhalb der Schwelle und Projektionen rechts oberhalb davon über der Schwelle.

Damit eine Trennfunktion bzw. eine geeignete Abbildung von Eingabevektoren auf die Ausgabevektoren *garantiert* bestimmbar ist, müssen Eingabevektoren aus unterschiedlichen Klassen linear unabhängig sein. Bei linear unabhängigen Klassen können dann die einzelnen Merkmale als Hinweise für eine Klasse monoton aufsummiert, bzw. kann die Bedeutung eines Merkmals unabhängig von seinem Kontext bewertet werden.

Da das XOR-Problem und ähnliche Probleme als Argumente gegen die Brauchbarkeit von konnektionistischen Netzen von Minsky und Papert (1969) vorgebracht wurden, soll dieses Problem genauer beleuchtet werden. Damit soll der Eindruck vermieden werden, dass das Delta Modell, welches in dieser Diplomarbeit als Erklärungsmodell benutzt wird, nur für

einen artifizell einfachen Aufgabenbereich eingesetzt werden kann. Eine einfache Möglichkeit, das oben diskutierte XOR-Problem zu lösen, liegt in einer Umkodierung der Reize, wie bereits mehrfach angedeutet. Durch die Repräsentation jedes Musters anhand einer eigenen Eingabeeinheit werden die Eingaben orthogonalisiert. In Abbildung 1.7 ist dies durch eine Verdopplung der Eingabeeinheiten geschehen.

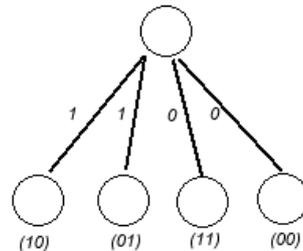


Abb. 1.7: Orthogonale Eingaberepräsentation um das XOR-Problem zu lösen.

Diese Repräsentationsform erlaubt mit der einfachen Trennfunktion $\sum_i w_i x_i$ und der Schwelle 0 die Stimuli in die gewünschten zwei Klassen einzuteilen. Eine weitere Möglichkeit besteht darin mittels „hidden units“ eine höherdimensionale Beschreibung zu finden. Im Wesentlichen besteht die Lösung darin, ein anderes Ähnlichkeitsmaß zwischen den Reizen zu schaffen, welches komplexere Zusammenhänge bzw. Strukturen erfasst. In Abbildung 1.8a ist eine Lösung für zwei „hidden units“ und in Abbildung 1.8b für eine „hidden unit“ gezeigt. Um „Halluzinationen“ bei Fehlen jeglicher Information zu vermeiden wird gewöhnlich eine Schwelle größer 0 für den Ausgabeknoten gewählt.

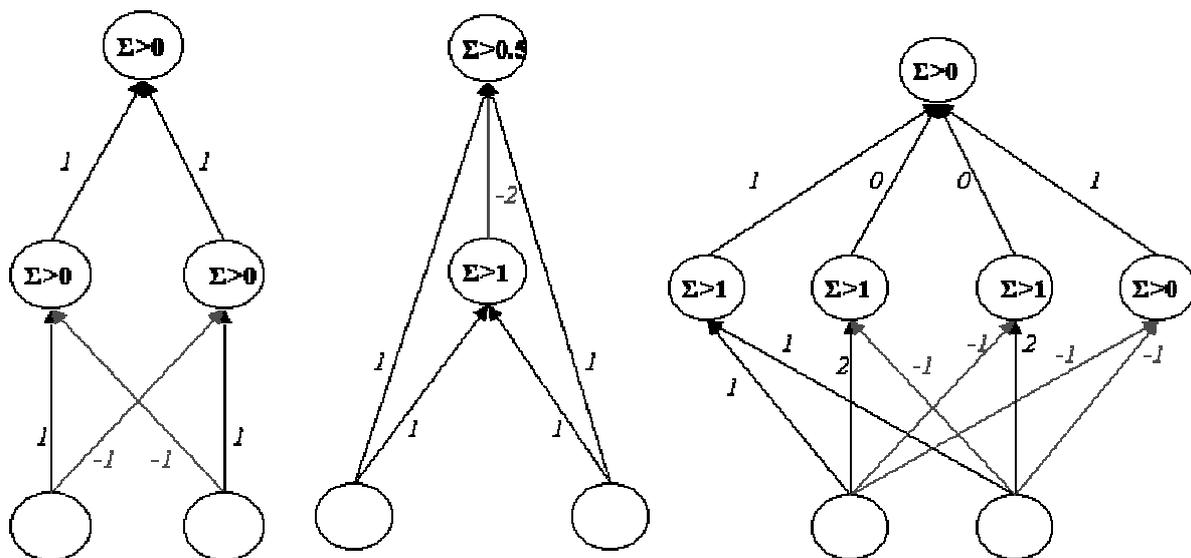


Abb. 1.8a-c: Lösung des XOR-Problems durch unterschiedliche Anzahlen an verborgenen Einheiten – „hidden units“.

Abbildungen 1.8a-c zeigen jeweils die Eingabeknoten A und B, die „hidden units“ und einen Ausgabeknoten. Die Eingabeknoten können die Aktivierung 1 (Merkmal vorhanden) und 0 (Merkmal nicht vorhanden) annehmen. Entlang der Pfeile sind die Gewichte eingetragen. Weiterhin ist der Schwellenwert der „hidden units“ aufgeführt. In Abbildung 1.8a wird die Lösung dadurch erreicht, dass jeder Eingabeknoten gleichzeitig eine „hidden unit“ hemmt und eine andere verstärkt. Für den Fall, dass nur eine Eingabeeinheit aktiv ist, fällt somit die Hemmung durch die andere Eingabeeinheit weg und die Ausgabeneinheit kann aktiviert werden. In Abbildung 1.8b wird der Ausgabeknoten nur für den Fall, dass beide Eingabeknoten aktiv sind, gehemmt. Abbildung 1.8c zeigt schließlich einen Fall, in dem die einzelnen Muster orthogonal in der Zwischenschicht repräsentiert werden.

Mathematisch lässt sich die Funktion der „hidden units“ als Transformation der Eingabevektoren beschreiben. Durch die Multiplikation des Gewichtvektors mit dem Eingabevektor entsteht abhängig von den Schwellenwerten der Zwischenschicht eine neue, „höher“ strukturierte, Repräsentation der Eingabe. Im Fall 1.8c wird aus einem zweidimensionalen Eingaberaum ein vierdimensionaler Repräsentationsraum. Diese neue Repräsentation der Muster erlaubt die Punkte durch eine Hyperebene in die gewünschten Klassen einzuteilen. Da dies im höherdimensionalen Fall schwer grafisch darstellbar ist, wird im Folgenden eine Lösung des XOR-Problems anhand von Beispiel 1.8a visualisiert werden, die nur zwei „hidden units“ benötigt. In Abbildung 1.9a ist die Verschaltung des Netzwerkes erneut gezeigt. Nur in dem Fall, dass die beiden Eingaben ungleich sind, wird eine „hidden unit“ aktiv. Bei Eingabe (11) werden beide „hidden units“ gleich stark aktiviert wie auch gehemmt. Im Falle (00) wird die Aktivierungsschwelle nicht überschritten.

Die Multiplikation der Eingabevektoren mit den Gewichtsvektoren zu den jeweiligen „hidden

units“ bildet den Vektor $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ nach Transformation durch die „hidden units“ ebenfalls auf $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$

ab. Für Vektor $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ gilt: $\left(\begin{array}{l} \begin{pmatrix} 1 \\ 1 \end{pmatrix} * \begin{pmatrix} 1 \\ -1 \end{pmatrix} \rightarrow 0 \\ \begin{pmatrix} 1 \\ 1 \end{pmatrix} * \begin{pmatrix} -1 \\ 1 \end{pmatrix} \rightarrow 0 \end{array} \right)$ Die Eingabevektoren $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ und $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ werden auf

sich selbst abgebildet (vgl. Abbildung 1.9b). Wie die in Abbildung 1.9b zeigt, wird durch diese Transformation das Problem linear separierbar. Eine weitere Möglichkeit, das XOR-Problem zu lösen besteht darin, kompetitive Netze vorzuschalten (vgl. Rolls & Treves, 1998, chapter 4).

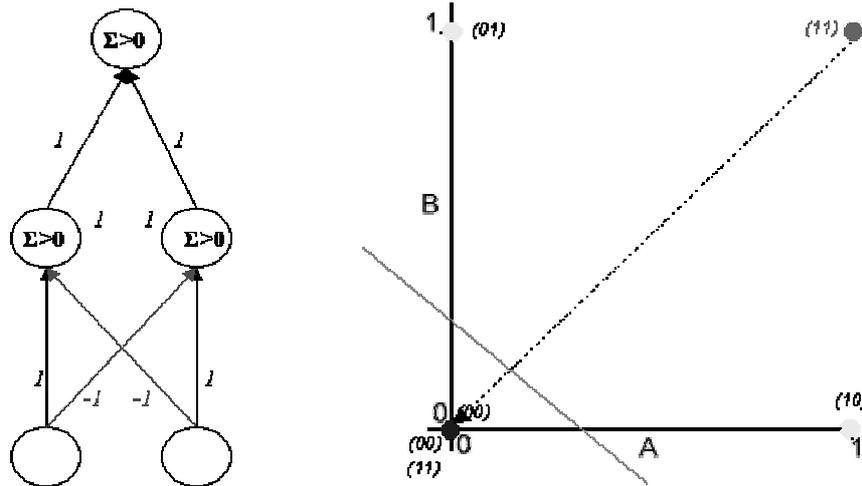


Abb. 1.9ab: Durch die Transformation der Eingabevektoren gemäß der Darstellung in (a) entsteht eine Repräsentation im Raum welche erlaubt die gewünschte Konfiguration durch eine Gerade zu erzeugen (b). Das Eingabemuster (11) wird dazu durch Multiplikation mit dem Gewichtsvektor auf dasselbe Aktivierungsmuster der „hidden units“ wie das Exemplar (00) abgebildet. Dies ist durch den Pfeil angedeutet.

Die gewählte Lösung eines Delta-Lernalgorithmus hängt von der Anzahl der „hidden units“, der Eingabekodierung und anderen Anfangszuständen des Netzes ab. Entspricht die Anzahl an „hidden units“ z.B. der Anzahl an Mustern, so können die Muster orthogonal repräsentiert werden, und die Lernvorgänge für die einzelnen Muster interferieren nicht (vgl. Abbildung 1.8c). Um Repräsentationen in mehreren Schichten zu erlernen, kann eine modifizierte Delta-Regel (Back-Propagation) benutzt werden, welche jedem Neuron einen gewissen Fehleranteil zuschreibt und dann die Gewichte dem Fehler entsprechend verändert. Der Fehler einzelner Neurone wird dann an aktivierende Neurone in früheren Schichten weitergegeben (propagiert). Die Neurone dieser Schicht ändern sich daraufhin entsprechend einer Delta-Regel, welche auf dem propagierten Fehler operiert (Rumelhart, Hinton & Williams, 1986). Im zweischichtigen Modell beweist das Perzeptronen Konvergenz Theorem, dass eine Lösung durch den Delta-Algorithmus gefunden wird, falls das Problem mit dem Netzwerk überhaupt lösbar ist (Rosenblatt, 1962). Dagegen kann die Lösung in mehrschichtigen Netzen mit dem gerade angedeuteten Backpropagation Verfahren nicht mehr garantiert werden, selbst wenn eine Lösung prinzipiell existiert. In mehrschichtigen Netzen kann sich der Lösungsalgorithmus in einem lokalen Minimum des berechneten Fehlers verfangen. Es gibt zwar heuristische Verbesserungen des Algorithmus, aber keine davon kann die Lösung garantieren. Für Netzwerke mit mindestens einer Zwischenschicht kann bewiesen werden, dass, genügend Einheiten in der Zwischenschicht vorausgesetzt, jede stetige Funktion berechnet werden kann (für eine Rückführung auf das Kolmogorov Theorem, z.B. Hecht-

Nielsen, 1987). Dies ist allerdings ein Existenzbeweis und ermöglicht weder die Voraussage, wie die benötigten Gewichte gefunden werden können, noch, ob mit einer größeren Anzahl an Zwischenschichten nicht eine effektivere Repräsentation des Problems erreicht werden könnte.

Wie oben anhand der orthogonalen Lösung des XOR-Problems gezeigt wurde, hat die Form der Kodierung einen großen Einfluss auf die Möglichkeiten eines konnektionistischen Netzes. Ein generelles Problem der geeigneten Kodierung besteht darin, dass ein Merkmal, welches durch „Nichtaktivität“ kodiert wird nicht berücksichtigt wird, da die Lernregel die Gewichtsänderung von einer Aktivität abhängig macht. Ein Beispiel wäre die Kodierung des Merkmals „kein Fell“ durch „0“. Diese Form der Kodierung macht nur dann Sinn, wenn ein Merkmal, welches aus der Ersetzung von „0“ durch „1“ oder umgekehrt entsteht auch tatsächlich das exhaustive logische Äquivalent als Reaktion erzeugen soll. Ansonsten sollten zwei verschiedene Eingabeknoten Verwendung finden. Für den Fall also, dass auf das Merkmal „kein Fell“ *nicht* mit der gegensätzlichen Handlung wie auf das Merkmal „Fell“ reagiert werden soll, sollte die Kodierung durch zwei Eingabeknoten erfolgen. Ein Knoten sollte dann das Merkmal „Fell“ durch den Aktivitätszustand „1“ und ein Knoten das Merkmal „kein Fell“ durch Aktivitätszustand „1“ kodieren. Das Problem wird an einem Delta-Netzwerk mit einem Eingabe- und einem Ausgabeknoten besonders plastisch. Soll z.B. der Ausgabeknoten sowohl bei Vorliegen des Merkmals „Fell“ als auch „kein Fell“ aktiv sein, so ist dies bei einer Kodierung von „kein Fell“ durch „0“ und „Fell“ durch „1“ nicht lösbar.

Delta-Netzwerke sind in der Lage Probability-Matching bei Versuchspersonen zu erklären, wenn die Aktivierungsstärke der Ausgabeknoten linear auf einen Wahrscheinlichkeitsraum für die Entscheidungen abgebildet werden (Gluck, 1992). Probability-Matching tritt auf, wenn in Experimenten die Häufigkeiten einzelner Entscheidungen den Wahrscheinlichkeiten mit denen diese Entscheidungen verstärkt werden entsprechen. Vom normativen Standpunkt aus wäre es hingegen effektiver nur mit der Entscheidung zu reagieren, welche mit der größten Wahrscheinlichkeit verstärkt wird¹¹.

Da ein Delta-Netzwerk bei C Eingabeneuronen höchstens C linear unabhängige Eingabevektoren haben kann, können maximal C verschiedene Abbildungen auf ein kontinuierliches Ausgabeneuron gelernt werden. Werden die Eingabevektoren kontinuierlich kodiert, so ist es sehr wahrscheinlich, dass bei Kodierung der Muster durch zufällige

¹¹ Jedenfalls wenn davon ausgegangen werden kann, dass die Ausführung der am häufigsten verstärkten Handlung keine größeren Kosten verursacht. Diese Annahme könnte aber in realistischen Situationen verletzt sein, in denen die mehrmalige Ausführung derselben Handlung starke negative Effekte haben kann, welche durch die seltenen Verstärkungen nicht ausgeglichen werden.

Ausprägungen der Eingabevektoren, die maximale Anzahl an unabhängigen Eingabevektoren erreicht wird. Die Kapazität kann noch auf $2C$ pro Ausgabeknoten erhöht werden¹², falls nur eine binäre Entscheidung benötigt wird (Rolls & Treves, 1998, chapter 5).

Prototypenabstraktion

Die Delta-Regel erlaubt Gemeinsamkeiten von „Exemplar“-Repräsentationen zu extrahieren. Verlässlich auftretende Eigenschaften der Eingabeschicht werden stärker gewichtet und erlauben eine richtige Klassifikation, selbst wenn einige, weniger reliable - zufällige - Merkmale falsch oder gar nicht dargeboten werden. Dies setzt allerdings voraus, dass Elemente einer Klasse so kodiert werden, dass ihre gemeinsamen Eigenschaften durch das Netzwerk extrahiert werden können. So macht es z.B. wenig Sinn eine „Pixelrepräsentation“ von Hunden einem Netz darzubieten, da es vermutlich schwer sein dürfte, aus dieser Repräsentationsform das Gemeinsame von Hunden gegenüber Katzen zu extrahieren¹³. Abstraktere Eigenschaften wie z.B. Bellen, Fell sträuben etc. dürften hingegen zu einer guten Diskrimination führen. Durch die höhere Gewichtung gemeinsamer Eigenschaften einer Kategorie entsteht eine Prototypenrepräsentation, welche klarer als Einzelexemplare erkannt werden kann, da sie der Gewichtsrepräsentation entspricht. Da die letzten Stimulus-Darbietungen die Gewichte beeinflussen können, ohne dass diese Beeinflussungen durch weitere Darbietungen überlagert werden, lässt sich auch der recency Effekt erklären. Exempleffekte können in begrenztem Maße darauf zurückgeführt werden, dass die Anpassung der Gewichte an die einzelnen Exemplare nicht mehr durch weitere Muster überlagert wurden (vgl. McClelland & Rumelhart, 1985). Shanks (1995) diskutiert diese Phänomene bei Menschen.

¹² Dies mag zwar unter informationstechnischen Gesichtspunkten interessant sein, für psychologische Fragen dürfte dies aber weniger Relevanz besitzen. So ist z.B. noch unklar, wie viele Neurone überhaupt bei Abspeicherungen verwendet werden, und ob es immer die gleiche Anzahl ist. Weiterhin spielt diese Aussage nur dann eine Rolle, wenn die Muster wieder fehlerfrei abgerufen werden sollen. Dies dürfte aber für menschliche Abrufvorgänge eine eher geringe Rolle spielen, da Gedächtnisinhalte häufig leicht verzerrt abgerufen werden. Diese Verzerrung ist bei Menschen wie bei konnektionistischen Netzen, in vielen Fällen von der Ähnlichkeit zu anderen Exemplaren beeinflusst (Roediger & McDermott 2000). Aus Sicht von konnektionistischen Netzen ist diese Verzerrung sogar erwünscht, um Abstraktion zu ermöglichen. Weiterhin korrelieren die gelernten Muster vermutlich sehr stark. Bei Kategorienbildung soll ja gerade auf hochkorrelierte Muster gleich reagiert werden. Das wiederum macht Aussagen zu orthogonalen Mustern irrelevant.

¹³ Erfolgreiche Lösungen dieses Problems beruhen vorwiegend auf einer sehr großen Rechenkapazität.

Biologische Plausibilität¹⁴

Delta-Netzwerke werden gegenwärtig als biologisch eher unplausibel angesehen, da sie einen „Lehrer“ voraussetzen, der „Fehler“ aus der Beobachtung *äußerer* Ereignisse (US) ermittelt, diese dann mit der Aktivität der Eingabeneuronen multipliziert und dann schließlich die Gewichte entsprechend modifiziert. Ein solcher Prozess erscheint im Moment biologisch recht unwahrscheinlich, jedenfalls ist kein zugrundeliegender Mechanismus im Gehirn bekannt (Rolls & Treves, 1998; chapter 5). Andererseits postulieren Gluck, Allen, Myers und Thompson (2001) einen Gehirnmechanismus für die Fehlerkorrektur gefunden zu haben. Die biologische Unplausibilität der direkten Umsetzung der Delta-Regel bedeutet auch nicht, dass nicht ein biologisch plausiblere Implementationsform des Algorithmus existiert.

Bewertung von Delta-Modellen

Delta-Modelle haben sich auch bei Menschen als fruchtbarer Ansatz zur Modellierung von Lösungsprozessen bei Klassifikationsaufgaben erwiesen. Das Rescorla-Wagner Modell wurde vorwiegend bei Fragestellungen eingesetzt, in denen ein Tier unterscheiden musste, welcher Reiz mit einer Belohnung verknüpft ist und welcher nicht. Delta-Modellen hingegen finden hauptsächlich in Situationen Verwendung, in denen Menschen aufgrund einer Stimulussituation Erwartungen an einen Zustand stellen, bzw. zwischen verschiedenen zu erwartenden Ereignisklassen diskriminieren sollen. Gluck und Bower (1988a) zeigten mithilfe verschiedener medizinischer Diagnoseaufgaben die Überlegenheit des Delta-Modells gegenüber „feature-frequency“- (Reed, 1972), Exemplar- (Nosofsky, 1986) und Prototypentheorien (Homa, Sterling, & Trepel, 1981). In den Experimenten von Gluck und Bower lernten Probanden den Zusammenhang unterschiedlicher Symptome mit einer selten und einer häufig auftretenden Krankheit. Schließlich wurden den Probanden die Symptome einzeln präsentiert und die Teilnehmer wurden gefragt, mit welcher Wahrscheinlichkeit die beiden Krankheiten vermutlich vorliegen. Nur das Delta-Modell konnte in allen 3 Experimenten die richtige Rangordnung der Krankheitswahrscheinlichkeiten bei Vorliegen der einzelnen Symptome vorhersagen. Das verwendete Delta-Modell bestand aus einer Eingabeschicht mit 4 Knoten. Jeder dieser Knoten repräsentierte genau ein Symptom durch einen Aktivationszustand von 1 bzw. 0 bei Nichtvorliegen des Symptoms. Alle

¹⁴ Ich denke nicht, dass Überlegungen zur biologischen Plausibilität die Suche auf der Algorithmenebene bereits stark beschränken sollten, da noch lange nicht klar ist, was eigentlich als biologisch plausibel anzunehmen ist. Allerdings dürfte es schon klar sein, dass Vorstellungen, welche möglichst wenig „Homunkulus“-Annahmen in Form von zentralen Prozessverarbeitungsinstanzen benötigen, sich generell, auch bei genauerem Verständnis biologischer Vorgänge, als vorteilhaft erweisen dürften. (vgl. aber auch Kommentar von Bösche 2002 zum Problem der Plausibilität).

Eingabeeinheiten waren über eine gerichtete Kante mit einem einzigen Ausgabeknoten verbunden. Aktivierung des Ausgabeknotens mit Stärke 1 wurde als Vorliegen der einen Krankheit und Aktivierung von -1 als Vorliegen der anderen Krankheit interpretiert. Als Lernalgorithmus wurde der Standard Delta-Algorithmus verwendet. Diese Modellvorstellung ist äquivalent zu einer, in der jede Krankheit durch einen separaten Ausgabeknoten repräsentiert wird, und der stärker aktivierte Knoten die Entscheidung bestimmt. Die Gewichtsveränderungen, die im Modell mit einem Ausgabeknoten direkt verrechnet werden, verteilen sich bei zwei Ausgabeknoten zwar auf verschiedene Gewichte, werden dann aber durch den Vergleich schließlich auch linear verrechnet (vgl. Gluck & Bower, 1988a, Fußnote 2).

Chapman und Robbins (1990) demonstrierten mittels eines simulierten Börsenszenarios die Überlegenheit von Modellen, in denen Lernen mit Rückmeldungen durch einen Prozess mit „historischen“ Wirkungen bestimmt wird, gegenüber normativen Modellen wie z.B. Δp (vgl. Abschnitt 1.2.1). In dieser Studie gelang es Chapman und Robbins Δp für die „geblockten“ und „ungeblockten“ Reize in der Kontroll- als auch in der Experimentalgruppe gleichzuhalten. Dennoch zeigte sich der Blocking-Effekt. Damit konnten Chapman und Robbins (1990) zeigen, dass ein normatives Modell, welches nur die objektive Paarung eines Hinweisreizes mit einem Ausgang sowie des Ausganges ohne Vorliegen eines Hinweisreizes beachtet, ihre Daten nicht erklären kann. Der Grund dafür besteht darin, dass sowohl Blocking als auch konditionierte Inhibition, welche sie in einem zweiten Experiment nachweisen konnten, von der Reihenfolge der Darbietung der Reize und nicht dem absoluten Zusammenhang abhängen. Weitere Plausibilität erhält das Delta-Modell durch Befunde von Gluck, Allen, Myers und Thompson¹⁵ (2001). Die Autoren postulieren neuronale Grundlagen des Rescorla-Wagner Modells für das Erlernen motorischen Verhaltens gefunden zu haben.

Durch die konkrete Implementierung erlauben Delta-Netzwerke ordinale Aussagen bezüglich der für ein Problem benötigten Zeit. Häufig operationalisierte Variablen stellen dafür die Anzahl an nötigen Operationen im Delta-Netzwerk bzw. die Aktivationsstärke der

¹⁵ Interessanterweise wird das Lernen in ihrem Modell nicht von einer Differenz zwischen einem theoretischen Wert λ und einer Erwartung aufgrund eines CS angetrieben, sondern die Passung zwischen US und CR bestimmt das Ausmaß des Lernens. Ihr neuroanatomisch inspiriertes Modell umgeht somit ein „höheres“ Modul, welches äußere Beobachtungen der „Außenwelt“ und deren Vorhersagen rückmeldet. Vielmehr verlagert es die Feedbackschleife in etwa auf das Verarbeitungsniveau des Stimulus und benutzt daher nur lokal vorhandene Information. Mit dieser Modellvorstellung ließe sich erklären, warum ein inhibitorischer Reiz seine Stärke nicht in der Extinktionsphase einbüßt, wie das Rescorla-Wagner Modell fälschlich vorhersagt. Da nur der CR und nicht die Vorhersage durch V_x den Lernvorgang beeinflusst, tritt in der Extinktionsphase, in der *kein* CR und kein US auftritt, auch keine lernrelevante Differenz auf. Durch die Einbeziehung von CR und US in die Bewertung dürfte eine Nähe zu instrumentellem Lernen vorhanden sein. Tatsächlich zeigen sich Blocking-Effekte auch in instrumentellen Paradigmen. In diesen Paradigmen verhindert ein Verhalten welches sich in früheren Durchgängen als erfolgreich erwiesen hat das Erlernen eines neuen Verhaltens (z.B. Rodrigo, Chamizo, McLaren & Mackintosh, 1997).

Ausgabeeinheiten dar. Der letzte Fall beruht auf der Vorstellung, dass Ausgabeeinheiten miteinander um die Antwort wettstreiten und bei einem klaren Vorteil einer Ausgabeeinheit die Antwort schneller erfolgen kann (vgl. Wills, Reimers, Stewart, Suret & McLaren, 2000). Die Idee des begrenzten Lernens von Signalen scheint gut mit neueren Entscheidungstheorien übereinzustimmen, die rationale Entscheidungsmodelle durch heuristische Entscheidungsmodelle ersetzen. In heuristischen Entscheidungsmodellen wird nicht eine vollständige Informationsverarbeitung sondern eine für die Zwecke ausreichende erwartet (Gigerenzer & Todd, 1999). Diese Bevorzugung nicht vollständiger Analysen der Umwelt entsteht dadurch, dass nicht nur der Informationsgewinn, der durch weitere Suche entsteht, beachtet wird, sondern auch, dass durch die längere Suche und Verarbeitung Kosten in Form von Zeit oder Ressourcen entstehen. Indem dieser Trade-off zwischen optimaler Lösung und Ressourcenverlust in die Kosten-Nutzen Kalkulation einbezogen wird stellen sich bei Annahmen über die Umwelt, wie sie heute als realistisch angesehen werden, die heuristischen Verfahren als überlegen heraus. Schließlich bietet sich mit der Delta-Regel ein genau spezifizierter Mechanismus an, nicht nur das Endresultat einer Kategorisierung zu beschreiben, sondern auch den Entstehungsprozess zu modellieren. Nicht zuletzt die Möglichkeit durch Gebrauch möglichst vieler Informationsquellen auch bei fehlerhaften oder unzureichenden Informationen, wie sie in realistischen Umwelten zu erwarten sind, dennoch relativ robuste Ergebnisse zu liefern, sprechen für distribuierte Modelle, wie z.B. Delta-Netzwerke.

1.2.3 Konfigurale Modelle

Um Lösungen für das oben diskutierte Problem nicht-linear separierbarer Aufgaben im Rahmen des Delta-Modells ohne eine zusätzliche Schicht zu ermöglichen, entwickelten sich im Wesentlichen zwei Modelle. Gluck und Bower (1988b, Gluck 1991,1992) blieben eng am Delta-Modell und forderte nur, dass die Reize nicht mehr nur in ihren einzelnen Merkmalen sondern auch in ihrer Kombination repräsentiert werden (siehe auch Goldstone, Steyvers, Spencer-Smith, & Kersten, 2000, p. 220). Reize wie z.B. Husten und Fieber sollen also nicht nur durch einzelne Merkmalseingabeknoten für Husten und Fieber kodiert werden, sondern auch durch konfigurale Knoten, welche sowohl das Merkmal Husten als auch Fieber gleichzeitig repräsentieren. Mit dieser Modellierung konnten erfolgreich ordinale Aussagen über die Schwierigkeit, die Versuchspersonen bei der Lösung unterschiedlicher Kategorisierungsaufgaben zeigen, gemacht werden (Gluck & Bower, 1988b).

Pearce (1994) hingegen entfernt sich etwas weiter von der Rescorla-Wagner Vorstellung, indem er fordert, dass erst bei Auftreten eines neuen Reizes ein eventuell konfiguraler Exemplarknoten angelegt wird. Die Aktivierungsstärke der bereits angelegten Exemplarknoten berechnet sich dann aus der Gemeinsamkeit mit dem gerade präsentierten Reiz. Lernen findet durch Fehlerminimierung zwischen der asymptotisch zu erreichenden Stärke und der über alle Exemplarknoten einlaufenden Aktivität statt. Geändert wird allerdings nur die Verbindung zum maximal aktivierten Exemplarknoten:

$$\Delta E_A = \alpha\beta(\lambda - V_A) \quad \text{Gleichung 13}$$

α und β geben wieder wie im Rescorla-Wagner Modell die Salienz bzw. Lerngeschwindigkeit an, E_A die assoziative Stärke des am stärksten aktivierten Exemplarknoten A, V_A entspricht der *aufsummierten* einlaufenden Aktivität bei Präsentation des Reizes A. Sie setzt sich zusammen aus der Verbindungsstärke von Exemplarknoten A und eventuell weiteren Exemplarknoten AB, AC ..., welche A enthalten. Diese Exemplarknoten werden entsprechend ihrer Ähnlichkeit zu A gewichtet.

Bösche (2002) konnte zeigen, dass die asymptotischen Gewichte eines konfiguralen Delta-Netzwerkes den Assoziationsstärken im Pearce Modell entsprechen, falls die asymptotischen Gewichte des Delta-Netzwerkes unabhängig von den Startgewichten sind. Die asymptotischen Gewichte sind dann unabhängig von den Startgewichten, wenn die Korrelationsmatrix invertierbar ist. Dies ist bei linear unabhängigen Eingabevektoren immer der Fall. Für den Fall, dass mehrere asymptotische Gewichte möglich sind, versucht die Delta-Regel die quadrierte Summe der Gewichte zu minimieren, während das Pearce-Modell höhere Assoziationsstärken für Merkmalskonfigurationen höherer Ordnung bevorzugt.

Da das Pearce-Modell, wie Bösche (2002) ausführt, keine wesentlichen Phänomene besser als ein konfigurales Delta-Modell, dafür aber nur mit zusätzlichen Annahmen, Effekte wie z.B. Summation erklären kann, dürfte das Delta-Modell das geeignetere Modell für Kategorisierungsaufgaben sein. Summation bezeichnet die stärkere Reaktion auf den kombinierten Reiz AB als auf die Einzelreize A und B, wenn in einer vorhergehenden Phase A und B nur alleine aufgetreten sind.

Nachdem nun die Grundlagen von Modellen beschrieben wurden, welche aufgrund von Fehlerrückmeldungen lernen, wird im folgenden Abschnitt auf ein Modell eingegangen, welches ohne jegliche Rückmeldung Informationen nach Ähnlichkeiten gruppiert.

1.2.4 Kompetitives Lernen

Modelle kompetitiven Lernens wurden vorwiegend in den 70 Jahren des letzten Jahrhunderts entwickelt und eingesetzt, um Muster nach ihrer Ähnlichkeit zu klassifizieren. Wesentliche Wegbereiter waren von der Malsburg, Grossberg und Kohonen. Worin diese Ähnlichkeit, nach der klassifiziert werden soll, besteht, wird durch die Form der Kodierung mitbestimmt. So können abstrakt repräsentierte Stimuli eine abstraktere Ähnlichkeit innerhalb einer Klasse aufweisen als dieselben Stimuli bei konkreter Kodierung. Angewendet auf das oben diskutierte XOR-Problem hängt die Lösbarkeit davon ab, ob die Stimuli orthogonal kodiert werden oder ob die Repräsentationen der verborgenen Schicht z.B eines Delta-Netzwerkes als Eingabe verarbeitet werden. Die Grundidee kompetitiven Lernens besteht darin, dass sich einzelne *typische* Repräsentanten von Gruppen dadurch bilden, dass die einzelnen Repräsentanten miteinander um die jeweiligen Stimuli streiten bzw. sich gegenseitig hemmen. Je nach Stärke und Form der Hemmung können unterschiedliche Repräsentantenkonfigurationen entstehen. Im Extremfall gewinnt nur ein typischer Repräsentant und alle anderen Repräsentanten sind inaktiv. Gewinnen bedeutet in diesem Kontext, dass ein Repräsentant ähnlicher als alle anderen Repräsentanten zu dem dargebotenen Stimulus ist. In diesem Fall verändert der Repräsentant sich etwas in Richtung des hinzugewonnenen Stimulus und wird dadurch bei der nächsten Präsentation desselben Stimulus diesen mit einer größeren Wahrscheinlichkeit wieder an sich binden. Dadurch entstehen einzelne Klassen ähnlicher Stimuli, in denen Redundanzen zusammengefasst werden können. Dadurch, dass sich der Gewinner in die Richtung eines neuhinzugewonnenen Stimulus verändert, wird er gleichzeitig den anderen Stimuli in seinem Cluster etwas unähnlicher. Damit erhält ein anderer Repräsentant eine Chance, diese mittlerweile eher „untypischen“ Stimuli an sich zu binden. Dieses Prinzip lässt sich relativ einfach in konnektionistischen Systemen realisieren. Das Ziel der Lernregel besteht dann darin, Gewichte zwischen den Eingabeeinheiten und den Ausgabeneinheiten zu finden, so dass ähnliche Stimuli möglichst stark mit nur einer Ausgabeneinheit assoziiert sind. Die Ausgabeknoten repräsentierten dann die entsprechenden Cluster. Da die Gewichte jeweils an den Eingabevektor angepasst werden, bestimmt die Ähnlichkeit bzw. Korrelation eines Eingabemusters mit den vorherigen Mustern die Aktivität an den Ausgabeneuronen. Etwas abstrakter lässt sich dieser Prozess dadurch beschreiben, dass die Summe der Eingabeknoten, multipliziert mit den Gewichten zu einem Ausgabeknoten, für ähnliche Stimuli maximal sein soll. Dies ist dann der Fall, wenn der Gewichtsvektor dem Eingabevektor dahingehend entspricht, dass aktive Merkmale mit einem positiven Gewicht multipliziert werden. Die Ähnlichkeit wird also mittels eines korrelativen

Maßes ermittelt. Wenn bei allen Eingabeneinheiten Inaktivität durch 0 repräsentiert wird, sollte der Gewichtsvektor bei allen von 0 verschiedenen Zuständen ein positives Gewicht haben. Um zu verhindern, dass eine Ausgabeinheit starke Gewichte zu allen Stimuli ausbildet, wird eine Gewichtsrestriktion eingeführt:

$$\sum_i w_{ij}^2 = 1$$

Hiermit wird gewährleistet, dass jede Ausgabeinheit nur eine beschränkte Assoziierbarkeit aufweist. Einfachheitshalber und da es für die Strukturfindung keine große Rolle spielt, wird diese Randbedingung häufig gelockert zu:

$$\sum_i w_{ij} = 1$$

Allerdings konnten Sutton und Reggia (1994) zeigen, dass die einfache Summenrestriktion zu Gewichtsvektoren unterschiedlicher Länge führt. Längere Gewichtsvektoren besitzen einen Vorteil, da bei gleicher Ähnlichkeit, gemessen in der Winkelgröße zwischen Eingabe- und Gewichtsvektor, ein größeres inneres Produkt erzeugt wird. Die Folge dieser Bevorzugung von längeren Vektoren ist, dass der Inputraum nicht so zwischen zwei Gewichtsvektoren aufgeteilt wird, dass der Gewichtsvektor gewinnt, welcher den kleinsten Winkelabstand zum Eingabevektor hat, sondern so, dass der gesamte Inputraum zur Hälfte partitioniert wird (vgl. Abbildung 1.10). Dies führt zu der überraschenden Situation, dass ein Gewichtsvektor selbst einen Eingabevektoren, welcher ihm entspricht, verliert. Seien z.B. w_1 und w_2 zwei Gewichtsvektoren und a ein Eingabevektor:

$$w_1 = \begin{pmatrix} 1 \\ 9 \end{pmatrix}, w_2 = \begin{pmatrix} 4 \\ 6 \end{pmatrix}, a = \begin{pmatrix} 4 \\ 6 \end{pmatrix}$$

so gilt:

$$w_1 * a = \begin{pmatrix} 1 \\ 9 \end{pmatrix} * \begin{pmatrix} 4 \\ 6 \end{pmatrix} = 4 + 54 > w_2 * a = \begin{pmatrix} 4 \\ 6 \end{pmatrix} * \begin{pmatrix} 4 \\ 6 \end{pmatrix} = 16 + 36.$$

Damit führt die simplere Gewichtsrestriktion zu falschen Reaktionen, falls die Gewichtsvektoren nicht gleichmäßig im Inputraum verteilt sind. Die Gewichtsvektoren passen sich nicht mehr „gleichberechtigt“ an die Inputvektoren an, sondern längere Gewichtsvektoren haben eine größere Wahrscheinlichkeit einen Eingabevektoren zu gewinnen.

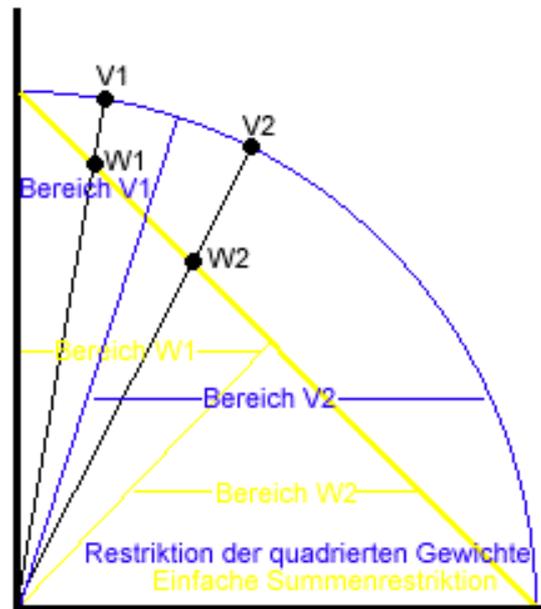


Abb. 1.10: nach Sutton und Reggia (1994). Die einfache Gewichtsrestriktion führt dazu, dass die verschiedenen Gewichtsvektoren auf einer Hyperebene liegen (helle Linie). Die Restriktion der quadrierten Gewichte hingegen zwingt alle Gewichtsvektoren auf eine Hypersphäre. Weiterhin eingezeichnet ist eine helle Linie, welche den dargestellten Bereich in zwei gleich große Teile partitioniert. Alle Eingabevektoren welche in den unteren Bereich fallen werden Vektor W2 zugeordnet, alle Gewichtsvektoren im oberen Bereich dem Vektor W1. Die dunkle Partitionierung findet hingegen zwischen den beiden Gewichtsvektoren statt, so dass der Winkelabstand zwischen Eingabevektor und Gewichtsvektor über den Gewinner entscheidet. Im Falle der hellen Partition würde selbst ein Eingabevektor welcher W2 entspricht W1 zugeordnet werden.

Für die Aktivierungsstärke, welche bei Clusterunit j eingeht, gilt:

$$o_j = \sum_i w_{ij} a_i = \sum_{a_i \neq 0} w_{ij} a_i$$

w_{ij} wieder das Gewicht der Eingabeeinheit i zur Ausgabeeinheit j , a_i die Aktivierungsstärke der Eingabeeinheit i .

Um bei *begrenzter* Assoziationsfähigkeit einen möglichst hohen Wert zu erzielen, sollte w_{ij} die relative Ausprägung der a_i widerspiegeln und nur bei $a_i \neq 0$ einen Wert >0 annehmen. Die Anpassung an die verschiedenen Eingabemuster gelingt dann möglichst gut, wenn die w_{ij} die „Typikalität“ des Merkmals i für die Mustermenge, die der Ausgabeeinheit j zugeordnet wird, widerspiegeln. Rumelhart und Zipser (1986) zeigten formal, dass die für eine Klasse typischen Merkmale bzw. Eingabeeinheiten höher bewertet werden als untypische. Die

Autoren konnten zeigen, dass das Gewicht w_{ij} im Equilibrium sogar proportional zu der Wahrscheinlichkeit ist, dass Eingabeeinheit i aktiviert ist, falls Ausgabeknoten O_j gewinnt. Als Equilibrium ist der Zustand definiert, in dem keine Veränderung der Gewichtsvektoren als Reaktion auf Präsentationen der Muster mehr stattfindet.

Rumelhart und Zipser (1986, pp. 168) versuchten den Lernvorgang dadurch zu veranschaulichen, dass die einzelnen Stimuli als Vektoren in einem durch die Merkmale der Stimuli aufgespannten Raum angesehen werden. Wenn sowohl die Quadrate der Eingabe- als auch der Gewichtsvektorenelemente sich zu 1 summieren, können auch die Gewichtsvektoren der einzelnen Ausgabeknoten als auf der Oberfläche einer Hypersphäre liegend vorgestellt werden. Der Effekt des Lernalgorithmus besteht darin, die Gewichtsvektoren möglichst nahe an die Stimulusvektoren heranzuführen. Herangeführt wird dabei immer nur der Gewichtsvektor, welcher einem gerade präsentierten Stimulus bereits am nächsten ist. Dies gilt aber, wie oben gezeigt, nur für den Fall gleicher Vektorlängen. Für diesen Fall lässt sich mathematisch zeigen (z.B. Hertz, Palmer & Krogh, 1991), dass der Gewichtsvektor die quadrierten Abstände zu den ihm zugeordneten Mustern minimiert. Formal ergibt sich die Lernregel zu (vgl. Rumelhart & Zipser, 1986):

$$\Delta w_{ij} = \begin{cases} 0 & \text{Falls Einheit } j \text{ bei Vorliegen von Stimulus } k \text{ verliert} \\ \beta \frac{a_i}{n} - \beta w_{ij} & \text{Falls Ausgabeknoten } j \text{ bei Stimulus } k \text{ gewinnt.} \end{cases} \quad \text{Gleichung 14}$$

$n = \sum_i a_i$; β Lernparameter; $a_i = 1$, falls Merkmal i aktiv, sonst 0.

„ n “ stellt eine Normierungskonstante dar, welche dafür sorgt, dass die Gewichtsvektoren sich annähernd unabhängig von der Anzahl aktivierter Eingabeeinheiten an die Muster anpassen. Dass mit dieser Formel das Gesamtgewicht zu einem Ausgabeknoten j gleich bleibt sieht man, wenn man sich verdeutlicht, dass für jede Stimuluspräsentation über alle i Eingabeknoten summiert wird: Gilt $\sum_i w_{ij} = 1$ so folgt:

$$\Delta w_{j+} = \sum_i \frac{\beta a_i}{n} = \beta \frac{\sum_i a_i}{n} = \beta \quad \text{nach Gleichung 14 für die positiven Gewichtsveränderungen}$$

und $\Delta w_{j-} = \beta \sum_i w_{ij} = \beta$ für die Gewichtsabnahmen

Ist am Anfang die Summenrestriktion $\sum_i w_{ij} = 1$ erfüllt, so gilt also $\Delta w_{j+} = \Delta w_{j-}$ und die Gesamtassoziationsstärke bleibt erhalten. Gilt die Summenrestriktion nicht bereits bei der Initialisierung des Netzes, so wachsen bzw. fallen die Gewichtsstärken bis die Summenrestriktion gilt¹⁶.

Da die Gewichte in gewisser Weise den Mittelwert der präsentierten Stimuli darstellen, wird durch ihre Subtraktion in Gleichung 14 eine gewisse Orthogonalisierung der einzelnen Eingabemuster erreicht, vor allem zwischen nicht sehr ähnlichen Mustern. Die Interferenz wird damit verringert.

Rumelhart und Zipser (1986) schlagen vor, den gewinnenden Output-Knoten o_{winner} „1“ und allen anderen die Aktivität „0“ zuzuweisen. Deshalb lässt sich Gleichung 14 auch umformulieren zu:

$$\Delta w_{ij} = \beta o_j \frac{a_i}{n} - \beta o_j w_{ij} \quad \text{Gleichung 15}$$

Der erste Teil von Gleichung 15 verwirklicht das Hebbsche Prinzip der Verstärkung gemeinsam feuender Neurone (o_j und a_i). Aktive Eingabeeinheiten, Knoten also die den Gewinner unterstützten, werden dadurch mit einem höheren Gewicht „belohnt“. Eingabeeinheiten, die nicht feuerten und somit nicht zum Gewinn beitrugen, werden hingegen nur abgeschwächt (rechter Teil von Gleichung 15). Da das innere Produkt zweier gleich langer Vektoren das Maximum bei Gleichheit annimmt, ist mit dem Hebbschen Prinzip eine möglichst starke Aktivierung eines Ausgabeknotens garantiert. Der zweite Teil der Gleichung führt zu einer allgemeinen Gewichtsreduktion und garantiert damit die Summenrestriktion.

Um die Strukturfindung durch Gewichtsvektoren der gleichen Länge zu verbessern, kann die Randbedingung zu $\sum_i w_{ij}^2 = 1$ verschärft werden. Der hierfür benötigte Lernalgorithmus entspricht der Regel von Oja (1982):

$$\Delta w_{ij} = \beta o_j \frac{a_i}{n} - \beta o_j^2 w_{ij} \quad \text{Gleichung 16}$$

Für einen formalen Nachweis, dass dabei die Länge der Gewichtsvektoren erhalten bleibt, siehe z.B. Hertz, Palmer und Krogh (1991). Auch nichtlineare Entscheidungsfunktionen

¹⁶ Dies folgt aus der Tatsache, dass Δw_{j+} unabhängig von den Gewichtsstärken ist und damit Δw_{j-} die Gesamtgewichtsstärke reguliert. Ist die Gesamtgewichtsstärke < 1 , so ist Δw_{j-} kleiner als Δw_{j+} und die Gesamtgewichtsstärke nimmt zu. Ist die Gesamtgewichtsstärke > 1 ist Δw_{j-} größer als Δw_{j+} und die Gesamtgewichtsstärke nimmt ab.

lassen sich durch eine weitere, von der Aktivierungsstärke o_j abhängige Lernkonstante implementieren.

Zusammenfassend lässt sich also sagen, dass in kompetitiven Modellen die Gewichte von den nichtaktiven Knoten zu den aktiven Knoten mit der Lernrate β umverteilt werden. Das Ziel besteht darin, den Gewichtsvektor zum Gewinner möglichst gut dem Eingabevektor anzupassen. Abweichungen davon lassen sich im Sinne des oben diskutierten Rescorla-Wagner Modells als *intern*¹⁷ generierte Fehler interpretieren. Die „Lernrate“ β gibt an, wie schnell diese Anpassung stattfindet. Das System versucht die Ähnlichkeit innerhalb der Gruppen maximal, und zwischen den Gruppen minimal werden zu lassen. Weiterhin werden gleich große Gruppen bevorzugt. Die erste Bedingung wird dadurch erreicht, dass sobald ein Muster eine größere Ähnlichkeit zu einem anderen Cluster aufweist, das andere Cluster gewinnt. Die zweite Bedingung wird dadurch realisiert, dass Ähnlichkeiten zwischen den Gruppen die Zuordnung eines Musters „springen“ lässt und eine stabile Konfiguration nur möglich ist, wenn die Zuordnungen nicht mehr springen. Da stabile Konfigurationen per Definition selten verlassen werden, sind stabile Konfigurationen häufig der Endpunkt einer Klassifizierung. Die annähernd gleiche Gruppengrößen werden durch die Beschränkung der Gesamtassoziationsstärke eines Ausgabeknotens erreicht. Wenn ein Ausgabeknoten mehr Muster bindet als ein anderer, so sind die „durchschnittlich angepassten“ Gewichtsstärken geringer und einzelne Muster gehen leichter verloren. Kompetitive Netze teilen Eingabemuster in Klassen entsprechend der Stimulusstruktur auf. Bei hoch strukturierten Stimuli werden die Klassen auch relativ stabil (Rumelhart & Zipser, 1986, p. 167). Im Falle der Kodierung unterschiedlicher Klassen durch orthogonale Eingabestimuli kann eine vollkommen stabile Abbildung der Eingabevektoren auf Ausgabevektoren erfolgen. Die gewählte Ähnlichkeitsstruktur hängt auch von der Anzahl an vorgegebenen Ausgabeneuronen bzw. Klassen ab. Rumelhart und Zipser (1986) verdeutlichen dies am Beispiel eines kompetitiven Netzes, dem Muster der Form AA, AB, BA und BB präsentiert wurden. Bei zwei vorgegebenen Klassen differenzierte das Netz nach dem Buchstaben an erster Position. Bei anderen Simulationsdurchläufen kann es natürlich auch zu einer Einteilung kommen, die nur auf den Buchstaben an der zweiten Stelle achtet. Bei 4 Klassen konnte jede Buchstabenkombination einen eigenen Gewichtsvektor für sich gewinnen. Ebenso hängt die Endkonfiguration auch von der ursprünglichen, meistens zufällig gewählten Gewichtsverteilung und der Präsentationsreihenfolge der Stimuli ab (ebda.). In welche Klasse

¹⁷ „Intern“ meint hier, dass die Fehlerbestimmung keine Beobachtung eines „äußeren“ US oder Rückmeldung durch einen „Lehrer“ benötigt.

also ein Stimulus eingeteilt wird, kann nicht alleine aufgrund der Stimulusstruktur vorhergesagt werden. Diese Sensitivität für Anfangsbedingungen und Präsentationsreihenfolge kommt durch einen Symmetriebruch zustande. Dieser Symmetriebruch entsteht dadurch, dass die Einteilung eines Stimulus in eine Kategorie die dazuführende Entscheidungsgrundlage (Distanz zum Gewichtsvektor) verändert. Diese Einteilung hat damit Einfluss auf weitere Einordnungen, was den „Wettkampf“ für die anderen Gewichtsvektoren verändert. Durch diese rückgekoppelte Schleife entstehen somit Verläufe, die aus den makroskopischen Bedingungen der Stimulusstruktur nicht vorhersagbar sind. Die Stabilität der Kategorisierung hängt dabei stark von der Stabilität der Merkmalskorrelationen über die verschiedenen Muster einer Klasse hinweg ab.

Um zu verhindern, dass sich die Gewichtsvektoren vom Anfang an im gleichen Raumbereich massieren, werden die Gewichtsvektoren zufällig initialisiert. Rolls und Treves (1998, chapter 4) sehen einen wesentlichen Vorteil von kompetitiven Netzen darin, dass zwei niedrig korrelierende Eingabevektoren auf unterschiedliche orthogonale Ausgabevektoren und gleichzeitig stark korrelierende Eingabevektoren auf noch stärker korrelierende Ausgabevektoren abgebildet werden können. Diese Eigenschaft erlaubt kompetitiven Netzen, Strukturen in Daten herauszufiltern, auf die in einem späteren Verarbeitungsschritt andere Netze zugreifen können. Die Orthogonalisierung ist eine Folge davon, dass eher unähnliche Stimuli durch unterschiedliche Gewichtsvektoren verarbeitet werden und damit eine noch größere Unähnlichkeit entsteht. Hingegen maximiert eine gleiche Kategorisierungsentscheidung aufgrund sehr ähnlicher Eingabevektoren deren Ähnlichkeit in der Eingabeform zur Identität in der Ausgabeform. Wann von Unähnlichkeit bzw. Ähnlichkeit zu sprechen ist, hängt jeweils von der Gesamt-Stimulusstruktur und der Anzahl an Ausgabeneuronen und damit verfügbarer Gewichtsvektoren ab. So ist es möglich nicht-linear-lösbare Probleme für Delta-Netzwerke vorzubereiten. Die Vektoren (01), (10) und (11) können durch ein kompetitives Netz mit zwei Eingabeknoten und drei Ausgabeknoten drei unterschiedlichen Kategorien zugewiesen (vgl. Rolls und Treves, 1998; chapter 4) und diese Lösung dann an ein Delta-Netzwerk weitergegeben werden. Dieses „Vorschalten“ eines kompetitiven Netzes erbringt einen starken Zeitvorteil gegenüber drei-schichtigen Netzen, welche mit einer abgeänderten Delta-Regel lernen (Hertz, Palmer & Krogh, 1991). Durch ihre Reduzierung der Eingabe auf wenige strukturerhaltende Ausgabeneurone eignen sich kompetitive Netze auch als Vorverarbeitungsstufe der Informationsspeicherung. Mittels einer kleinen Veränderung in der Verknüpfungsregel zwischen den Ausgabeknoten können topologische Karten erzeugt werden. Verstärken aktivierte Ausgabeneurone räumlich oder funktional benachbarte Knoten

und werden gleichzeitig weit entfernte Ausgabeneurone gehemmt, so führt dieser Mechanismus dazu, dass ähnliche Eingabevektoren auf benachbarte Ausgabeneurone abgebildet werden.

Biologische Plausibilität

Kompetitive Netzwerke sind prinzipiell biologisch plausibel, da sie nur „lokale“ Informationen benötigen. Die Längennormierung bzw. daraus resultierende Aktivitätsnormierung des Gewichtsvektors sowie des Eingabevektors kann durch feedforward Inhibition mittels Interneuronen als auch durch Inhibition in vorverarbeitenden Neuronen gedacht werden (vgl. Rolls & Treves, 1998; chapter 4). Der „hebbische“¹⁸ Mechanismus lässt sich leicht durch heterosynaptische Langzeitunterdrückung bzw. Langzeitpotenzierung verstehen. Im ersten Fall wird eine geringere postsynaptische Aktivierung erzielt, falls die präsynaptische Aktivierung geringer als die postsynaptische ausfällt, im zweiten Fall wird, wenn die präsynaptische und postsynaptische Aktivierung hoch ist, die postsynaptische verstärkt. Rolls und Treves (1998; chapter 4) vermuten, dass kompetitive Netze im Hippocampus Informationen für die Gedächtnisabspeicherung vorbereiten.

Prototypenabstraktion

Da der Gewichtsvektor für Muster, welche in eine Klasse gehören, eine Art Mittelwert darstellt, reagiert das Netzwerk mit der stärksten Aktivierung auf den Mittelwert, bzw. den Prototypen einer Kategorie. Das gilt selbst dann, wenn der Prototyp selbst noch nicht präsentiert wurde. Wenn einzelne Features reliabel eine Kategorie vorhersagen, während andere Merkmale unreliabel sind, bewertet der Gewichtsvektor die reliablen Eigenschaften stärker. Aus diesem Grund können Muster, die nur unvollständig oder mit „falschen“ Werten an unreliablen Stellen präsentiert werden, dennoch klassifiziert werden. Diese Phänomene stimmen gut mit Beobachtungen an Menschen überein (z.B. Shanks, 1995). Recency- und Exemplar-Effekte beruhen bei kompetitiven Netzwerken auf denselben Mechanismen wie bei Delta-Netzwerken.

Bewertung kompetitiver Modelle

Kompetitive Modelle eignen sich sehr gut dazu, Muster aus Eingabeinformationen zu extrahieren. Diese Musterextraktion erfolgt auf der Basis von Ähnlichkeiten der

¹⁸ „hebbische“ Mechanismus wird hier in Anführungszeichen gesetzt, da es sich dabei nicht um den ursprünglich von Hebb postulierten Mechanismus handelt. Hebb (1949) erwartete nur eine Verstärkung zwischen kausal sich beeinflussenden Neuronen. Die Modellierung hier verlangt aber auch eine Modifizierung der Verbindungsstärke zwischen aktiven Ausgabeknoten und *nichtaktiven* Eingabeknoten.

Aktivierungsmuster bzw. Eingabevektoren. Wenn die Grundannahme, dass von der Ähnlichkeit im Eingabevektor auf die Ähnlichkeit im Ausgabevektor geschlossen werden kann, jedoch nicht gilt, so muss entweder eine neue Eingabekodierung erfolgen oder ein „Lehrer“ im Sinne des Delta-Netzwerkes eingesetzt werden. Damit dürften kompetitive Netzwerke bei komplizierten Strukturen (i.e. Strukturen, die nicht aus der Anzahl gemeinsamer Features entstehen) weniger geeignet sein. Inwieweit solche „komplizierten“ Strukturen bei menschlichen Klassifikationsaufgaben eine Rolle spielen ist unklar.

Kompetitive Netzwerke eignen sich gut als Vorverarbeitungsstufe von Delta-Netzwerken. Kompetitive Netze können als Vorverarbeitung eingesetzt vorhandene Strukturen extrahieren und, da der Ausgabevektor sparsamer als der Eingabevektor kodiert, Eingaben orthogonalisieren (z.B. Rumelhart & Zipser, 1986). Durch den Wettstreit zwischen den Ausgabeknoten werden verschiedene Parameter gleichzeitig optimiert. So werden die Gruppen annähernd gleich groß gewählt, die Heterogenität innerhalb der Gruppen minimiert und zwischen den Gruppen maximiert. Während in Delta-Netzen eine „Umgewichtung“ des Eingabevektors erlernt werden muss, lernen kompetitive Netze den Eingabeknoten wieder in ihren Gewichten zu reproduzieren. Da bei dieser Tatsache das Lernen des Gewichtvektors unabhängig von der Prädiktivität eines Reizes erfolgt und das Ziel darin besteht mit dem Gewichtvektor den Eingabevektor möglichst genau anzupassen, sind kompetitive Netzwerke generell nicht in der Lage Blocking-Phänomene zu zeigen. Im technischen Bereich werden kompetitive Modelle vor allem zur Mustererkennung verwendet.

1.3 Kategorisierungsmodelle

Eine umfassende Darstellung der wesentlichen Kategorisierungsmodelle für Paradigmen mit Rückmeldung würde den Rahmen dieser Arbeit sprengen. Vielmehr soll nur kurz erwähnt werden, dass eine Vielzahl sowohl beschreibender Modelle (z.B. Anderson 1991) als auch Prozessmodelle (Kruschke & Johansen, 1999; Gluck & Bower, 1988a) existiert. Wichtig für den Zweck dieser Diplomarbeit ist lediglich, dass die wesentlichen Prozessmodelle eine Form des fehlergetriebenen Lernens zeigen, welches typisch für das Rescorla-Wagner Modell ist. Selbst regelbasierte Kategorisierungsmodelle (z.B. Anderson & Betz, 2001) ließen sich noch unter diese Kategorie fassen. In vielen instanzbasierten und konnektionistischen Modellen gehen alle Stimulusdimensionen in die Entscheidung über die Kategorienzuordnung ein. Allerdings kann eine Gewichtung der Dimensionen aufgrund von Erfahrung mit dem Material erfolgen (z.B. Nosofsky, 1986; Kruschke & Johansen, 1999).

Um das Rescorla-Wagner Modell auf Kategorisierungsphänomene anzuwenden, kann der Präsentationsstimulus als Aktivierung der Eingabeknoten, die Entscheidung der Versuchspersonen als Aktivierung eines Ausgabeknotens und die Rückmeldung seitens des Experimentators als die Differenzrückmeldung verstanden werden. Dadurch, dass um ein Ziel zu erreichen die Konsequenzen von *Handlung* verglichen werden, ändert sich das Paradigma vom klassischen Konditionieren zum operanten Konditionieren. Obwohl das Rescorla-Wagner Modell im Wesentlichen als Erklärung für klassische Konditionierung eingesetzt wurde, lässt es sich leicht auf operante Konditionierung übertragen. Nahegelegt wird diese Erweiterung auch durch das Vorhandensein von Blockingphänomenen bei instrumentellem Lernen (z.B. Rodrigo, Chamizo, McLaren & Mackintosh, 1997). In der Studie ignorierten Ratten in einem Labyrinth einen vierten Orientierungspunkt, wenn in einer vorhergehenden Phase nur drei Orientierungspunkte benötigt worden waren. Auch die neurophysiologische Verankerung durch Gluck, Allen, Myers und Thompson (2001) legt diese Interpretation wie oben bereits ausgeführt nahe.

1.4 Klassifikationsmodelle

Anders sieht es bei Modellen für unüberwachte Kategorienbildung aus. Hier finden sich nur sehr wenige Modelle. Prozessmodelle stellen das Modell von Love, Medin und Gureckis (eingereicht) bzw. das Rumelhart und Zipser Modell (1986) dar. Pothos und Chater (2001) entwickelten ein auf einem Einfachheitskriterium beruhendes beschreibendes Modell.

Studien im Paradigma der freien Klassifizierung verwenden häufig auf Karteikarten abgebildete einfache Stimuli, welche in von den Versuchspersonen selbst zu bildende Kategorien einsortiert werden sollen (z.B. Preusser & Handel, 1970; Ahn & Medin, 1992; Medin, Wattenmaker, Hampson, 1987). Medin, Wattenmaker und Hampson (1987) stellten fest, dass natürliche Kategoriensysteme Ähnlichkeitsstrukturen besitzen, welche sich über die Anzahl gemeinsamer Merkmale erklären lassen. Hingegen klassifizierten Versuchspersonen in ihrem Experiment aufgrund nur einer Dimension und zeigen somit keine „overall similarity“. Diese Bevorzugung aufgrund nur einer Dimension zu urteilen zeigte sich sowohl bei Variation der Stimuli (visuell, linguistisch) als auch bei expliziter Instruktion alle Eigenschaften bei der Klassifizierung zu benutzen. Erst durch Einführung einer Theorie, welche die einzelnen Merkmale von Objekten in einer Kategorie verband (z.B. Tiermerkmale, welche der Flugfähigkeit dienen), zeigten sich Klassifizierungen aufgrund der Ähnlichkeit über Dimensionen *hinweg* und die Klassifizierung aufgrund *einer* Dimension unterblieb

(siehe auch Kaplan & Murphy, 1999). Verschiedene Erklärungen der Diskrepanz zwischen natürlichen Kategoriestructuren und denen in Experimenten wurden angeboten:

Medin, Wattenmaker und Hampson (1987) berichten über Studien, welche eine ontologische Entwicklung von einer alle Dimensionen beachtenden Ähnlichkeitsstruktur zu einem eher analytischen Vorgehen nahe legen. Ähnlich berichten Goldstone und Barsalou (1998), dass ein „overall similarity“ Effekt bei Unerfahrenen bzw. unter Zeitdruck eine größere Rolle spielt.

Goldstone und Barsalou (1998) als auch Shiffrin und Schneider (1977) gehen davon aus, dass „overall similarity“ eine Form von assoziativen Prozessen darstellt und somit häufiger bei automatischen Prozessen eine Rolle spielen sollte. Bei kontrollierter Verarbeitung hingegen sollten Regelmechanismen bzw. Selektion einzelner Dimensionen in den Vordergrund treten.

Regehr und Brooks (1995) demonstrierten, dass die Entscheidung, aufgrund einer einzigen Dimension oder aufgrund einer Integration mehrerer Dimensionen zu klassifizieren, davon beeinflusst wird, ob alle Stimuli auf einmal, oder ob die Stimuli jeweils sequentiell dargeboten werden. Die Autoren erklärten den Befund damit, dass bei gleichzeitiger Darbietung der Stimuli leichter nach einer Regel gesucht werden kann. Allen und Brooks (1991) zeigten, dass aber selbst bei Anwendung einer Regel, die Ähnlichkeit zu bereits verarbeiteten Stimuli eine Rolle spielt. Der Konflikt, welcher bei widersprüchlichen Urteilen der Regel- und Ähnlichkeitsentscheidung auftritt, spiegelt sich in den Experimenten von Allen und Brooks (1991) in einer längeren Reaktionszeit und erhöhter Fehlerrate wider.

Shanks und Darby (1998) konnten zeigen, dass der Fortschritt im Lernprozess einen Einfluss darauf hat, nach welchen Gesetzmäßigkeiten neue Stimuli klassifiziert werden. Die Versuchspersonen konnten in einem medizinischen Szenario neue Stimuli entweder aufgrund ihrer Merkmalsähnlichkeit oder entsprechend einer Regel der Form „Zwei Elemente zusammen führen zu der gegensätzlichen Diagnose als die Elemente alleine“ klassifizieren. Shanks und Darby (1998) fanden einen Übergang von instanzbasierter Generalisierung auf Merkmalsebene zu regelbasierter Generalisierung, abhängig von dem erreichten Lernniveau bezüglich der präsentierten Stimuli.

Love (eingereicht) unterscheidet zwischen intentionalem und nicht-intentionalem unüberwachtem Klassifizieren. In seiner empirischen Studie zeigte sich, dass intentionales Lernen eher zum Erwerb einer Kategorienstruktur aufgrund einer Regel führt, während bei nicht-intentionalem, zufälligem Lernen, assoziative Prozesse überwiegen. Dies erscheint konsistent mit der obigen Feststellung, dass erst längere Verarbeitung bzw. mehr kognitive

Ressourcen zu einer Regelinduktion führen. Interessanterweise berichtet Love, dass Versuchspersonen in der intentionalen Kategorienlernumgebung bei späteren arithmetischen Aufgaben schlechter abschnitten als Versuchspersonen in der nicht-intentionalen Gruppe. Dieser Abfall der Leistung könnte dadurch erklärt werden, dass die intentionale Bedingung die kognitiven Ressourcen stärker erschöpfte. Allerdings unterschieden sich auch die Instruktionen und Anforderungen zwischen den beiden Gruppen. In der intentionalen Bedingung mussten die Versuchspersonen die Farbe des Rahmens der dargebotenen Stimuli vorhersagen, während in der nicht-intentionalen Bedingung lediglich beurteilt werden musste, wie ansprechend die Stimuli waren. Unter der Annahme, dass Versuchspersonendaten normalerweise unter der intentionalen Bedingung erhoben werden, erklärt sich die höhere Prävalenz von auf einzelnen Dimensionen basierenden Urteilen.

Ahn und Medin (1992) entwickelten ein 2 Phasen Modell, welches Klassifizierungen aufgrund „overall similarity“ produziert und damit konsistent mit dem Verhalten von Versuchspersonen ist. In der 1. Phase wird eine Dimension ausgesucht und aufgrund dieser 2 Kategorien gebildet. Stimuli, welche sich nicht aufgrund dieser Dimension klassifizieren lassen, werden in der 2. Phase aufgrund einer weiteren Ähnlichkeitsbegutachtung einer der beiden Kategorien zugeordnet. Bei für eine Klassifizierung suffizientem Merkmal führt dies zu einer „overall similarity“ Klassifizierungsstruktur. Unter „suffizientem“ Merkmal verstehen Ahn und Medin ein Merkmal, das in der ersten Phase erlaubt die meisten Stimuli so einzuteilen, dass die Einteilung der „overall similarity“ entspricht. Die restlichen Stimuli, welche nicht aufgrund des Merkmals eingeteilt werden können, da z.B. deren Ausprägung des Merkmals keiner Gruppe entspricht, werden dann in der 2. Phase entsprechend der „overall similarity“ zugeteilt. Eine gewisse Nähe zu einem Zirkelschluss dürfte aber bei dieser Begründung bestehen.

Inwieweit die obigen Modelle erklären können, warum natürliche Kategorien andere Ähnlichkeitsmaße zu besitzen scheinen als Versuchspersonen im Experiment zeigen, hängt im Wesentlichen von den Annahmen über den natürlichen Objektbereich und den dort vorherrschenden kognitiven Verarbeitungsstrategien (intentional vs. unintentional; unter Zeitdruck vs. nicht unter Zeitdruck etc.) ab. Besonders deutlich wird dies im 2 Phasen Modell von Ahn und Medin (1992). Die Erklärung von Ahn und Medin baut darauf auf, dass die Kategorien suffiziente Merkmale besitzen, aufgrund derer die Kategorien im Sinne der „overall similarity“ unterschieden werden können.

Generell dürfte die Benutzung von artifiziellen Stimuli aus zwei Gründen zu einer Überschätzung der Tendenz nur eine Dimension zu beachten führen: Erstens weisen

artifizielle Stimuli nur ein geringes Maß an Komplexität auf, erfordern also keine Einbeziehung vieler Dimensionen und zweitens variieren die Merkmale artifizierlicher Stimuli nicht stark über die Stimuli hinweg. Im Gegensatz dazu sind in der Natur Merkmale wie z.B. Eßwerkzeuge selten in gleicher Weise bei unterschiedlichen Trägern anzutreffen. Durch diese starke Variation in der Natur, im Gegensatz zu den Stimuli in Experimenten, wird erst eine abstrakte Ähnlichkeitsstruktur, die mehrere Merkmale berücksichtigt, erforderlich.

Glücklicherweise muss für die vorliegende Arbeit nicht zwangsläufig von einer der beiden Möglichkeiten ausgegangen werden. Wenn auch die Beachtung aller Dimensionen eher der Ursprungsintention des Rescorla-Wagner Modells entsprechen dürfte, so können doch durch eine größere Salienz auf nur einer Dimension auch regelbasierte Klassifizierungsstrategien erklärt werden. Diese Salienzzuweisung würde dann allerdings, falls sie nicht erst durch direkt vorhergehende Lernerfahrungen entstanden ist, auf außerhalb des Modells liegende Gründe zurückzuführen sein. Die generelle Hypothese, dass Klassifizierungslernen von dem Vorliegen von Diskrepanzen getrieben wird, bleibt dabei unangetastet. Das Blockingphänomen würde sich unter dieser Modellvorstellung von der Blockierung einfacher Stimulusinformationen auf die abstraktere Ebene der Blockierung der weiteren Suche nach Hypothesen über weitere Merkmale verlagern.

Wie Love (eingereicht) andeutet, mag es fruchtbar sein, Modelle für überwachtetes Kategorienlernen auf unüberwachte Paradigmen zu übertragen. Zwickel und Wills (2002) haben mit einem ähnlichen wie dem in dieser Diplomarbeit verwendeten Versuchsaufbau gezeigt, dass der Blocking-Effekt auch bei Klassifizierungsvorgängen mit minimaler Rückmeldung auftritt. Allerdings wurde den Versuchspersonen in der Studie von Zwickel und Wills nach jeweils 24 Stimuli mitgeteilt, wie viel Prozent der Stimuli sie richtig klassifiziert hatten. Somit handelte es sich um ein Experiment, welches den Blocking-Effekt nur für solche Situationen belegen konnte, in denen nur sehr wenige Rückmeldungen vorliegen. Der Informationsgehalt der Rückmeldung dürfte zwar ökologisch valide sein, da nach mehrmaligem Ausführen einer Handlung vermutlich immer eine gewisse Form der Rückmeldung im Sinne von besser oder schlechter auftritt, aber der Blocking-Effekt konnte somit nicht unter Fehlen von jeglichem Feedback belegt werden. Obwohl die Klasse der unüberwachten Netzwerkmodelle sensu Rumelhart und Zipser (1986) intuitiv eigentlich die geeignetere für Klassifizierungsprozesse ohne Rückmeldung zu sein scheint, soll in den folgenden Experimenten gezeigt werden, dass diese intuitive Annahme vermutlich falsch ist. Davor wird zuerst noch ein Modell vorgestellt, welches mittels einer modifizierten Delta-Regel Klassifizierung ohne Rückmeldung erklären könnte.

1.5 Die Delta Regel und Klassifikation ohne Rückmeldung

Es erscheint zunächst widersprüchlich ein Modell, welches sich dadurch auszeichnet, dass es in Abhängigkeit des rückgemeldeten Fehlers lernt, für Klassifizierung ohne Rückmeldung zu verwenden. Dies erscheint nur solange als Widerspruch, wie tatsächlich gefordert wird, dass die Rückmeldung von „Außen“ erfolgen muss. Wenn hingegen die Bedingung für Lernen davon abhängt, wie „sicher“ oder „konsistent“ eine vorliegende Kategorisierung erfolgt, so ist kein äußerer „Lehrer“ mehr nötig. Allerdings führt dies zu der Voraussage, die allgemein für Hebbsches, bzw. kompetitives Lernen sensu Rumelhart und Zipser (1986) gilt: Es führt nur dann zu Erfolg, wenn sich die richtige Kategorisierung aus der Anzahl gemeinsamer Merkmale ergibt. Hängt hingegen die Ähnlichkeitsstruktur von *einzelnen* Merkmalen ab, nicht aber von der *generellen* Unähnlichkeit, so hat ein Algorithmus ohne Rückmeldung von Außen keine Chance diese Struktur zu erlernen. Andere Ergebnisse würden aber auch überraschen, da sie bedeuten würden, dass ein Algorithmus angewendet auf eine Struktur ohne jegliche weitere Information (außer eventuell, Anzahl der zu bildenden Kategorien) die Struktur nicht nach Ähnlichkeit im Sinne übereinstimmender Merkmale aufteilen würde. Eine Aufteilung nach überlappender Merkmale dürfte aber, bei Fehlen von invalidierender Information, rational sein. Eine Einteilung aufgrund übereinstimmender Merkmale wäre deshalb rational, weil davon ausgegangen werden kann, dass ähnliche Objekte mit vielen gemeinsamen Merkmalen auch ähnliche Konsequenzen haben. Deshalb sollten Objekte mit vielen übereinstimmenden Merkmalen als Exemplare einer gemeinsamen Kategorie behandelt werden, bis weitere Information in Form von Rückmeldung durch die Umwelt auftritt. Daher erscheint es vernünftig, Elemente des kompetitiven Algorithmus auch in ein neues Modell für Klassifikationsprobleme ohne Rückmeldung aufzunehmen. Um das Lernen wie eingangs gefordert von der bereits erreichten „Sicherheit“ bzw. „Erfolg“ abhängig zu machen, muss aber noch eine Operationalisierung für „Sicherheit“ bzw. „Erfolg“ gefunden werden. Diese Operationalisierung sollte möglichst nur lokale Informationen verarbeiten, um biologisch plausibel zu erscheinen (vgl. aber auch Fußnote 14). Dies wird in dem vorgeschlagenen Algorithmus dadurch gelöst, dass nur dann gelernt wird, wenn die Präsentation eines Musters nicht zur maximalen Aktivierung eines Ausgabeknotens führt, man also sagen könnte, dass das Netzwerk noch eine „gewisse Unsicherheit verspürt“¹⁹. Die Veränderung des

¹⁹ Es könnte auch ein Intervall um den Zielzustand angegeben werden, in dem überhaupt nicht mehr gelernt wird. In der hier verwendeten Implementierung nimmt aber lediglich das Ausmaß an Gewichtsveränderungen automatisch mit der Annäherung an den Zielzustand ab. Da die Unterscheidung durch empirische Daten aufgrund der benötigten Genauigkeit der Messung schwierig sein dürfte, wurde die einfachere Implementation ohne zusätzliches Intervall, bevorzugt.

Gewichtsvektors erfolgt also ähnlich der Delta-Regel, wobei in dem gerade vorgeschlagenen Modell die Differenz zwischen maximaler und tatsächlicher Aktivierung das Ausmaß des Lernens bestimmt. Gleichzeitig wird der kompetitive Mechanismus dadurch aufgegriffen, dass einzelne Gewichtsvektoren um die Stimuli konkurrieren, um zu gewährleisten, dass unähnliche Muster im Sinne der Merkmalsüberlappung unterschiedlichen Kategorien zugeteilt werden. Formal ergibt sich damit für den gewinnenden Ausgabeknoten j der folgende Algorithmus:

$$\Delta w_{ij} = \frac{\beta(1-o_j)a_i}{n}; i \in \{\text{aktive Eingabeknoten}\} \quad \text{Gleichung 17}$$

und

$$\Delta w_{ij} = -\beta(1-o_j)\frac{n}{m}; i \in \{\text{inaktive Eingabeknoten}\} \quad \text{Gleichung 18}$$

Hierbei gibt „ n “ bzw. „ m “ die Anzahl aktiver bzw. inaktiver Eingabeknoten an.

Bei Annahme, dass die Eingabeknoten eine Aktivationsstärke von 0 oder 1 besitzen, reduziert sich dieser Algorithmus zu dem von Zwickel und Wills (2002) vorgeschlagenen. Das Ziel der Ad-hoc-Annahme des Bruches $\frac{n}{m}$ bzw. $\frac{1}{n}$ von Zwickel und Wills war, durch die stärkere Hemmung als Verstärkung der Gewichte eine schnelle Separierung der Muster zu ermöglichen. Dies war für eine Aufgabe mit zwei Kategorien ausreichend, da nur zwischen zwei Ausgabeknoten entschieden werden musste und damit die Hemmung von Merkmalen durch ein Ausgabeneuron automatisch zum „Gewinn“ des Eingabemusters durch das einzige alternative Ausgabeneuron führte. Für die in dieser Diplomarbeit vorliegende Aufgabe mit drei Kategorien muss hingegen zwischen drei Ausgabeknoten differenziert werden, was durch am Anfang des Lernens auftretende häufige Verwechslungen zu sehr hohen negativen Gewichten führt. Entscheidungen für oder gegen eine Kategorie werden damit nicht mehr durch positive Assoziationen, sondern durch negative Assoziationen bestimmt. Um dieses Aufschaukeln zu verhindern, muss die Inhibition inaktiver Eingabeknoten stärker abgeschwächt werden, aber gleichzeitig stark genug bleiben, um eine Separation der Muster zu ermöglichen. Bei einer zu kleinen Separationswirkung würde das erste gewinnende Ausgabeneuron alle anderen Stimuli aus anderen Kategorien auch gewinnen. Eine angemessene Inhibitionsrate könnte

- (a) durch Division der Gewichtsveränderung für nichtaktive Eingabeknoten durch eine Konstante,
- (b) durch Multiplikation der Gewichtsveränderung mit dem bereits bestehenden Gewicht, oder
- (c) durch eine generelle Begrenzung der negativen Gewichtsstärken erreicht werden.

Da alle Annahmen ad-hoc ohne ausreichende Datenbasis, aufgrund der eine Entscheidung zwischen den drei Varianten möglich wäre, getroffen werden mussten, wurden folgende zwei Kriterien zur Beurteilung herangezogen:

- a) Konnten mit den Veränderungen die Daten von Zwickel und Wills (2002) immer noch erklärt werden?
- b) Wurde ein weiterer freier Parameter eingeführt?

Alle 3 Verfahren konnten die Daten von Zwickel und Wills (2002) replizieren, wenn auch das zweite Verfahren nur in einem kleinen Parameterbereich. Damit konnte das erste Kriterium nicht zwischen den „Begrenzungs“-Varianten entscheiden. Da aber die Verwendung des ersten und dritten Verfahrens eine weitere willkürliche Setzung des Divisors bzw. der Schwelle bedeutet hätte, wurde die zweite Möglichkeit bei den Simulationen im folgenden Abschnitt verwirklicht. Allerdings werden für das erste Experiment auch Simulationsergebnisse für den ersten Algorithmus, bei dem die Begrenzung durch Division durch 100 erreicht wurde im Anhang unter „divisionsbegrenzter Algorithmus“ aufgeführt.

Der damit für die Studie gewählte, „gewichtsbegrenzte“ Algorithmus unterschied sich nur im Falle inaktiver Eingabeknoten von dem in Gleichung 17 und 18 aufgeführten. Für inaktive Eingabeknoten wurde das Gewicht zu dem Gewinnerausgabeknoten j nach folgender Regel verändert:

$$\Delta w_{ij} = -\beta(1 - o_j)w_{ij}; i \in \{ \text{inaktive Eingabeknoten} \} \quad \text{Gleichung 19}$$

Erneut spielt der genaue Algorithmus für diese Diplomarbeit keine Rolle, solange er die Idee implementiert, dass Lernen solange erfolgt, wie noch nicht die maximale Verbindungsstärke erlernt wurde. Die Vorstellung des internen Abgleichens zwischen einer Handlung und einem Erwartungswert auch ohne Rückmeldung wird durch Experimente von Scheffers und Coles (2000) bestärkt: Die Autoren zeigten, dass Versuchspersonen auch ohne Rückmeldung eine

fehlerbezogene Negativierung im EEG aufwiesen, wenn die Antwort bei einem späteren Rating subjektiv als falsch eingeschätzt wurde. Scheffers und Coles vermuten, dass im Wesentlichen zwei Faktoren an einer Diskrepanz zwischen einer Reaktion und der Erwartung an die Reaktion verantwortlich sind. Unter Zeitdruck bzw. einem subjektiven Zeitlimit kann eine Reaktion ausgelöst werden, welche nicht mit Ergebnissen einer weiteren Verarbeitung der Stimulusinformationen konsistent ist und deshalb als Fehlreaktion eingeschätzt wird. Eine andere Möglichkeit besteht darin, dass eine degradierte Darstellung der Stimuli zur Aktivierung mehrerer Reaktionsmöglichkeiten führt und deshalb ein partieller Mismatch zwischen Reaktion und Erwartung auftritt. Das von Scheffers und Coles verwendete Zeitlimit betrug 1300ms für die Unterscheidung zwischen zwei Buchstaben, die zusammen mit kompatiblen bzw. inkompatiblen Distraktoren präsentiert wurden. Die einfache Übertragbarkeit der Ergebnisse wird allerdings dadurch erschwert, dass im Paradigma von Scheffers und Coles (2000) die Versuchspersonen sehr klar erkennen konnten - jedenfalls bei ausreichender Zeit für die kategoriale Antwort - wie sie sich entscheiden sollten. Dies dürfte aber auf die später geschilderten Experimenten nicht zutreffen (vgl. die Schwierigkeiten auch ohne Zeitdruck in Experiment 2). Trotzdem zeigen die Studien, dass ein gewisser Abgleich zwischen einer Reaktion und einer wie auch immer generierten Erwartungshaltung plausibel ist.

Mit dem vorgeschlagenen Modell ist es möglich den Kategorienerwerb sowohl mit als auch ohne Rückmeldung zu modellieren. Es erlaubt die Entscheidung, ob Rückmeldungen aus der Umwelt integriert werden auf eklektische Weise. Im Falle *vorhandener* Rückmeldungen werden diese integriert. Im Falle *fehlender* Rückmeldungen wird die Konsistenz durch „intern“ vorliegende Informationen erhöht, indem die Entscheidung mit der höchsten Aktivierung, die also in der Vergangenheit bereits gewählt wurde, weiter verstärkt wird.

2 Experimente

Wie durch die obige Diskussion klar geworden sein sollte, stellen Modelle, welche nur solange „lernen“ wie eine Differenz zwischen der bisherigen Vorhersage und einer Rückmeldung existiert, plausible Erklärungen für Kategorisierungsvorgänge bei Menschen dar. Weiterhin sollte es nun plausibel erscheinen, ähnliche, von Fehlern „getriebene“ Modelle auch für Klassifizierungsaufgaben ohne Rückmeldung zu untersuchen, um eine Vereinheitlichung der Modellvorstellungen zu ermöglichen. Damit können dann Vorstellungen, welche verschiedene Modelle für Klassifizierungs- und Kategorisierungslernen fordern, mit Occam's Razor²⁰ beseitigt werden. Dieses bis jetzt nur mittels Plausibilitätsargumenten begründete Ziel soll nun durch empirische Daten erhärtet werden. Dazu wurden Versuchspersonen instruiert, am Bildschirm dargebotenen Stimuli entsprechend ihrer Ähnlichkeit in verschiedene Kategorien einzuteilen. Sollte sich in diesen Einteilungen ein Blocking-Effekt finden, so wäre dies ein starker Hinweis auf fehlergetriebenes Lernen auch bei Klassifizierungsaufgaben ohne Rückmeldung. Die übergeordneten Hypothesen lauten somit:

H₀: Kategorisierung mit Rückmeldung und Klassifizierung ohne Rückmeldung erfordern prinzipiell unterschiedliche Modelle. Während Kategorisierungsmodelle abhängig von der Differenz zwischen einer Rückmeldung und einer Performanz lernen, laufen Klassifizierungsvorgänge unabhängig von der gezeigten Performanz ab.

H₁: Kategorisierungs- und Klassifizierungsphänomene können durch sehr ähnliche Modelle erklärt werden. Beide Modelle beinhalten einen Vergleich zwischen gezeigter Performanz und einem Erwartungswert.

Natürlich folgt H₁ nicht logisch aus der Ablehnung von H₀, da H₁ noch weitere Annahmen, nämlich über die generelle Ähnlichkeit enthält. Allerdings erhöht eine Ablehnung von H₀ die Plausibilität für H₁. H₁ wurde aus Gründen der Anschaulichkeit so stark formuliert.

Als testbare, operationale Hypothesen ergeben sich daraus:

O₀: Ein Blocking-Effekt als typisches Phänomen von differenzgetriebenen Lernmodellen tritt nicht bei Klassifizierungsaufgaben auf.

O₁: Auch bei Klassifizierungsaufgaben tritt der Blocking-Effekt auf.

²⁰ vgl. Fußnote 1 von Chater (1999) zur Wandlung des Verständnisses des Einfachheitskriteriums von der Anzahl benötigter Objekte hin zur allgemeinen Einfachheit der Erklärung.

Im 1. Experiment sollten Stimuli in eine von drei verschiedenen Kategorien eingeordnet werden. Wie Homa und Cultice (1984) zeigten, ist für eine erfolgreiche Klassifizierung ohne Rückmeldung eine hohe Strukturierung der Stimuli notwendig ist. Da Unsicherheit darüber bestand, ob die Stimuluskomplexität bei drei verschiedenen Kategorien die Versuchspersonen nicht überfordert, wurde eine Pilotstudie mit sechs Versuchspersonen durchgeführt. Das Design der Pilotstudie wurde im Wesentlichen für die Experimente übernommen. Aus diesem Grund erfolgt die Darstellung der Pilotstudie sehr detailliert.

2.1 Pilotstudie

Das Experiment gliederte sich in 1²¹ Aufwärm-, 2 Lern- und 1 Testphase. In der 1. Lernphase wurden Stimuli präsentiert, welche sich auf der Basis gemeinsamer Merkmale in 3 verschiedene Kategorien einteilen ließen. In der zweiten Lernphase wurden Stimuli präsentiert, die in unterschiedlichem Grade aus Elementen der in der ersten Lernphase gesehenen Stimuli und neuen Elementen bestanden. Das Vorhandensein von Elementen aus der ersten Lernphase in einem Stimulus sollte damit als Klassifizierungsgrundlage ausreichen und neue, ebenfalls in dem Stimulus vorhandene Elemente, „blocken“. Abgeschlossen wurde das Experiment durch eine Testphase. In der abschließenden Testphase wurden 2 Stimuli gezeigt, welche zu gleichen Anteilen aus Elementen der 3 in Phase 2 gezeigten Kategorien generiert wurden. Der 1. Teststimulus beinhaltete sowohl „geblockte“ als auch „ungeblockte“ Informationen. Der 2. Teststimulus beinhaltete Informationen, die sich lediglich hinsichtlich ihrer Darbietungshäufigkeit unterschieden. Da die reine Anzahl an gemeinsamen Elementen mit den Prototypen der vorhergehenden Lernphasen keine Auskunft über die Zuordnung zu einer Kategorie gab, ergeben sich die Experimentalhypothesen zu:

E_0 : Für *Teststimulus 1*, ist über die Versuchspersonen hinweg, eine Gleichverteilung über alle 3 Kategorien zu erwarten. Es besteht keine Bevorzugung der „ungeblockten“ Kategorie.

E_1 : *Teststimulus 1* wird nicht allen 3 Kategorien mit der selben Häufigkeit zugeordnet. Am häufigsten wird Teststimulus 1 der Kategorie zugeordnet, der die „ungeblockten“ Elemente entstammen.

²¹ Um die Lesbarkeit zu erhöhen, wurde für Zahlenangaben, die sich auf das Design beziehen, die numerische Schreibweise gewählt.

Während die 3. Kategorie in der 2. Lernphase 4 Elemente aus der ersten Lernphase beinhaltete, besaß die 1. Kategorie 8 Elemente aus der ersten Phase. Die 2. Kategorie bestand aus keinen Elementen der 1. Phase. Strikte Anwendung des Delta-Modells führt zur Vorhersage, dass Teststimulus 1 am häufigsten in die 2. Kategorie, und am seltensten in die 1. Kategorie eingeordnet wird. Es wurde erwartet, dass die 3. Kategorie mit mittlerer Häufigkeit gewählt werden sollte. Diese Vorhersagen folgen aus der Eigenschaft von Delta-Modellen, da in diesen Modellen bei größerem Fehler stärker gelernt wird. Für Stimuli aus Kategorie 2 sollte somit in der 2. Lernphase aufgrund der größten Anzahl an neuen Elementen der größte Fehler auftreten. Es wurde erwartet, dass dieser größere Fehler in Kategorie 2 als in den anderen Kategorien, zu stärkeren Verbindungen zwischen den Elementen der Kategorie und der Kategorienrepräsentation als bei den anderen Kategorien führt. Die Assoziation der Elemente zu Kategorie 1 sollte in der 2. Phase, aufgrund des kleinen Fehler, relativ gering sein. Die Elemente des 1. Teststimulus, welche aus der 3. Kategorie der 2. Phase stammten sollten eine mittlere Verknüpfungsstärke zu der 3. Kategorie besitzen, da in Phase 2 für die 3. Kategorie ein mittlerer Fehler, im Sinne noch nicht vorhergesagter Elemente entstanden sein sollte.

Teststimulus 2 sollte allen 3 Kategorien mit gleicher Wahrscheinlichkeit zugeordnet werden, da von einem asymptotischen Lernen der Verbindungen ausgegangen wurde. Am Anschluss an Teststimulus 2 wurden einzelne Stimuluselemente präsentiert und ihre Bekanntheit von den Versuchspersonen eingeschätzt. Es wurde erwartet, dass „ungeblockte“ Elemente als bekannter eingeschätzt werden als „geblockte“ Elemente.

2.1.1 Methode

Eine anfallende Stichprobe von 6 normalsichtigen oder korrigierten Teilnehmern, vorwiegend Psychologiestudierende beider Geschlechter, wurden zufällig einer von 3 Bedingungen zugeteilt. Ihre Aufgabe bestand darin, unterschiedliche Stimuli in 3 exklusive und exhaustive Gruppen durch Betätigen einer von 3 Tasten einzuteilen. Die Versuchspersonen erhielten keine Rückmeldung über ihre Performanz. Das Experiment wurde an einem Tag vom selben männlichen Versuchsleiter in einem abgetrennten und ruhigen Raum durchgeführt.

Materialien

Die Stimuli wurden an vier Standard PCs dargeboten. Die Zuordnung der Versuchspersonen zu einem der Rechner erfolgte zufällig. Die Versuchspersonen konnten sich den Abstand zum Monitor selbst einstellen. Die Stimuli wurden den 3 Kategorien durch die Zifferntasten „1“, „2“ und „3“ auf dem Ziffernblock zugewiesen.

Instruktionstext

Der Instruktionstext wurde ebenfalls mittels PC-Bildschirme dargeboten. Die Versuchspersonen wurden aufgefordert, selbständig 3 Kategorien zu bilden und sich nicht von der anfänglichen Komplexität entmutigen zu lassen. Für den genauen Wortlaut, siehe Anhang 1.

Stimuli

Alle Stimuli setzten sich aus 12 Icons aus einem Set von 72 farbigen Icons (Wills & McLaren, 1997) zusammen. In Anhang 2 ist das vollständige Set aufgeführt. Die Icons wurden in einem Rechteck, bestehend aus 4 Icons in der Breite und 3 in der Höhe, angeordnet. Ein grauer Rahmen umgab die Stimuli, welche im Zentrum des Monitors die Fläche eines Rechtecks mit ca. 5 cm Breite und ca. 4 cm Höhe einnahmen. Die Stimuli wurden mittels der DMDX Software (Forster & Forster, 2002, Version 2.9) dargeboten.

Prototypen

Für jeden Prototyp wurden 12 Icons zufällig und exklusiv ausgewählt.

Exemplar-Stimuli

Von jedem der 3 Prototypen wurden 7 Stimuli generiert, indem die Icons zufällig ihre Plätze im Rechteck permutierten. Zusätzlich wurde jedes Icon zufällig mit einer Wahrscheinlichkeit von 10 Prozent mit einem ebenfalls zufällig bestimmten Icon eines anderen Prototyps aus der gleichen Phase ausgetauscht. Der Erwartungswert an ausgetauschten Icons ergab sich somit zu 1.2 Icons pro Stimulus. Abbildung 2.1 zeigt 2 Prototypen und 2 aus den Prototypen generierte Stimuli.

Maske

Eine graue Maske verdeckte nach jeweils 1800 ms den Stimulus vollständig. Zweck der Maske war, die Stimuli aus dem ikonischen Gedächtnis zu löschen (Turvey, 1973). Die Maske sollte weitestgehend eine Interferenz zwischen aufeinanderfolgenden Stimuli verhindern und eine Regelgenerierung durch bewusstes „Durchscannen“ des Stimulus erschweren. Durch die Zeitbeschränkung sollte ebenfalls eine eventuelle Regelgenerierung erschwert werden. Mit diesen Maßnahmen sollte ein „natürlicher“ Anwendungsbereich von konnektionistischen Modellen geschaffen werden.

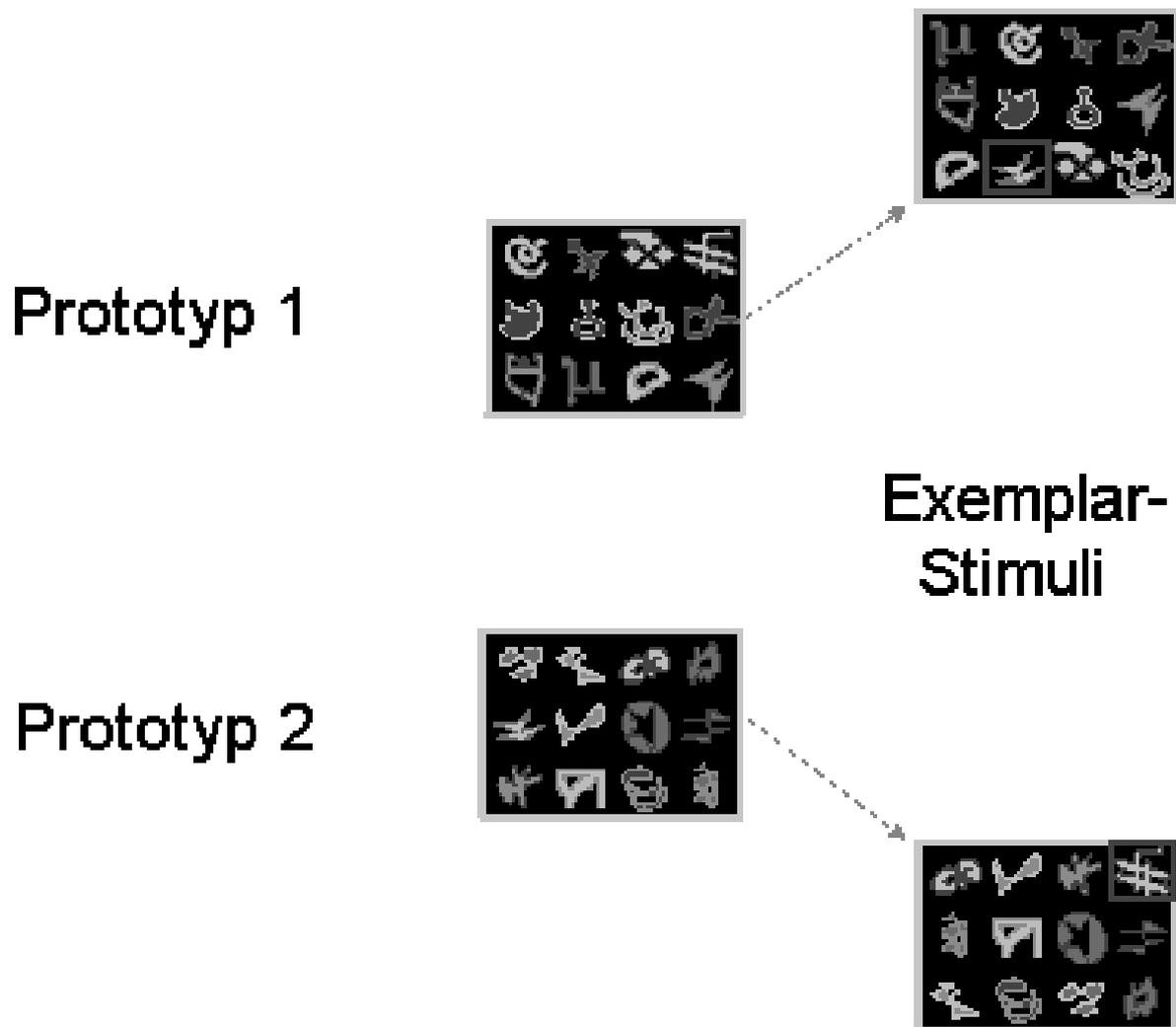


Abb. 2.1: 2 Prototypen und 2 aus den Prototypen generierte Stimuli. Die roten Rechtecke umranden 2 zufällig ausgetauschte Icons.

Design

Im Folgenden werden die einzelnen Kategorien entsprechend Abbildung 2.2 mit „1.“, „2.“ und „3.“ Kategorie bezeichnet, um die Zusammensetzung der Prototypen zu beschreiben. Diese Bezeichnung hatte für die Versuchspersonen keine Bedeutung. Um das Design vorzustellen werden jeweils 6 Icons durch einen Buchstaben referenziert. Die Zuordnung der Icons zu den Buchstaben erfolgte für $\frac{1}{3}$ der Versuchspersonen zufällig (Gruppe 1). Die anderen $\frac{2}{3}$ erhielten „gematchte“ Zuordnungen (Gruppe 2 und 3). Die Zuordnungen für „gematchte“ Versuchspersonen erfolgte so, dass „Buchstaben“, welche bei der Referenzversuchsperson geblockt werden sollten, bei der entsprechend „gematchten“ Person ungeblockte Icons repräsentierten. Die Zuordnung der Icons zu geblockten Kategorie erfolgte also ausbalanciert. Für jeweils 3 „gematchte“ Versuchspersonen bedeuten damit gleiche Buchstaben auch gleiche Icons. Diese Äquivalenz gilt nicht über „ungematchte“

Versuchspersonen hinweg. Für Versuchspersonen in der 2. Phase wurden die Stimuli so generiert, dass für Exemplare aus der ersten Kategorie, $\frac{2}{3}$ der Icons bereits in der 1. Phase präsentiert worden waren. Für Stimuli aus der 2. Kategorie waren alle Icons neu, in der 3. Kategorie war nur $\frac{1}{3}$ der Icons bekannt. In der Testphase schließlich wurde den Versuchspersonen ein Stimulus präsentiert, welcher zu gleichen Anteilen aus Icons der 3 Kategorien zusammengesetzt war. Allerdings waren die Anteile aus den einzelnen Kategorien in einem unterschiedlichen Maße „geblockt“. Daran anschließend wurde erneut ein Stimulus zu gleichen Teilen aus den 3 Kategorien gebildet. Dieses Mal sollte aber keines der Icons geblockt sein. Vielmehr bestand der einzige Unterschied zwischen den Icons in der Häufigkeit, mit der sie über das ganze Experiment dargeboten worden waren. $\frac{2}{3}$ der Icons waren bereits in Phase 1 präsentiert worden. Die 4 Icons jedes Buchstabens des ersten Teststimulus wurden anschließend in über die Versuchspersonen hinweg ausbalancierter Reihenfolge dargeboten und die Versuchspersonen befragt, wie bekannt ihnen die Icons vorkamen. Abbildung 2.2 zeigt das Design.

		<u>Phase 1</u>	<u>Phase 2</u>	<u>Test 1</u>	<u>Test 2</u>	<u>Bekanntheit</u>	
Gruppe 1	K1	ABC	ABE				
	K2	DFG	HIJ	EJO	AHK	(E,J,O)	
	K3	KLM	KNO				
Gruppe 2	K1	ABC	ABJ				
	K2	DFG	HIE	E O	AHK	(E,I,O)	
	K3	KLM	KNO				
Gruppe 3	K1	ABC	ABE				
	K2	DFG	HIO	EJO	AHK	(E,J,O)	
	K3	KLM	KNJ				
		Zeit					
		→					

Abb. 2.2: Design der Pilotstudie: Die Prototypen der 3 Kategorien (K1-K3) sind in den ersten 2 Phasen für jede Gruppe untereinanderstehend aufgeführt. Die Versuchspersonen wurden zufällig in eine von 3 Gruppen eingeteilt. Die Gruppen unterschieden sich nur darin, welche 4 Icons, repräsentiert durch eine Buchstaben, „geblockt“ wurden. Geblockte Icons, welche durch 8 bekannte Icons geblockt wurden sind hell markiert. Geblockte Icons, die zusammen mit 4 aus der vorhergehenden Phase bekannte Icons auftraten, sind mittelhell gekennzeichnet. Zum Beispiel erhielt Gruppe 1 in Phase 2 aus der Kategorie 1 die Icons der „Buchstaben“ „A“, „B“ und „E“, sowie aus der Kategorie 3 „K“, „N“ und „O“. Im ersten Fall waren die Icons der „Buchstaben“ „A“ und „B“ bereits bekannt, im zweiten Fall nur die von „K“. Die Klammern in der Spalte „Bekanntheit“ bedeuten eine ausbalancierte Reihenfolge der Präsentation der einzelnen „Buchstaben“.

Datengewinnung

Die Daten wurden mittels Standard PC-Tastaturen erhoben. Die Teilnehmer der Studie konnten ihre Entscheidung, einen Stimulus in eine von 3 Kategorien einzuordnen, durch Betätigen der nebeneinanderliegenden Tasten „1“, „2“ und „3“ auf dem Ziffernblock anzeigen. Die Versuchspersonen wurden instruiert, nur einen Finger für die Antwortwahl zu benutzen, um eventuell auftretende Gruppierungsprozesse zu verhindern. Die Einschätzung der Bekanntheit der einzelnen Stimuli erfolgte ebenfalls durch den Ziffernblock. Für das Rating standen 6 verschiedene Ausprägungen zur Verfügung. „1“ wurde dabei mit „überhaupt nicht bekannt“ und „6“ mit „absolut bekannt“ bezeichnet. Unter der Präsentation der Icons war die Zuordnung nochmals aufgeführt. Das Rating wurde mit folgenden Worten eingeführt: „Wenn du jetzt die Leertaste drueckst werden Dir 4 Icons gleichzeitig gezeigt. Versuche einen allgemeinen Eindruck von dem Bekanntheitsgrad zu erhalten. Waehle dann eine von 6 Kategorien mit den Zifferntasten 1-6 aus. 1 steht fuer ueberhaupt nicht bekannt, 6 fuer absolut bekannt. 2 fuer vage bekannt, 3 fuer kaum bekannt, 4 fuer etwas bekannt, 5 fuer gut bekannt.“

Durchführung

Die Versuchspersonen wurden zufällig einer der 3 Gruppen hinsichtlich „gematchter“ Icons unter der Randbedingung zugeteilt, dass jede der Bedingungen bei Abschluss des Experimentes gleichhäufig präsentiert worden sein sollte. Das Experiment gliederte sich in 1 Übungs-, 2 Experimental- und 1 Testphase. In der Übungsphase wurden zuerst 2 Stimuli aus derselben Kategorie dargeboten und die Versuchsperson, nachdem sie beide Stimuli eingeordnet hatte, darauf hingewiesen, dass sie für beide Stimuli die gleiche Taste hätte benutzen sollen. Danach wurde den Probanden mitgeteilt, dass die nächsten 3 Stimuli aus unterschiedlichen Kategorien entstammen und sie deshalb 3 verschiedene Tasten benutzen sollte. Die Stimuli der Übungsphase wurden aus den ersten 12 Icons der im Anhang 2 abgebildeten Icons generiert. Diese Icons wurden im Experiment nicht mehr verwendet.

Jede Experimentalphase bestand maximal, falls die Stimulusstrukturen nicht erlernt wurden, aus 9 Lernblöcken. Die Versuchspersonen erhielten keinen Hinweis auf die Blockstruktur. Während jedes Blockes wurden 24 Stimuli (3 Prototypen und 21 Exemplare) in zufälliger Reihenfolge präsentiert und sollten von den Probanden klassifiziert werden. Die Versuchspersonen mussten dabei selbst die Kategorien entwickeln und erhielten keine Hilfe. Die Teilnehmer blieben solange in einer Phase, bis sie entweder während eines Blockdurchganges mindestens 6 richtige Exemplare aus jeder Kategorie konsistent einer Taste zugeordnet hatten, oder die maximale Anzahl an Blöcken dargeboten worden war. Diese letzte Erleichterung wurde gewählt, um Frustration zu vermeiden. Das Lernkriterium

wurde also auf 75% konsistenter Klassifizierungen festgelegt. Unter der Annahme, dass die Versuchspersonen die Stimuli zufällig beantworteten, d.h. die Zuordnung unabhängig vom präsentierten Reiz erfolgte, aber jeder Kategorie gleich viele Stimuli zugeteilt wurden, beträgt die Wahrscheinlichkeit in einer der beiden Phasen das Kriterium zu erreichen 0.4% (für Herleitung siehe Anhang 3). Die ersten 3 Stimuli in beiden Phasen waren jeweils die Prototypen, um den Versuchspersonen einen Eindruck von der Unterschiedlichkeit der Stimuli zu vermitteln. Die Versuchspersonen wurden nicht darauf hingewiesen. Als erster Stimulus in Phase 2 wurde der Prototyp aus Kategorie 1 präsentiert. Auch dies wurde den Versuchspersonen nicht mitgeteilt. Die Versuchspersonen wurden weder auf einen Wechsel zwischen den Blöcken noch zwischen den Phasen aufmerksam gemacht. Wenn nach 3000 ms noch keine Reaktion erfolgt war, wurde der Text „Du hast zu langsam reagiert. Versuche bitte etwas schneller zu antworten!“ eingeblendet und mit dem nächsten Item fortgefahren. Nach Abschluss der Phasen 1 und 2 wurden die beiden Teststimuli präsentiert. Für diese Teststimuli bestand kein Zeitlimit, um eine Entscheidung zu erzwingen. Anschließend sollten die Teilnehmer mit den Tasten „1“ bis „6“ angeben, wie bekannt ihnen die 4 Icons jeweils eines Buchstabens vorkamen. Die Reihenfolge der Buchstabenpräsentation war über die Versuchspersonen und innerhalb der Gruppen ausbalanciert. Nach Abschluss des Experimentes wurden die Versuchspersonen gebeten informell kurz schriftlich festzuhalten, nach welchen Gesichtspunkten die Klassifizierung erfolgte. Sie wurden aufgefordert, sich dabei weder um grammatikalische noch orthografische Regeln zu kümmern. Die Aufforderung wurde informell gehalten, um auf eventuelle Nachfragen der Versuchspersonen reagieren zu können. Wichtig war bei dieser Erhebung nicht so sehr die Objektivität der Antwort in dem Sinne, dass bei unterschiedlichen Versuchsleitern dieselbe Antwort produziert werden würde, sondern dass die Versuchsperson ihren eigenen Eindruck möglichst offen schildert. Das Experiment dauerte zwischen 20 und 35 Minuten. Die Versuchspersonen wurden entweder mit Schokoriegeln oder Versuchspersonenstunden entlohnt.

2.1.2 Ergebnisse und Diskussion

Nur 2 der 6 Versuchspersonen erreichten in den ersten 2 Phasen das Kriterium mindestens 6 Stimuli aus jeder Kategorie mit einer exklusiven Taste zu beantworten. Dieser niedrige Erfolgsrate würde dann keine Rolle spielen, wenn der Zweck der Studie darin bestünde, eine realistische und korrekte Beschreibung der Versuchspersonenleistung zu ermöglichen. Da diese Studie aber eine konkrete Annahme über Blocking untersuchen sollte und dieser Effekt nur dann auftreten kann, wenn die einzelnen Elemente in der Trainingsphase überhaupt erlernt wurden, müssten 4 der 6 Versuchspersonen von der Analyse ausgeschlossen werden. Es

besteht natürlich die Möglichkeit, dass diese 4 Versuchspersonen doch Wissen über die Kategorienstruktur erworben, aber eine andere Einteilung vorgenommen haben als es die Anzahl gemeinsamer Elemente nahe legen würde. Eine mögliche Zuteilung wäre z.B. bei Auftreten eines blauen Icons in der linken oberen Ecke eine bestimmte Taste zu betätigen. Auch bei dieser Klassifizierungsstrategien wäre Blocking zu erwarten. Bei dieser Strategie würde sich allerdings der Blocking-Effekt nicht auf einzelne Icons sondern auf Regeln beziehen. Eine erworbene Regel würde, solange sie für die Klassifizierung ausreicht, verhindern, dass andere Regeln erworben werden oder die Regel würde nur auf Teilbereiche des Stimulus angewendet werden (vgl. Diskussion in Allen & Brooks, 1991, p. 17). Allerdings lässt sich dieser „Hypothesen-Blocking-Effekt“ nicht mit dem hier verwendeten Paradigma testen, da für das vorliegende Design davon ausgegangen werden muss, dass die Versuchspersonen die erwarteten Kategorien erlernt haben. Andernfalls würde in der 3. Phase kein Konflikt mit gleichen „objektiven“ Anteilen aus allen Kategorien auftreten.

Ein Paradigma ohne Rückmeldung ist zwangsläufig dem Problem ausgesetzt, dass erwartete Strukturen eventuell nicht erlernt werden. Für die vorliegende Fragestellung muss dies aber gewährleistet sein. Aus diesem Grunde wurde versucht die Aufgabe zu erleichtern. Eine Möglichkeit bestand darin die Anzahl der Kategorien zu reduzieren. Während bei 3 Kategorien die Ähnlichkeit zu den einzelnen Kategorien verglichen werden muss, reicht bei 2 Kategorien die Einschätzung, dass ein gerade präsentierter Stimulus einer gewissen Kategorienvorstellung nicht entspricht aus, um die Kategorienstruktur zu erwerben. Da aber durch die Verwendung dreier Kategorien, das Verhalten der Versuchspersonen nicht so stark beschränkt würde, könnte ein genaueres Bild der Klassifizierungsvorgänge erhalten werden. Deshalb wurde die Möglichkeit der Kategorienreduzierung fallengelassen. Eine andere Möglichkeit bestand darin, den Versuchspersonen mehr Zeit für die Kategorisierung zu gewähren. Da aber vermutet wurde, dass dies zu stärkerer Regelinduktion führen würde (vgl. Abschnitt 1.4) und dies gewinnbringender in einem zweiten Experiment erforscht werden sollte, wurde auch diese Möglichkeit fallengelassen. Die 3. Möglichkeit bestand darin, die Stimulusähnlichkeit zwischen den Stimuli einer Kategorie weiter zu erhöhen. Diese Möglichkeit wurde als am wenigsten verfälschend angesehen. Daher wurde, um die Stimulusähnlichkeit zu erhöhen, der Platz der Icons innerhalb der Phasen nicht mehr permutiert, sondern die einzelnen Icons erschienen jeweils an derselben Stelle im Rechteck. Dies sollte eine Regelinduktion nur leicht verstärken (vgl. Regehr & Brooks, 1995). Da die erfolgreichen Versuchspersonen weniger als 7 Blöcke in beiden Phasen benötigten, wurde die maximale Anzahl an Blöcken in Experiment 1 auf 6 Blöcke festgelegt.

2.2 Experiment 1

Das Design von Experiment 1 war im Wesentlichen identisch mit dem der Pilotstudie. Lediglich wurde den Icons innerhalb jeder Phase eine fixe Position zugewiesen und die maximale Blockanzahl auf 6 Blöcke reduziert.

2.2.1 Methode

36 weibliche und männliche Teilnehmer wurden an 5 aufeinanderfolgenden Werktagen im selben Raum vom Versuchsleiter der Pilotstudie erhoben.

Stimuli

Die Stimuli wurden nach der gleichen Methode und aus denselben Icons der Pilotstudie konstruiert. Allerdings wurden die Plätze der Icons *innerhalb* einer Phase nicht mehr permutiert.

Design

Die Phasen bestanden maximal aus 6 Blöcken. Nach 24 Versuchspersonen wurden für jede neue Versuchsperson von den DMDX-Präsentationsfiles, die von den Versuchspersonen bisher nicht gelöst wurden, zufällig ein Präsentationsfile ausgewählt (im Folgenden durch "Zweite Erhebungsstufe" bezeichnet). Es wurde darauf geachtet, dass jede Buchstaben-Icon-Zuordnung, wie sie in den DMDX-Files festgelegt war, höchstens einmal wiederholt wurde. Aus Versehen wurde aber eine Präsentation wiederholt, obwohl sie bereits im ersten Durchgang gelöst worden war (Vpn 6 und Vpn 7). Der übrige Ablauf entsprach dem der Pilotstudie.

2.2.2 Ergebnisse

In Anlehnung an Hays (1994, p. 328; siehe auch Bortz, 1979, S. 166) wurde bei Vorliegen gleicher Stichprobengrößen die Varianzhomogenitäts- und Normalverteilungsannahmen für t-Tests nicht überprüft. Berichtete p-Werte beziehen sich, mit Ausnahme von Binomial und Fisher-Exact-Tests, auf 2-seitige Tests. Marginal signifikante Ergebnisse mit p-Werten zwischen 0.05 und 0.10 werden mit „ $p < 0.10$ “ berichtet.

15 der 36 Versuchspersonen schafften in beiden Trainingsphasen 75% der Stimuli entsprechend der Anzahl gemeinsamer Elemente zu klassifizieren und erreichten damit das Lernkriterium. Aus den in der vorhergehenden Diskussion vorgebrachten Gründen wurden für die hier betrachtete Fragestellung alle Versuchspersonen ausgeschlossen, welche in einer der beiden Phasen das Kriterium nicht erreicht hatten. Dies stellt ein nicht unübliches Verfahren

bei Blockingexperimenten mit Rückmeldung dar (vgl. z.B. Kruschke, eingereicht). Versuchspersonen, welche in mindestens einer Phase das Kriterium nicht erreichten, wurden als „Nicht-Löser“ analysiert. Ihre Daten werden, wenn nötig, im Text erwähnt und sind vollständigshalber in Anhang 4 für Experiment 1 und Anhang 5 und 6 für die beiden anderen Experimente aufgeführt. Das Verhältnis erfolgreiche Löser zu Nicht-Löser der 1. Erhebungsstufe entspricht dem Verhältnis bei den 12 Versuchspersonen in der 2. Erhebungsstufe (10:14 bzw. 5:7).

Anzahl benötigter Blöcke

Nur einer der „erfolgreichen“ Teilnehmer benötigte mehr als 5 Blöcke in einer der 2 Phasen. Die Anzahl benötigter Blöcke in den 2 Phasen ist in Abbildung 2.3 dargestellt. Es findet sich kein signifikanter Unterschied zwischen den zwei Phasen ($t(14)=-0.63$, $p>0.10$) die Korrelation zwischen der Anzahl benötigter Blöcke in Phase 1 und 2 ist minimal (Pearson $r=-0.05$, $p>0.10$).

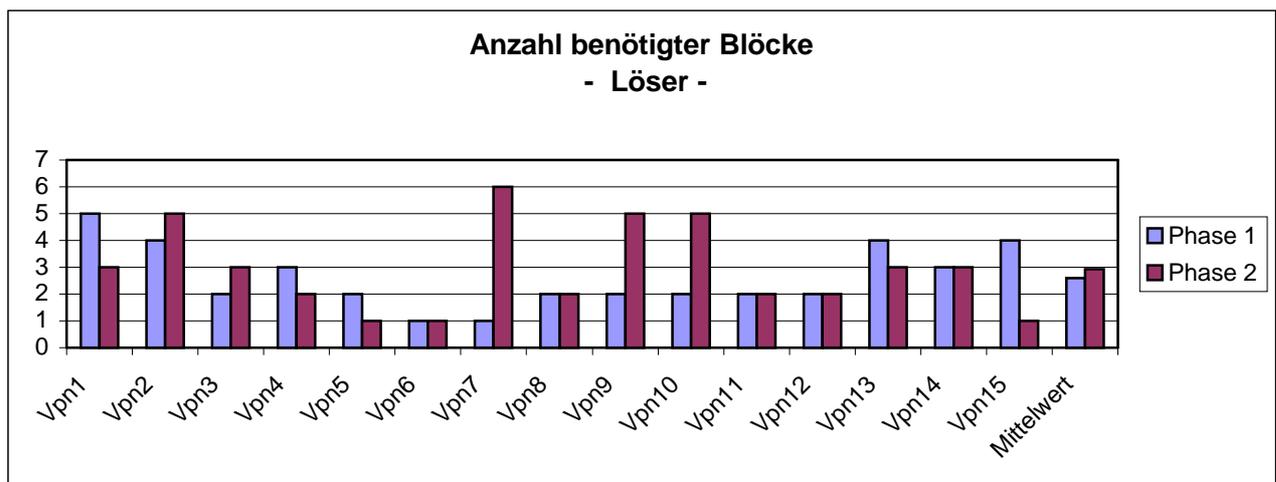


Abb. 2.3: Anzahl benötigter Blöcke der Löser in Phase 1 und Phase 2.

Reaktionszeit

Items auf die nicht in der vorgegebenen Zeit reagiert wurde, wurden aus den Mittelwertberechnungen für die Reaktionszeit ausgeschlossen. Dies geschah, da unklar war, welche Reaktionszeit unbeantworteten Items zugeordnet werden sollte. Außerdem wurde davon ausgegangen, dass es sich bei nicht beantworteten Items um Ausreißer und damit eine schlechte Informationsquelle für die *mittlere Reaktionsgeschwindigkeit* einer Versuchsperson handelte. Die „Ausreißer“ sind vermutlich auf nicht im Experiment begründete Störeinflüsse zurückzuführen. Der Ausschluss veränderte die Mittelwerte nur unwesentlich, da nur bei einer Person in beiden Phasen zusammen mehr als 3 Items ausgeschlossen werden mussten.

Insgesamt wurden über alle Versuchspersonen und Phasen hinweg 16 Items von der Mittelwertsbildung ausgeschlossen.

Die erfolgreichen Versuchspersonen blieben im Mittelwert mit ihrer Reaktionszeit weit unter dem Zeitlimit von 3000ms. Ein t-Test zwischen den individuell mittleren Reaktionszeiten in allen Blöcken und Experimentalphasen und dem Zeitlimit von 3000ms war hoch signifikant ($t(83)=0.00$, $p<0.001$). Der Mittelwert der Reaktionszeiten der Versuchspersonen im 1. und letzten Block einer Phase unterschied sich sowohl in der 1. als auch 2. Phase signifikant ($t(14)=3.60$, $p<0.01$; $t(14)=2.91$, $p<0.05$)²². Kein signifikanter Unterschied ist hingegen zwischen dem letzten Block der ersten Phase und dem ersten Block der zweiten Phase zu belegen ($t(14)=-1.43$, $p>0.10$). In Abbildung 2.4 ist in schwarz der Mittelwert der mittleren Reaktionszeiten der Versuchspersonen aufgetragen. Für Teilnehmer, die bereits vor dem 6. Block das Kriterium erreichten, wurde der zuletzt erreichte Wert im *Gesamtmittelwert* aller Versuchspersonen, aber nicht als *Versuchspersonenmittelwert* im Diagramm weitergeführt²³. Im jeweils letzten Block der ersten 2 Phasen sind im Diagramm wieder alle Löser aufgeführt.

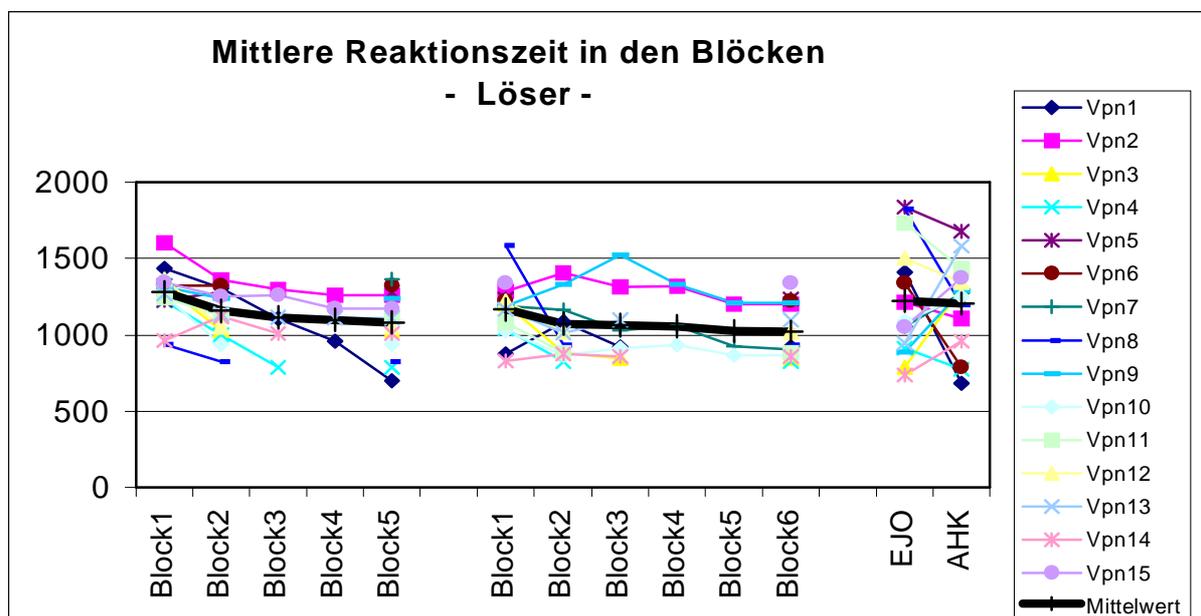


Abb. 2.4: Mittlere Reaktionszeit in Sekunden und Mittelwert der einzelnen Versuchspersonen, aufgetragen über die Blöcke in den 2 Lernphasen und über die Testphase.

²² Auf die Bonferroni-Korrektur wurde hier und bei den im Folgenden berichteten Ergebnissen aufgrund der geringen Anzahl an abhängigen Testungen verzichtet.

²³ Würden die Werte nicht im Gesamtmittelwert fortgeführt, so würde der Gesamtmittelwert aufgrund abnehmender Datenpunkte instabiler werden. Weiterhin würde der Gesamtmittelwert künstlich erhöht, da nur noch „langsamere“ Lerner in späteren Blöcken verrechnet würden. Allerdings führt diese Darstellung dazu, dass eventuelle Veränderungen in der Reaktionszeit unterdrückt werden, da der Gesamtmittelwert durch die bereits erfolgreichen Löser, die sich nicht mehr verbessern können, stabilisiert wird.

Alle Löser blieben in ihrer Reaktionszeit auf die Testitems unterhalb des Zeitlimits für die Experimentalitems von 3000ms. Die Reaktionszeit auf das Testitem EJO war nicht signifikant langsamer als die Reaktionszeit auf das jeweilige letzte Item der Phase 2 ($t(14)=-1.50$, $p>0.10$). Die Reaktionszeit auf das Testitem AHK unterschied sich signifikant von dem individuell letzten Item ($t(14)=-2.25$, $p<0.05$). Die Reaktionszeit auf AHK unterschied sich nicht signifikant von der Reaktionszeit auf Testitem EJO ($t(14)=0.19$, $p>0.10$).

Gesamt-Konsistenz

Die Gesamt-Konsistenz der Klassifizierungsstrategie wurde dadurch operationalisiert, dass für jede Kategorie bestimmt wurde, welcher Taste sie am häufigsten zugewiesen wurde. Die Anzahl dieser Zuweisungen wurden dann für jeden Block über die 3 Kategorien aufsummiert. Abbildung 2.5 zeigt die Zunahme der Gesamt-Konsistenz über die Blöcke hinweg.

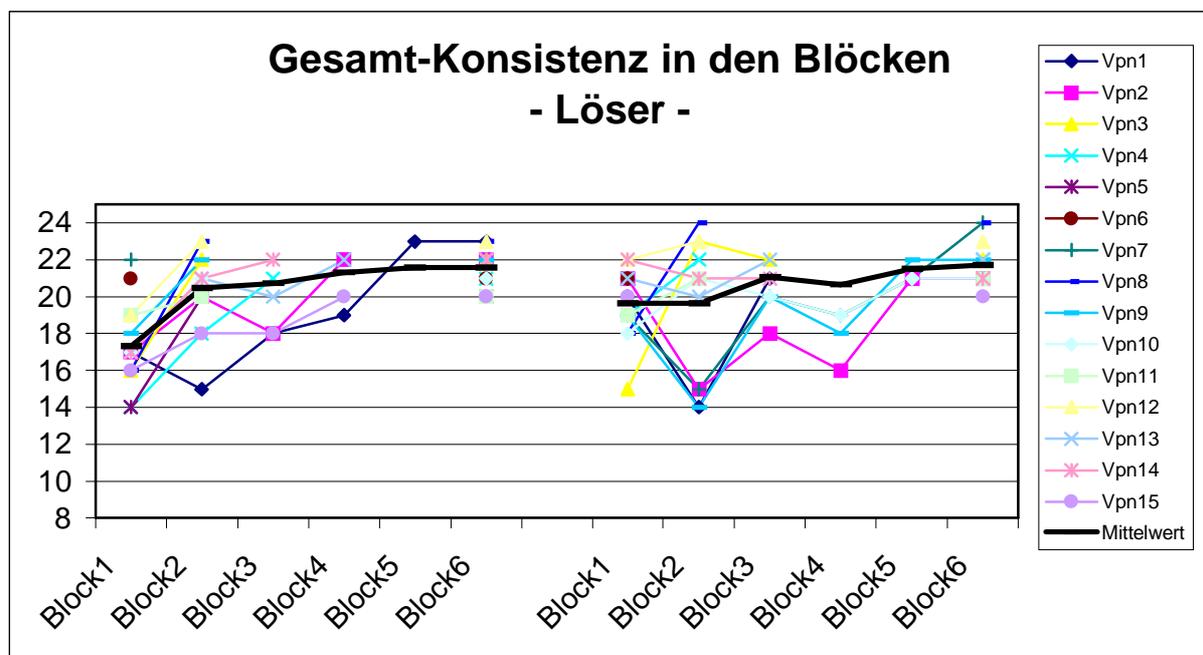


Abb. 2.5: Zunahme der Gesamt-Konsistenz über die Blöcke hinweg. Für den Gesamtmittelwert wurde das bei der Reaktionszeit diskutierte Verfahren verwendet.

Da eine höhere Konsistenz in einer Kategorie eine höhere Konsistenz in den anderen Kategorien begünstigt kann davon ausgegangen werden, dass die Normalverteilungsannahme auch theoretisch verletzt sein muss. Deshalb wurde der Wilcoxon Rangsummentest verwendet. Die Gesamt-Konsistenz, welche im 1. Block der 2. Phase gezeigt wurde, ist signifikant höher als die Konsistenz im 1. Block der 1. Phase (Wilcoxon $Z=-2.36$, $p<0.05$). Zwischen den letzten Blöcken der beiden Phasen besteht kein signifikanter Unterschied mehr (Wilcoxon $Z=-0.58$, $p>0.10$). Die Zunahme an Gesamt-Konsistenz innerhalb beider Phasen vom ersten zum individuell letzten Block ist signifikant (Phase 1: Wilcoxon $Z=-3.20$, $p<0.01$;

Phase 2: $Z=-2.73$, $p<0.01$). Der letzte Block der ersten Phase unterscheidet sich signifikant vom ersten Block der zweiten Phase (Wilcoxon $Z=-2.62$, $p<0.01$).

Unterschiede zwischen den Kategorien

Es sollte auch untersucht werden, ob Unterschiede in der Geschwindigkeit, mit der die Kategorien erworben wurden, bestanden. Dazu wurde untersucht, nach wie vielen Blöcken in der 2. Phase die Stimuli einer Kategorie zum erstenmal konsistent einer Taste zugeordnet wurden. Es zeigt sich kein reliabler Unterschied zwischen den Kategorien (Kat1-Kat2: $t(14)=0.27$, $p>0.10$; Kat1-Kat3: $t(14)=0.00$, $p>0.10$; Kat2-Kat3: $t(14)=-0.20$; $p>0.10$). Für die ersten konsistenten Blöcke der jeweiligen Kategorien wurde dann die Anzahl an konsistenten Zuweisungen zu den 3 Kategorien berechnet. Auch hier zeigt sich kein signifikanter Unterschied (Kat1-Kat2: $t(14)=0.18$; Kat1-Kat3: $t(14)=-0.25$; Kat2-Kat3: $t(14)=-0.38$; für alle: $p>0.10$). Da dieses Maß auch in den folgenden Experimenten keine verlässlichen Unterscheidungen ermöglicht, wird es im folgenden nicht mehr berichtet.

Tastenzuweisung

13 der 15 erfolgreichen Versuchspersonen ordneten den ersten Stimulus der Taste „1“ zu, 2 Versuchspersonen ordneten den ersten Stimulus der Taste „2“ zu. Der Unterschied zu einer Gleichverteilung auf allen 3 Tasten ist signifikant ($\chi^2(2)=19.60$, $p<0.001$; $B_{15;0.66}(2)<0.001$; $B_{15;0.33}(2)>0.05$). Für Teststimulus EJO wich die Zuordnung mit 6 mal zu Taste 1, 4 mal zu Taste 2 und 5 mal zu Taste 3 nicht mehr signifikant von einer Gleichverteilung ab ($\chi^2(2)=0.40$, $p>0.10$). Bei den Nicht-Lösern ergaben sich für den ersten Stimulus die folgenden, signifikant von der Gleichverteilung abweichenden Zahlen: 12, 4 und 4 ($\chi^2(2)=6.40$, $p<0.05$; $B_{20;0.66}(8)<0.05$; $B_{20;0.33}(4)>0.10$). 1 Versuchsperson reagierte nicht in der gegebenen Zeit. Für Teststimulus EJO ergaben sich bei den Nicht-Lösern die folgenden, nichtsignifikant von der Gleichverteilung abweichenden Zuordnungen: 8 mal zu Taste „1“, 6 mal zu Taste „2“ und 7 mal zu Taste „3“ ($\chi^2(2)=1.64$, $p>0.10$).

Konsistenz der Tastenzuweisung

Im letzten Block der 2. Trainingsphase ordneten 13 Versuchspersonen die Stimuli der Kategorie 1 derselben Taste zu, welche sie für Stimuli der Kategorie 1 im letzten Block der 1. Phase benutzen. ($B_{15;0.66}(2)<0.001$). Für Kategorie 2 und 3 zeigten 11 Versuchspersonen diese Konsistenz ($B_{15;0.66}(4)<0.01$). Die 1. Kategorie zeigte nicht signifikant häufiger eine Zuordnungskonsistenz als die anderen 2 Kategorien ($B_{35;0.66}(22)>0.10$). Für die

Rechtfertigung der Wahrscheinlichkeitsannahme von $\frac{2}{3}$ für Kategorie 2 und 3 zusammen, obwohl eine Abhängigkeit zwischen den Kategorien besteht, siehe entsprechende Diskussion zur Abhängigkeit bei der Konsistenzbestimmung der Tastenzuweisung. Die Zuordnungskonsistenz ist in Abbildung 2.6 dargestellt. Bei den Nicht-Lösern behielten 15 der 21 Versuchspersonen die 1. Kategorie bei ($B_{21;0.66}(6) < 0.001$), 5 die 2. ($B_{21;0.66}(16) > 0.1$) und 7 die 3. ($B_{21;0.66}(14) > 0.1$).

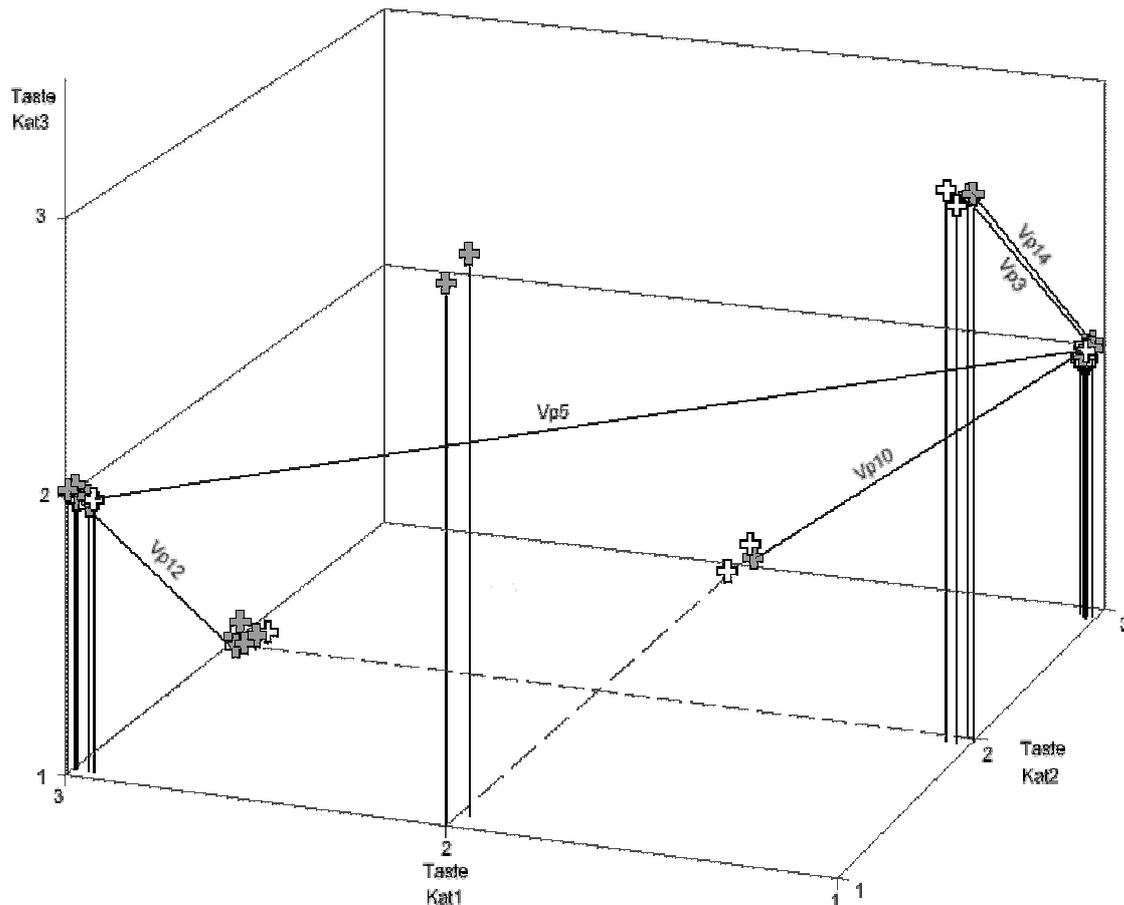


Abb. 2.6: Die drei Dimensionen repräsentieren die Tasten, welche den Stimuli aus den 3 Kategorien zugeordnet wurden. Jede Versuchsperson ist durch zwei Kreuze repräsentiert. Senkrechte schwarze Linien sollen die Interpretation der Lage der Kreuze im Raum erleichtern. Dem gleichen Zweck dienen die eingezeichneten Hilfslinien. Ein Kreuz zeigt die Tastenzuordnung zu den 3 Kategorien am Ende der 1. Phase, das zweite Kreuz am Ende der 2. Phase. Wenn ein Wechsel der Zuordnung erfolgte, ist dies durch eine Linie dargestellt die mit der Versuchspersonennummer gekennzeichnet ist. So repräsentiert ein Kreuz am Punkt (1,3,2) eine Versuchsperson die entweder in der 1. oder 2. Phase die 1. Kategorie der Taste 1, die 2. Kategorie der Taste 3 und die 3. der Taste 2 zugeordnet hat. Da hier nur erfolgreiche Versuchspersonen dargestellt sind, bleiben Punkte mit gleichen Werten auf mehr als einer Dimension unbesetzt. Dunklere Kreuze repräsentieren Versuchspersonen welche EJO in Kategorie 3 eingeordnet haben. Die Punkte liegen alle auf diskreten Werten 1-3. Lediglich zu Darstellungszwecken wurden die Punkte hier wie auch in den folgenden Darstellungen durch Addition von Zufallszahlen leicht verschoben.

Kategorienwahl

Als eigentliche abhängige Variable des Designs dient die Zuordnung des Teststimulus EJO in der Testphase zu einer Kategorie. 10 Versuchspersonen ordneten EJO in die Kategorie 3 ein, 3 in Kategorie 1, welche durch AB geblockt sein sollte und 2 in die ungeblockte Kategorie 2. Ein Chi-Quadrat Test ergab eine signifikante Abweichung von der Gleichverteilung ($\chi^2(2)=7.6$, $p<0.05$; $B_{15;0.66}(5)<0.01$; $B_{15;0.33}(3)>0.10$; $B_{15;0.33}(2)<0.10$). Die Daten sind in Abbildung 2.7 aufgeführt.

Zuordnung EJO			
	Kategorie 1	Kategorie 2	Kategorie 3
Löser	3	2	10
Analysierbare Nicht-Löser	2	5	2

Zuordnung AHK			
	Kategorie 1	Kategorie 2	Kategorie 3
Löser	4	8	3
Analysierbare Nicht-Löser	2	5	2

Abb. 2.7: Zuordnung der Teststimuli EJO und AHK zu den Kategorien. Aufgegliedert in Löser und auswertbare Nicht-Löser.

Der Stimulus AHK wurde 4 mal der Kategorie 1, 8 mal Kategorie 2 und 3 mal Kategorie 3 zugeordnet (vgl. Abbildung 2.7). Ein Chi-Quadrat Test ergab keine signifikante Abweichung von der Gleichverteilung ($\chi^2(2)=2.80$, $p>0.10$). Von den Nicht-Lösern konnten nur diejenigen in der Analyse erfasst werden (analysierbare Nicht-Löser), die in der letzten Phase eine Präferenz zeigten, die 3 Tasten aus Sicht des Designs unterschiedlichen Kategorien zuzuordnen. Dies war nur für 9 der Nicht-Löser der Fall. Die meisten der nicht-analysierbaren Nicht-Löser bildeten weniger als 3 Kategorien. Einige wenige der nichtanalysierbaren Nicht-Löser ordneten dieselbe Anzahl an Stimuli aus einer Kategorie 2 verschiedenen Tasten zu. Die Daten der analysierbaren Nicht-Löser wurden in Abbildung 2.7 übernommen, auf eine separate statistische Auswertung wurde verzichtet.

Bekanntheitsrating

Anschließend wurde untersucht, ob die Icons des Buchstabens aus der am häufigsten gewählten Kategorie 3 als bekannter als die Icons aus den anderen beiden Kategorien eingeschätzt wurden (vgl. Abbildung 2.8). Es zeigte sich kein signifikanter Unterschied zwischen den Icons aus Kategorie 3 verglichen mit Icons aus Kategorie 1 und 2 ($t(14)=1.35$, $p>0.10$; $t(14)=1.78$, $p<0.10$).

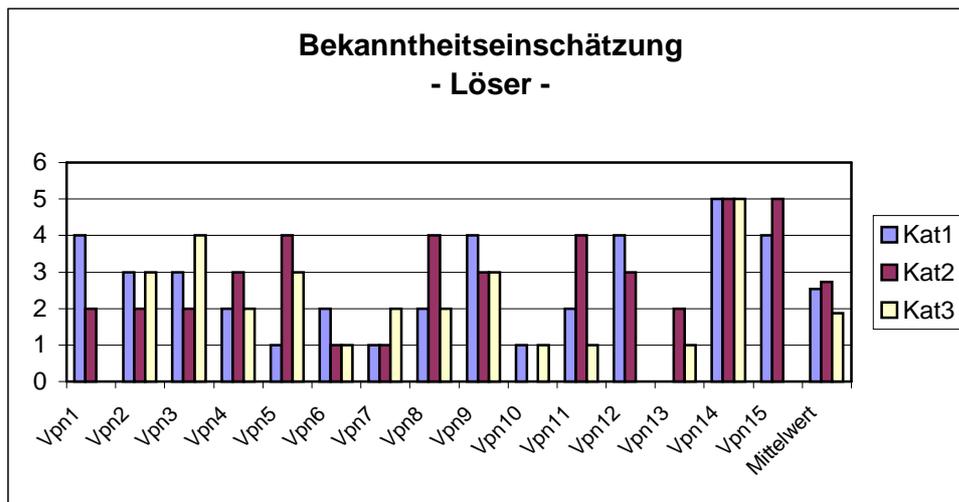


Abb. 2.8. Bekanntheitseinschätzung der Icons aus den Kategorien der 2. Phase.

Auskünfte der Teilnehmer

Da der Fokus der Auswertung auf die Löser ausgerichtet war, sollte das Auswertesystem aus den Beschreibungen der Nicht-Löser gewonnen werden, um eine Zirkelbestätigung der eigenen Auswertekategorien möglichst zu vermeiden. Deshalb wurden die Beschreibungen, welche die Versuchspersonen nach dem Experiment gaben, zuerst für die Nicht-Löser durchgesehen, um sinnvolle Kategoriensysteme zu finden. Die Äußerungen wurden daraufhin danach unterschieden, ob sie nahe legen, dass die Versuchsperson auf die Farben oder die Form der Icons geachtet hat. Eine zweite Einordnung unterschied, ob die Beschreibung darauf hindeutete, dass die Kategorisierung eher aufgrund eines einzelnen oder mehrerer Icons erfolgte. Beide Unterscheidungsmerkmale wurden jeweils als exhaustiv und exklusiv aufgefasst um zu verhindern, dass bei den häufig vorhandenen Mehrdeutigkeiten keine oder alle Möglichkeiten gewählt wurden. Natürlich war es möglich, dass die Versuchspersonen ihre Klassifizierungsgrundlage im Laufe des Versuches änderten. Es sollte aber die überwiegende Art der Klassifizierung herausgelesen werden. Eine nicht exhaustive Kategorie, welche erst nach Lesen der Nicht-Löser- als auch der Löser-Kommentare formuliert wurde, erfaßte, ob die Versuchspersonen vermutlich eine Restkategorie gebildet haben. Eine Restkategorie war dadurch definiert, dass dieser alle Stimuli zugeordnet wurden, welche nicht in eine der 2 anderen Kategorien zu passen schienen. Abschließend wurde entschieden, ob die Versuchsperson die Art ihrer Einordnung strukturell verändert hat. Wechsel von einem Icon zu einem anderen wurden dabei nicht berücksichtigt, sondern nur strukturelle Veränderungen wie z.B. der Wechsel von Farbenklassifizierung zu Formklassifizierung. Anhang 8 führt die

schriftlichen Äußerungen und ihre Einordnung auf²⁴. Die Klassifikation wurde von einer nicht über den Hintergrund der Einordnung informierten Person wiederholt. Ihr Instruktionstext ist in Anhang 7 aufgeführt.

Löser

Die Erst- und Zweittrater stimmten bezüglich der Frage, ob die Klassifizierung aufgrund der Iconstruktur oder Farbe erfolgte, 13 mal überein. Nur einmal kam es zu einer widersprüchlichen Entscheidung ($B_{14;0.5}(1) < 0.001$)²⁵. 1 Versuchsperson hatte keine Angaben gemacht. Für die Entscheidung über die Anzahl der verwendeten Icons gab es 11 Übereinstimmungen und 3 unterschiedliche Urteile ($B_{14;0.5}(3) < 0.05$). Bei der Entscheidung über die Verwendung einer Restkategorie kam es nur zu einer Differenz. Ebenso bei der Entscheidung über das Vorliegen eines Strategiewechsels ($B_{14;0.5}(1) < 0.001$).

Nicht-Löser

Bei 21 Urteilen zeigten sich 3 Nichtübereinstimmungen bei der Entscheidung zwischen Iconstruktur und Farbe ($B_{21;0.5}(3) < 0.001$), 1 Nichtübereinstimmung bei der Entscheidung über die Anzahl der beachteten Icons ($B_{21;0.5}(1) < 0.001$), 4 bei der Restkategorieentscheidung ($B_{21;0.5}(4) < 0.01$) und 2 bei der Einschätzung, ob ein Strategiewechsel stattgefunden hat ($B_{21;0.5}(2) < 0.001$).

Damit erwiesen sich die Einschätzungen der Strategien für die Löser und Nicht-Löser als hinreichend reliabel, um die Einschätzungen statistisch auszuwerten.

Strategien der Löser und Nicht-Löser

Die Anzahl an Einstufungen in den einzelnen Bereichen wurden über die beiden Bewerter gemittelt. Die berichteten Zahlen beziehen sich dabei auf die Löser. In Klammern sind die Zahlen für die Nicht-Löser aufgeführt. Durchschnittlich 46% (88%) der ausgewerteten Kommentare wurden dahingehend interpretiert, dass nach Farben klassifiziert wurde. In 39% (2%) der Fälle, dass ein einzelnes Icon die Klassifizierungsgrundlage darstellte. 18% (33%) der Aussagen wurden als Beschreibung der Verwendung einer Restkategorie aufgefasst und 7% (24%) der Versuchspersonen schienen die Strategie gewechselt zu haben.

²⁴ Die Texte sind im Anhang, was Rechtschreibung und Grammatik angeht, unverändert wiedergegeben worden. Die Sprache spiegelt vermutlich den hohen Anteil an Austauschstudenten unter den Teilnehmern wider.

²⁵ Es wurde hier die vereinfachende Annahme getroffen, dass die Übereinstimmungswahrscheinlichkeit zwischen Erst- und Zweittrater 0.5 beträgt. Es wurde also davon ausgegangen, dass keine Urteilstendenzen bei den Ratern bestehen.

Vergleich Löser mit Nicht-Löser

Ein Chi-Quadrat Test auf Unterschiede zwischen den Lösern und Nicht-Lösern zeigte einen signifikant von der Gleichverteilung abweichenden Chi-Wert für die 1. Beurteilungsdimension²⁶ bezüglich Farb- oder Formbeurteilung ($\chi(1)=7.15$, $p<0.01$), ebenso für die 2. Beurteilungsdimension, welche die Anzahl der genutzten Icons erfasste ($\chi(1)= 8.05$, $p<0.01$). Auf eine exakte Bestimmung des p-Wertes wurde aufgrund des hohen Chi-Wertes und des Vorliegens nichtganzzahliger Zellohäufigkeiten verzichtet, obwohl die erwartete Zellohäufigkeit in einer Zelle kleiner als 5 war. Bei den folgenden Tests waren die Chi-Werte nicht ausreichend extrem ausgeprägt. Aus diesem Grund wurde ein Fisher-Exact-Test gerechnet. Dazu wurden nichtganzzahlige Zellohäufigkeiten zu Ungunsten eines Unterschiedes auf oder abgerundet wurden. Der Fisher-Exact-Test ergab keine signifikant häufigere Restkategorieeinstufung bei den Nicht-Löser als bei den Lösern (Fisher-Exact-Test $p>0.10$). Für die Wechsel-Dimensionen ergab sich ebenfalls kein signifikanter Unterschied zwischen den Lösern und Nicht-Lösern (Fisher-Exact-Test, $p>0.10$).

2.2.3 Diskussion Experiment 1

Nahezu die Hälfte der Teilnehmer erwarben die erwartete Kategorienstruktur in den 2 Lernphasen. Trotz der komplizierten Kategoriestructur und dem Verzicht auf jegliche Rückmeldung war die Ähnlichkeit von Stimuli aus denselben Kategorien hoch genug, um das Erlernen zu ermöglichen. Da die Aussage des Experimentes sich nur auf Lernvorgänge bezieht, sollte der Ausschluss von nahezu der Hälfte der Versuchspersonen keine Verzerrung darstellen. Selbst unter der Annahme, dass die ausgeschlossenen Versuchspersonen eine gewisse Strukturierungsform gelernt haben, spielt dies nur dann eine Rolle, wenn das Experiment die verschiedenen Strategien von Versuchspersonen aufdecken soll. In diesem Fall müssten alle Versuchspersonen berücksichtigt werden.

Die Tatsache, dass das Verhältnis der Löser zu den Nicht-Lösern in der 1. und 2. Erhebungsstufe konstant blieb, deutet darauf hin, dass die Schwierigkeit und damit die Wahrscheinlichkeit eine spezielle Kategorienstruktur zu erlernen, nicht objektiv gegeben ist, sondern nur in Interaktion mit Subjektvariablen zu verstehen ist. Andernfalls hätten die in der 1. Erhebungsstufen nicht gelösten Kategorienstrukturen bei der erneuten Präsentation in der 2. Erhebungsstufe die Lösewahrscheinlichkeit vermindern müssen.

²⁶ Der Begriff einer Dimension wird hier nicht im mathematischen Sinne einer kontinuierlichen Ordnung, sondern als binäre Beurteilungsdimension verstanden.

Anzahl benötigter Blöcke

Da die Versuchspersonen theoretisch in der 2. Phase Wissen über Kategorien 1 und 3 fortführen konnten, sollte die 2. Phase einfacher als die 1. zu lösen sein. Dies spiegelt sich aber nicht in einer unterschiedlichen Anzahl benötigter Blöcke wider. Auch die niedrige Korrelation legt den Schluss nahe, dass kein Lerntransfer zwischen der 1. und 2. Phase stattfand. Dass dies aber ein Artefakt eines zu wenig sensitiven Maßes darstellen könnte, zeigen die durchgeführten Schwierigkeitsanalysen zur Reaktionsgeschwindigkeit und Gesamt-Konsistenz.

Reaktionsgeschwindigkeit und Gesamt-Konsistenz

Auffallend ist, dass die Versuchspersonen in ihrer Reaktionszeit im Mittelwert weit unter dem Zeitlimit von 3000ms blieben. Unter der Annahme, dass 3000ms eine nicht ausreichende Zeit für eine ausgiebige Analyse der Stimuli darstellt (vgl. teilweise höhere Reaktionszeiten in Experiment 2), deutet dies darauf hin, dass Versuchspersonen ihre Analysestrategien so abändern können, dass ein Zeitlimit eingehalten werden kann (vgl. auch Wills, 2002 und Unterschied in den Ergebnissen in Experiment 1, 2 und 3 bei unterschiedlichen Zeitlimits).

Sowohl in der Reaktionszeit als auch in der Gesamt-Konsistenz zeigte sich eine Verbesserung innerhalb der Phasen wie bei kontinuierlichem Lernen zu erwarten wäre. Die Tatsache, dass kein signifikanter Unterschied in der Reaktionszeit zwischen dem letzten Block der 1. Phase und dem 1. Block der 2. Phase gefunden wurde, scheint einen gewissen Lerntransfer zu belegen. Die Nicht-Signifikanz könnte aber auch darauf zurückgeführt werden, dass die Reaktionszeit von der Anzahl der präsentierten Blöcke, nicht aber vom Kompetenzniveau abhängt. Da beide Blöcke aufeinander folgen, wäre unter dieser Annahme kein signifikanter Unterschied zu erwarten. Allerdings zeigte sich bei einem Vergleich derselben Blöcke hinsichtlich ihrer Gesamt-Konsistenz ein signifikanter Unterschied. Da Konsistenz ein Maß für die Kompetenz darstellt, bedeutet der signifikante Abfall an Konsistenz zu Beginn der 2. Phase, dass eine in der 1. Phase erworbene Kompetenz nicht vollständig in die 2. Phase übernommen werden konnte. Diese Hypothese wird aber durch die Beobachtung abgeschwächt, dass die Konsistenz im ersten Block der 2. Phase signifikant höher als im ersten Block der 1. Phase ist. Die Versuchspersonen begannen also in Phase 2 auf einem höheren Niveau und zeigten damit vermutlich doch eine Form von Lerntransfer. Allerdings war kein wesentlicher Unterschied zwischen den letzten Blöcken der beiden Phasen festzustellen. Dies liegt aber vermutlich daran, dass das Kriterium eine gewisse Mindestkonsistenz als Randbedingung vorschreibt.

Der Widerspruch zu der Anzahl benötigter Blöcke, die keinen Lerntransfer vermuten lässt, kann folgendermaßen aufgelöst werden: Die benötigte Blockanzahl könnte damit zusammenhängen, wie häufig Icons präsentiert werden müssen, um eine ausreichende Verbindung zu einer Kategorie zu bilden. Eine Verbesserung des Konsistenz- und Reaktionsmaßes lässt sich aber auch ohne diese gefestigte Bindung dadurch erklären, dass gelernt wird, auf einzelne Icons selektiv zu achten. Dieser der Aufmerksamkeitszuteilung verwandte Prozess würde eine schnellere Reaktion ermöglichen, da weniger Icons beachtet werden müssen. Gleichzeitig könnte es durch die Selektion der Informationsquellen möglich werden, die geringere Menge an relevanter Information innerhalb des Zeitlimits zu extrahieren und damit die Konsistenz zu erhöhen. Dies müsste aber nicht zwangsläufig zu einer höheren Rate an Kategorisierungen im Sinne des Designs, gemessen durch die benötigte Blockanzahl, führen. Das Metawissen wäre nach dieser Hypothese in die zweite Phase übernommen worden, während die Verbindungen selbst nicht schneller gelernt werden konnten (vgl. auch Annahmen zur Salienz im weiter unten diskutierten Simulationsmodell).

Da Stimulus EJO gleichzeitig Elemente aus verschiedenen Kategorien enthält, wäre für den Fall, dass die Informationseinheiten aus der 1. und 3. Kategorie nicht vollständig geblockt sind, ein gewisser Konflikt zu erwarten. Dieser Konflikt sollte die Reaktionszeiten auf Stimulus EJO verglichen mit dem jeweilig letzten Item aus Phase 2 verlängern. Aufgrund des kurzen zeitlichen Abstandes des Stimulus EJO zum letzten Experimentalitem kann eine Konfundierung durch den Grad der Beherrschung der Kategorisierungsaufgabe oder der Anzahl gesehener Reize ausgeschlossen werden. Auch die Aufhebung des Zeitlimits für EJO spielte keine Rolle, da alle Reaktionszeiten auf das Testitem unter 3 Sekunden blieben. Dass dieser Unterschied in der Reaktionszeit zwischen EJO und dem letzten Trainingsitem nicht signifikant wurde, kann an der relativ großen Varianz von Reaktionszeitmaßen allgemein liegen. Die große Varianz dürfte vor allem bei diesem Test relevant sein, da jeweils nur 2 Reaktionszeiten pro Versuchsperson verglichen wurden. Den nicht vorhandenen Effekt allein auf die geringe Power zurückzuführen wird aber der Tatsache nicht gerecht, dass der gleiche Test mit Teststimulus AHK aufgrund der geringeren Standardabweichung signifikant wurde. Es scheint also theoriekonsistent ein geringerer Konflikt durch Blocking bei Stimulus EJO als bei Stimulus AHK, der mit allen Kategorien über nicht geblockte Icons verbunden ist, aufzutreten.

Tastenzuweisung

Die Abweichung von der Gleichverteilung bei der Tastenzuweisung für den ersten Stimulus nicht aber für den Teststimulus zeigt, dass die Händigkeit oder andere vor dem Experiment bestehende Präferenzen für einzelne Tasten durch das Lernen besserer Entscheidungsstrategien in Form von Kategorien, irrelevant wurden. Diese Präferenz, bei einem überraschenden Ereignis, bzw. bei Unentschlossenheit eine spezielle Taste zu betätigen, lässt erwarten, dass bei Unsicherheit eine Bevorzugung der Kategorie 1 entstehen sollte. Der Grund hierfür ist, dass der 1. stark veränderte und deshalb mit hoher Unsicherheit verbundene Stimulus in Phase 2 aus Kategorie 1 entstammt. Wird bei dem 1. Teststimulus aufgrund von erhöhter Unsicherheit, hervorgerufen durch den großen „Konflikt“, die gleiche Taste bevorzugt, so wird der Teststimulus Kategorie 1 zugeordnet. Diese Bevorzugung sollte umso ausgeprägter sein, je unsicherer die Zuordnung des 1. Stimulus in der 2. Phase erfolgte (vgl. Ergebnisse der Kategorienzuordnung der Nicht-Löser in Experiment 3).

Konsistenz der Tastenzuweisung

Die Stabilität der Tastenzuweisungen validiert die Messung der Kategorienzuordnung, da sie zeigt, dass bei gleicher Kategorienempfindung dieselbe Taste gewählt wird. Numerisch wird die Tastenzuordnung vom letzten Block der 1. Phase zum letzten Block der 2. Phase am häufigsten in Kategorie 1 übernommen. Dies ist konsistent mit dem vorgeschlagenen Delta-Modell, da Kategorie 1 die größte Anzahl gemeinsamer Icons in der 1. und 2. Phase besitzt. Diese stärkere Konsistenz für die 1. Kategorie lässt sich aber statistisch nicht nachweisen. Dies ist unmittelbar einsichtig, wenn man sich verdeutlicht, dass eine konsistente Zuordnung einer Taste zu einer Kategorie eine gewisse Konsistenz bei den anderen Tasten-zu-Kategorien Zuordnungen erzwingt. Die signifikanten Ergebnisse für die einzelnen Kategorien müssen daher mit Vorsicht betrachtet werden, da die Annahme, dass die Wahrscheinlichkeit zufällig dieselbe Kategorie zu wählen $\frac{1}{3}$ beträgt, voraussetzt, dass die Wahl unabhängig von den anderen Kategorienzuordnung erfolgt. Da spätestens nach 2 Kategorienzuordnungen und der Randbedingung, dass alle 3 Tasten benutzt werden sollen, die 3. Kategorie festgelegt ist, ist diese Annahme verletzt. Unter der Annahme, dass aber keine Kategorie bevorzugt die Kategorienzuordnung festlegt ergibt sich die Wahrscheinlichkeit, dass eine Kategorie x „blind“ erneut einer Taste „y“ zugeordnet wird aus der Summe der Wahrscheinlichkeiten, dass Kategorie x mit Taste „y“ belegt wird, wenn Kategorie x als 1., 2. oder 3. Kategorie einer Taste fest zugeordnet wird zu:

$$\frac{1}{3} * \binom{1}{3} + \frac{1}{3} * \binom{2}{3} * \binom{1}{2} + \frac{1}{3} * \binom{2}{3} * \binom{1}{2} * 1 = \frac{1}{3}$$



Kategorie x wird
als erste Kategorie
fest zugeordnet



Kategorie x wird
als zweite Kategorie
fest zugeordnet



Kategorie x wird
als dritte Kategorie
fest zugeordnet

was aus Symmetriegründen auch unmittelbar einsichtig ist. Ein signifikantes Ergebnis kann somit zweierlei bedeuten: Erstens, dass die Tastenzuordnung valide Aufschluss darüber gibt, dass *diese* Kategorie konsistent behandelt wurde oder zweitens, dass die anscheinend konsistente Kategorienbehandlung durch die konsistente Behandlung einer *anderen* Kategorie erzwungen wurde.

Es wurde nur der letzte Block der 1. Phase mit dem letzten Block der 2. Phase verglichen, um zu garantieren, dass die Struktur erlernt worden war. Die näherliegende Testung des 1. Blocks der 2. Phase gegen den letzten Block der 1. Phase brächte die Schwierigkeit mit sich, dass, im Extremfall, eine Versuchsperson, welche auf alle Stimuli in Block 1 der 2. Phase mit derselben Taste reagiert maximale Konsistenz für eine Kategorie zeigen würde, obwohl nicht wirklich einer Kategorie eine Taste zugewiesen wurde. In diesem Fall hätte die Versuchsperson keine klare Zuordnung der Tasten zu den Kategorien entwickelt, selbst wenn bereits eine gewisse Kategorienvorstellung vorhanden sein mag. Aus denselben Gründen muss das Ergebnis für die Nicht-Löser vorsichtig interpretiert werden.

Kategorienzuordnung – Stimulus EJO

Um zu untersuchen, ob die erhaltenen empirischen Daten mit den Vorhersagen eines Delta-Netzwerkes übereinstimmen, wurden 24 Versuchspersonen simuliert. In den 24 Simulationen wurden alle Präsentationskombinationen des empirischen Experimentes einmal verwendet. Auf eine 2. Erhebungsstufe wie in der empirischen Studie wurde verzichtet, da das Netzwerk genügend „Löser“ simulierte. Für eine genauere Beschreibung des Modells siehe Anhang 9 „Konnektionistisches Delta-Modell I“.

Ein ernstes Problem stellt für das Delta-Modell die häufige Zuweisung des Stimulus EJO zur 3. Kategorie im Experiment dar. Im Gegensatz zu den empirischen Daten ergab die Simulation mit dem Delta-Modell eine Bevorzugung der 2. Kategorie für einen weiten Lernparameterbereich. Eine Ausnahme stellten Simulationen mit sehr kleinen Lernparametern dar. Wenn kleine Lernparameter die asymptotische Verknüpfung von Icons mit einer

Kategorie verhinderten, konnte nur ein sehr kleiner Blocking-Effekt entstehen und die Zuweisung des Teststimulus EJO erfolgte relativ zufällig.

Andererseits scheint in den empirischen Daten doch ein starker selektiver Lerneffekt vorzuliegen, da ansonsten keine signifikante Bevorzugung *einer* Kategorie für den Teststimulus EJO erfolgen dürfte. Dieser selektive Lerneffekt kann nicht durch die 2. Phase erklärt werden, da hier alle Stimuli gleichberechtigt präsentiert wurden. Eine zufällige Bevorzugung einer Kategorie in dieser Ausprägungsstärke scheint unplausibel. Vor allem, da das „Matching“ von Versuchspersonen garantierte, dass diejenigen Icons, welche für eine Versuchsperson aus Kategorie 2 stammten, für die beiden anderen „gematchten“ Versuchspersonen aus der 1. und 3. Kategorie stammten. Eine allgemein größere Auffälligkeit der Icons aus Kategorie 3 kann somit als Erklärung nahezu ausgeschlossen werden. Gegen eine Erklärung durch Zufall spricht auch, dass die Löser ein Ähnlichkeitsmaß verwendet hatten, welches die Anzahl gemeinsamer Icons als stark determinierende Komponente enthält. Ansonsten hätten die Versuchspersonen das Kriterium in der 1. und 2. Phase mit großer Wahrscheinlichkeit nicht erreicht und wären damit nicht in die statistische Auswertung aufgenommen worden. Auffallend ist auch, dass EJO so gut wie nie der geblockten Kategorie zugewiesen wurde. Lediglich schien Kategorie 3, welche im Vergleich mit Kategorie 1 nur durch halb so viele bekannte Icons geblockt wurde, eine noch stärkere Bindung zu EJO als die überhaupt nicht geblockte Kategorie 2 aufzuweisen. Dies deutet auf eine größere Aufmerksamkeit für die 3. Kategorie, welche sich stärker als Kategorie 1 änderte, hin. Zwei Erklärungen bieten sich für das erhaltene Ergebnis an²⁷:

Durch den Blocking-Effekt bedingt werden Kategorie 2 und 3 gegenüber Kategorie 1 bei Zuweisung des Teststimulus bevorzugt. Zu erklären bleibt also nur noch, warum Kategorie 3 häufiger als Kategorie 2 gewählt wird. Dazu wird davon ausgegangen, dass die Teilnehmer der Studie nur ein begrenztes Gedächtnis besitzen und nicht mehr als 3 Prototypen im Gedächtnis halten können. In diesem Fall können Prototypen der Kategorien 1 und 3 auch noch in der 2. Phase repräsentiert werden, da diese den Prototypen aus Phase 1 genügend ähnlich sehen. Der Prototyp aus Kategorie 2 kann aber, da er aus vollkommen neuen Informationen besteht, nicht mehr im Gedächtnis gehalten werden. Daher resultiert eine geringere Stärke der Repräsentation von Kategorie 2, was eine Zuordnungstendenz zu dieser Kategorie verhindert.

²⁷ Auf die erste Erklärung hat mich Andy Wills (Email am 08.01.2003) hingewiesen.

Die Setzung der Kapazitätsgrenze, auf welche die Erklärung aufbaut, erscheint mir relativ willkürlich und es wäre zu erwarten, dass sie über Versuchspersonen hinweg variiert. Weiterhin scheint die Grenze sehr tief angesetzt zu sein, wenn man sie mit der „magischen Nummer 7“ von Miller (1956) vergleicht. Dies kann natürlich damit begründet werden, dass Miller sich auf im Langzeitgedächtnis bereits vorhandene Chunks bezieht und in der Lernphase die Prototypen noch nicht als Chunks, sondern als einzelne Icons repräsentiert werden. Da die Anzahl 3-5 erinnerter Icons in Experimenten gewöhnlich berichtet wird (Email Wills, 08.01.2003), erhält die Zahl eine größere Plausibilität. Allerdings erscheint diese Erklärung, welche darauf aufbaut, dass entsprechende Chunks noch nicht gebildet wurden, durch das Erreichen des Kriteriums und damit verbundenem erfolgreichen Lernen relativ unwahrscheinlich. Außerdem dürften Menschen in der Lage sein, mit metakognitiven Strategien solche Engpässe des Arbeitsgedächtnisses zu kompensieren, vor allem, da Kategorie 2 aus Phase 1 nicht mehr in Phase 2 auftrat. Weiterhin würde diese Erklärung eine ebenso magische Zahl der erinnerten Icons multipliziert mit 3 einführen. Gegen die Erklärung spricht auch, dass der Prototyp aus Kategorie 3 sich in der 2. Phase hinreichend von dem Prototypen der 1. Phase unterscheiden dürfte, so dass auch für diesen eine neue Repräsentation benötigt werden würde.

Unter der Annahme, dass Kategorie 2 in Phase 2 schwächer als die anderen Kategorien repräsentiert ist, sollten die Stimuli aus Kategorie 2 nur nach Ausschluss der anderen Kategorien klassifiziert werden. Empirisch schienen die Versuchspersonen laut ihren Beschreibungen aber nur sehr selten Restkategorien zu benutzen. Andererseits zeigte sich eine signifikant erhöhte Reaktionszeit auf die Stimuli der Kategorie 2 verglichen mit den anderen beiden Kategorien. Über alle Löser hinweg war die Reaktionszeit auf Stimuli der Kategorie 2 signifikant höher als auf Stimuli der Kategorie 3 ($t(346)=2.83$, $p<0.01$) und marginal höher als Stimuli der Kategorie 1 ($t(348)=-1.94$, $p<0.10$). In der 1. Phase unterschieden sich die Reaktionszeiten auf Stimuli der Kategorie 1-3 hingegen nicht signifikant von einander²⁸ (Kat1-Kat2: $t(305)=-1.58$, $p>0.10$; Kat1-Kat3: $t(306)=0.31$, $p>0.10$; Kat2-Kat3: $t(302)=1.41$, $p>0.10$). Diese Ergebnisse sind konsistent mit der Vorstellung, dass Stimuli aus Kategorie 2 nur mit größeren Schwierigkeiten als Stimuli aus anderen Kategorien

²⁸ Antworten, die nicht im gesetzten Zeitlimit erfolgten, gingen nicht in die Analyse ein. Dieses Ausschlusskriterium brachte eine leichte Bevorzugung für kürzere Reaktionszeiten in der daraus entstehenden kleineren Stichprobe mit sich, da die relativ längsten Antworten bereits ausgeschlossen wurden. Da die Anzahl an Ausschlüssen aber, bezogen auf die Gesamtstichprobe, sehr klein war, konnte dieser Fehler vernachlässigt werden. Die Anzahl an Ausfällen durch zu langsame Reaktion unterschied sich zwischen den Kategorien maximal um 4. Damit konnte die Selektion niedriger Werte in der kleineren Stichprobe diese Signifikanz nicht erklären.

klassifiziert werden konnten. Die langsamere Reaktion auf Stimuli der Kategorie 2 muss aber nicht auf eine mangelnde Gedächtnisfähigkeit zurückgeführt werden. Sie kann auch in der Tatsache begründet liegen, dass für Kategorie 2 vollkommen neue Informationen erlernt werden mussten. Eine andere Erklärung für die Bevorzugung von Kategorie 3 dürfte deshalb näher liegen. Für diese Erklärung sind zwei Annahmen nötig:

1. Die Versuchspersonen sahen in Phase 2 Stimuli aus Kategorie 1 als Verzerrungen der Stimuli aus Kategorie 1 in Phase 1 an. Stimuli aus Kategorie 3 in Phase 2 wurden hingegen nicht als Verzerrungen aus Phase 1 wiedererkannt (Wiedererkennungshypothese).
2. Icons, welche bereits in der 1. Phase präsentiert wurden, werden weniger beachtet als neue Icons (Habituationshypothese).

Für die 1. Annahme spricht, dass Stimuli der 1. Kategorie in der 2. Phase 66% der Icons mit der Kategorie 1 in Phase 1 gemeinsam hatten und deshalb vermutlich gut als Verzerrungen des ursprünglichen Prototyps erkannt werden konnten. Die Wiedererkennungshypothese wird weiterhin durch die Beobachtung gestützt, dass die Tastenzuordnung für die 1. Kategorie numerisch am konsistentesten über beide Phasen blieb. Stimuli der 3. Kategorie hingegen hatten nur 33% bzw. 4 Icons mit dem Prototypen aus Phase 1 gemeinsam und konnten aufgrund der geringen Ähnlichkeit laut Hypothese nicht mehr als Verzerrung angesehen werden. Vielmehr wurden sie als eine ganz neue Kategorie aufgefasst. Aus diesem Grunde wurde der neue Prototyp in Phase 2 aus Kategorie 3 wie ein ganz neuer Prototyp, zu dem keine Vorerfahrung bestand, behandelt und die Versuchspersonen mussten die Verbindungen aller Icons zu dem Prototypen neu lernen.

Die erste Forderung impliziert die weitere Annahme, dass die Verbindungen nicht zwischen den Icons und einzelnen Tasten geknüpft werden, sondern über einen Prototypen vermittelt werden. Dies erscheint deshalb sehr plausibel, weil es ermöglicht zu erklären, wie Versuchspersonen eine Regelumkehr einfach vollziehen können. Erst durch diese Abstraktion des Prototyps wird eine Entscheidung, den Prototypen einer anderen Taste zuzuweisen, modellierbar (vgl. auch Gluck & Bower, 1988b). Eine strikt behaviorale Erklärung der Assoziation zwischen einem Stimulus und einer Reaktion ohne zwischengeschaltete mentale Repräsentation dürfte eventuell für tierisches Verhalten angemessen sein, menschlichen kognitiven Leistungen dürfte sie aber nicht gerecht werden.

Die zweite Forderung ist mit verschiedenen Modellvorstellungen vereinbar. Im ursprünglichen Modell der Simulation wurde der Salienzparameter nur aus Sparsamkeits-

gründen weggelassen, da er als nicht für diese Simulation entscheidend angesehen wurde. Für ein „vollständiges“ Modell würde dieser Parameter aber benötigt (vgl. auch Abschnitt 1.2.1, Problem (c) des Rescorla-Wagner Modells aus Sicht von Miller, Barnet und Grahame, 1995). Die Formalisierung der Salienz in dieser Form der Habituationshypothese könnte für das Rescorla-Wagner Modell problematische Beobachtungen erklären. In seiner bisherigen Form sagt das Rescorla-Wagner Modell z.B. vorher, dass die Verbindungsstärke konditionierter Inhibitoren bei Nicht-Vorliegen des US sich vermindert und schließlich 0 erreicht. Diese Abnahme ist aber empirisch nicht zu beobachten (Miller, Barnet & Grahame, 1995). Mit Aufnahme der oben besprochenen Salienzformulierung könnten dieser und vermutlich ähnliche Befunde mit dem Modell vereinbart werden: Da durch die häufige Präsentation des Inhibitors während der Lernphase die Salienz abgenommen hat, vermindern sich die Assoziationsstärken nur noch sehr langsam bis auf 0. Dass diese Formalisierung der Salienz noch nicht häufiger gewählt wurde, mag in der Tatsache begründet liegen, dass sie das Rescorla-Wagner Modell mathematisch schwerer handhabbar macht und Voraussagen daher Computersimulationen nötig machen. Da Menschen aber vermutlich die Aufgaben ebenfalls durch Prozesse lösen und diese nicht zwangsläufig durch einfache mathematische Gleichungen beschreibbar sein müssen, sollte dies kein Gegenargument darstellen.

Aus dem Bereich der Wahrnehmungsmodelle lassen sich sowohl Habituationsmodelle als auch Modelle, die eine größere Aufmerksamkeit für neue Informationen vorhersagen, als Möglichkeiten ansehen, um diese Salienzvorstellung zu verwirklichen. Aus dem Bereich der Kategorisierungsmodelle sind sowohl Vorstellungen, die auf Pearce (1994), Pearce und Hall (1980) als auch Kruschke und Blair (2000) zurückgehen, geeignet. Pearce (1994) geht davon aus, dass neue Einheiten mit einer hohen Assoziativität eingeführt werden. Pearce und Hall (1980) fordern, dass die Salienz abnimmt, wenn eine Informationsquelle reliabel genutzt wurde, was für die K-Icons in der 1. Phase gilt. Kruschke (2001) fordert explizit: “learners will shift attention away from cues that already have been learned as indicative of different diseases [...]”²⁹ (p. 823; vgl. auch Kruschke & Blair, 2000).

Die Habituationshypothese steht nicht im Widerspruch zum Phänomen des Blocking-Effektes. Der Blocking-Effekt bezieht sich darauf, ob Hinweisreize noch gelernt werden müssen. Die Habituationshypothese hingegen legt fest, wie schnell die einzelnen

²⁹ Kruschke und Blair verwenden den Begriff der Aufmerksamkeit für das Konzept, das in dieser Arbeit mit Salienz bezeichnet wurde. Die Terminologie in dieser Arbeit entstammt dem Bereich des Rescorla-Wagner Modells bzw. der Tierstudien in denen eine passive Salienz der Stimuli der geeigneteren Ausdruck gegenüber einer aktiven, vom Subjekt bestimmten Aufmerksamkeit darstellte. Dieser Salienzbegriff wurde aus Konsistenzgründen beibehalten. Kruschke und Johansen (1999) unterscheiden zusätzlich Aufmerksamkeit von der relativ unveränderlichen Salienz eines Stimulus, die in dieser Arbeit überhaupt nicht modelliert wird.

Hinweisreize gelernt werden. Der Blocking-Effekt selektiert also im Experiment die Hinweisreize, welche noch gelernt werden müssen, und die Habituationsannahme bestimmt die Stärke des Lernens.

Die zwei Annahmen führen zu der folgenden Situation:

Phase 1:

Da alle Buchstaben³⁰ neu sind und keiner der Prototypen bevorzugt wird, sollten dem Delta-Modell zufolge alle verwendeten Buchstaben die gleiche Assoziationsstärke zu den Prototypen ausbilden. Da die Gesamtverbindungsstärke nach dem Delta-Modell für jeden Prototypen auf das gleiche Niveau begrenzt ist, wird im folgenden ohne Beschränkung der Allgemeingültigkeit von einem Gesamtassoziationslimit von 1 ausgegangen. Nach erfolgreichem Erlernen der Stimuli in Phase 1 sollten alle Verbindungsstärken der präsentierten Buchstaben 0.33 bzw. 33% der Gesamtverbindungsstärke betragen.

Phase 2:

Die Stimuli der *Kategorie 2* bestehen aus neuen Buchstaben. Sie werden genauso wie die Stimuli der 1. Phase erlernt und enden mit einer Assoziationsstärke von ca. 0.33.

Die Buchstaben AB der *Kategorie 1* haben bereits aus der 1. Phase eine Verbindungsstärke von 0.33. Da E noch keine Verbindungsstärke aufweist und somit das Assoziationslimit noch nicht ausgeschöpft ist, können alle 3 Buchstaben ihre Verbindungen weiter verstärken. Da AB bereits über 0.66 startet, kann E aufgrund des Assoziationslimits nur unter 0.33 enden.

Alle Buchstaben der *Kategorie 3* beginnen mit einer Assoziationsstärke von 0, da ein neuer Prototyp erlernt werden muss (Wiedererkennungshypothese). Da Buchstabe K bereits in Phase 1 vorkam, erhält er eine geringere Salienz (Habituationshypothese) und wird schlechter erlernt als die beiden anderen Buchstaben der 3. Kategorie. Dies ist eine nahezu parameterunabhängige Aussage. Jegliche, noch so kleine Benachteiligung von K führt dazu, dass der Anteil des 1. Teststimulus aus der 3. Kategorie eine Assoziationsstärke größer als 0.33 erhält, da die kleinere Gewichtsstärke als 0.33 von K noch „Spielraum“ lässt.

Die *Größe* der Differenz zu 0.33 hängt natürlich von der Schwächung von K und damit von der Modellierung des Salienzparameters ab. Wie groß diese Differenz aber sein muss, um den Entscheidungsprozess für EJO zu beeinflussen, hängt vom Modell der Entscheidung ab. Bei einer Entscheidung, die strikt linear mit der Assoziationsstärke wächst, führt jegliche Schwächung von K zu dem erhaltenen Ergebnis der Bevorzugung von Kategorie 3, da der Anteil am Teststimulus aus Kategorie 3 eine stärkere Verbindung als der Anteil aus den anderen beiden Kategorien aufbauen kann (vgl. Abbildung 2.9).

³⁰ Da *alle* Icons eines Buchstabens bis auf wenige zufällige Vertauschungen immer gleichzeitig präsentiert wurden, wird in der folgenden Erklärung nur auf Buchstaben verwiesen.

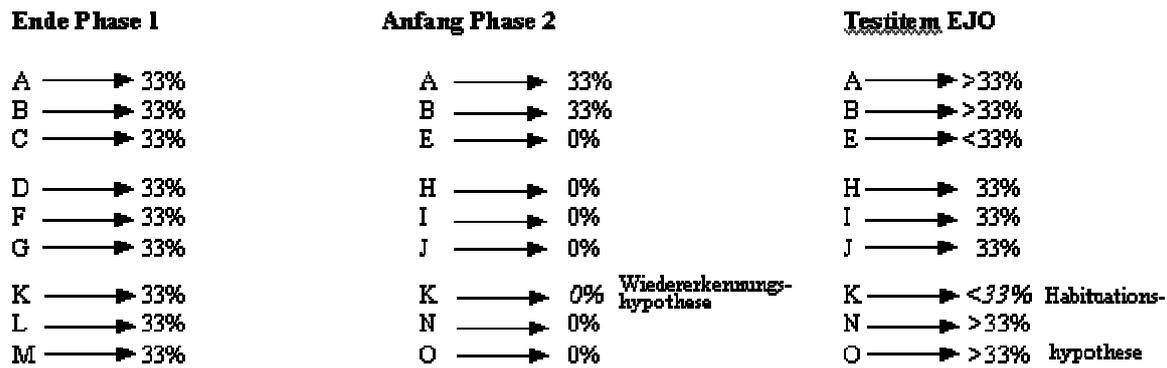


Abb. 2.9: Stärke der Verbindungen zwischen den einzelnen Buchstaben und den Prototypen in Prozent der möglichen Gesamtverbindungsstärke. Angegeben sind die Verbindungsstärken für Gruppe 1 am Ende von Phase 1, Anfang von Phase 2 und bei Teststimulus EJO. Die Salienz ist nicht in die Abbildung aufgenommen.

Für Entscheidungen, welche die Assoziationsstärken nicht fehlerfrei bestimmen können, ist eine entsprechend stärkere Benachteiligung von K notwendig. In der Computersimulation wurde ein linearer Entscheidungsprozess implementiert, der die am höchsten aktivierte Ausgabereinheit selektiert. Andere Entscheidungsmodelle (z.B. Jones, Wills & McLaren, 1998) hätten weitere Parameterannahmen nötig gemacht, ohne dabei die ordinale Vorhersage des Netzwerkes wesentlich zu verändern. Um das Modell nicht unnötig aufzublähen, wurde auf die Modellierung verschiedener Entscheidungsmodelle daher verzichtet. Für eine genauere Beschreibung des Modells und der Simulation, siehe Anhang 9, "Konnektionistisches Delta-Modell II". Es zeigte sich, dass das Netzwerkmodell für weite Parameterbereiche die Präferenz von Kategorie 3 vorhersagen konnte. In den meisten Simulationen lösten nahezu alle 24 simulierten Versuchspersonen die Aufgabe. Das Netzwerk stellte somit die perfekte, von Nichts abgelenkte, konzentrierte Idealversuchsperson dar.

Ein Modell, welches genau die gleiche Datenkonstellation erzeugt wie sie empirisch erhalten wurde, müsste auch die „Fehler“ im Sinne der Abweichung vom Idealverhalten modellieren. Einem einfachen konnektionistischen Netzwerk die Fülle an kognitiven Vorgängen aufzubürden, dürfte vermessen sein. Mir erscheint es problematisch mit einer Parametersuche empirische Daten, welche die beobachtbaren Endprodukte von verschiedenen Prozessen darstellen, exakt reproduzieren zu wollen und damit eine Exaktheit der Modellierung vorzutäuschen, die nicht gegeben ist. Es wurde daher nicht versucht, mittels Parameteranpassung die *exakte* Anzahl an Kategorienzuordnungen zu reproduzieren, da diese Anzahl als empirisch zufällig und vermutlich in einem erneuten Versuch in der exakten Anzahl als auch nicht empirisch replizierbar angesehen wurde. Aus diesem Grund sollte das Netzwerk nicht dasselbe Kategorienverhältnis wie die empirische Beobachtung erzeugen,

sondern für weite Parameterbereiche dasselbe qualitative Verhalten zeigen. Die Lösung sollte für weite Parameterbereiche gefunden werden, da wenn ein Netzwerk in weiten Parameterbereichen qualitativ dasselbe Phänomen zeigt, der Schluss plausibler ist, dass das Phänomen nicht nur zufällig durch das Netzwerk beschrieben ist, sondern eine inhärente Eigenschaft des Modells darstellt. Außerdem sollte, da verschiedene Individuen vermutlich unterschiedliche „Parameterwerte“ aufweisen, leichte Parameterveränderungen qualitativ kein anderes Verhalten bewirken. Auch für den Fall, dass doch die quantitative Nähe zur empirischen Beobachtung eine Rolle spielt, hat eine Lösung über einen weiten Parameterbereich eine größere Wahrscheinlichkeit die Daten exakt widerzuspiegeln.

In Abbildung 2.10 sind jeweils 24 Simulationen durch einen Kreis repräsentiert. Gefüllte Datenpunkte repräsentieren Simulationsdurchgänge, in denen die Zuweisungshäufigkeit zu einer Kategorie signifikant erhöht war. Die „Schwellen-Anzahl“ wurde jeweils über einen an die Löseranzahl angepassten Binomialtest errechnet. Schwarze Datenpunkte bedeuten, dass die Mehrzahl der Zuordnungen des Stimulus EJO zu Kategorie 1 erfolgten, blaue, dass eine Präferenz für Kategorie 2 vorlag und rote stellen schließlich eine Präferenz für Kategorie 3 dar. Simulationen, bei denen die simulierten Versuchspersonen keine Kategorie bevorzugten oder weniger als 3 Löser simuliert wurden, sind grün eingefärbt.

Wie aus Abbildung 2.10 ersichtlich, zeigt das Netzwerk für weite Parameterbereiche ein ähnliches Verhalten wie die Versuchspersonen, symbolisiert durch gefüllte rote Kreise. Damit sich der Blocking-Effekt zeigen kann, müssen die Kategorien hinreichend gut gelernt sein. Aus diesem Grund werden vorwiegend Parameterkombinationen mit hohen Lern- und Salienswerten signifikant. 681 simulierte Versuchspersonen präferierten signifikant die 3. Kategorie über die 1. und 2. Kategorie. Die entsprechende Zahl betrug für die 2. Kategorie 152 und nur 14 simulierte Versuchspersonen bevorzugten die 1. Kategorie signifikant. Der Erwartungswert an signifikanten Ergebnissen pro Kategorie war die Anzahl an Simulationen multipliziert mit dem Signifikanzniveau geteilt durch die Anzahl an Kategorien: $2550 \cdot 0.05 / 3 = 42.5$. In Anhang 10a sind die Simulationen der Kategorienwahl für den „divisionsbegrenzten“ Algorithmus (vgl. Abschnitt 1.5) grafisch repräsentiert. Es zeigen sich kleinere „erfolgreiche“ Parameterbereiche.

Um einschätzen zu können, wie viele Löser jeweils bei den einzelnen Simulationsdurchläufen auftraten, sind in Abbildung 2.11 die Datenpunkte zusätzlich gegen die Gesamtanzahl an Lösern aufgetragen. Wie man sieht, produziert das Netzwerk in den meisten Fälle ausschließlich Löser. Aber auch bei Simulationsläufen, in denen weniger Löser auftreten, kommt es zu signifikanter Bevorzugung der Kategorie 3.

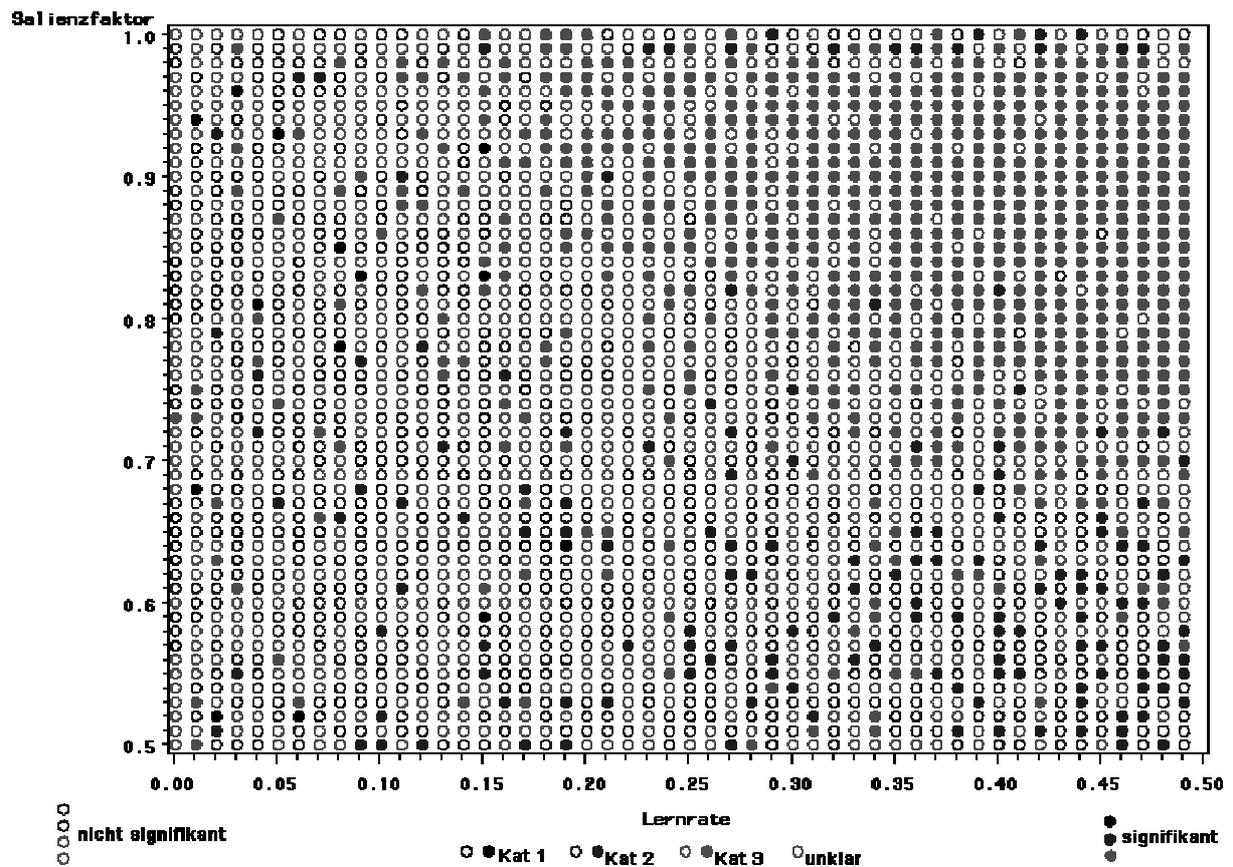


Abb. 2.10. Jeder Punkt repräsentiert 24 simulierte „Versuchspersonen“. Gefüllte Punkte zeigen eine signifikante Bevorzugung einer Kategorie für Stimulus EJO über die 24 „Versuchspersonen“ an. Die 3 Kategorien werden durch schwarz, blau und rot repräsentiert. Grün symbolisiert Simulationsdurchläufe in denen entweder weniger als 3 Löser auftraten oder keine Kategorien signifikant bevorzugt wurden.

Die folgende Abbildung 2.12 zeigt die Stärke der Bevorzugung der Kategorie 3. Es sind die Kategorienpräferenzen gegen die Anzahl an Zuordnungen zu Kategorie 3 aufgetragen. Wie aus Abbildung 2.12 deutlich wird, wählten in keinem Simulationsdurchlauf alle 24 simulierten Versuchspersonen die Kategorie 3. Die Anzahl an Zuweisungen schwankte in einem weiten Bereich. Diese Unterschiede zwischen den simulierten „Versuchspersonen“ in einem Versuchsdurchlauf beruhen darauf, dass die *unterschiedlichen* DMDX-Präsentationsdateien aus der empirischen Studie übernommen wurden.

Daher wurden die Stimuli für unterschiedliche simulierte „Versuchspersonen“ in einer unterschiedlichen Reihenfolge und mit unterschiedlich vertauschten Icons dargeboten. Aus einem ähnlichen Grund führte ein Netzwerk, welches dem Rumelhart und Zipser (1986) Modell folgte (Anhang 9, Konnektionistisches Rumelhart & Zipser-Modell I), bei hohen Lernraten zu einer Präferenz der 1. und 2. Kategorie gegenüber der 3. Kategorie.

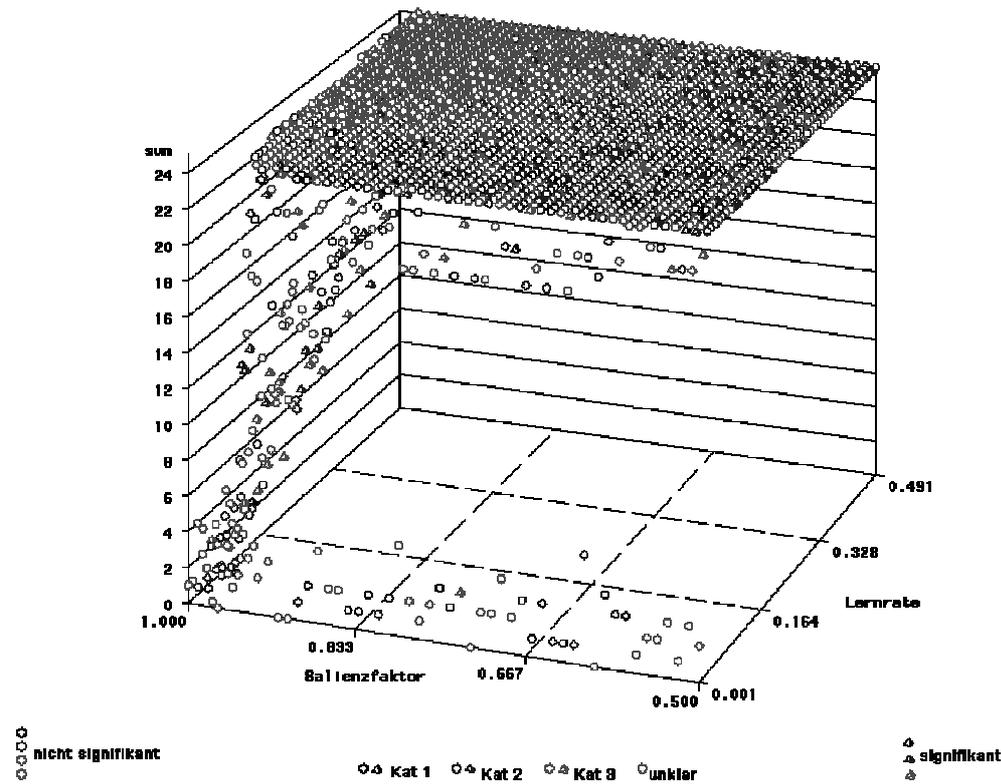


Abb. 2.11: Verteilung der Kategorienwahl abhängig von der Anzahl an Lösern pro Simulationsdurchlauf (sum), den Saliencyfaktor und die Lernrate.

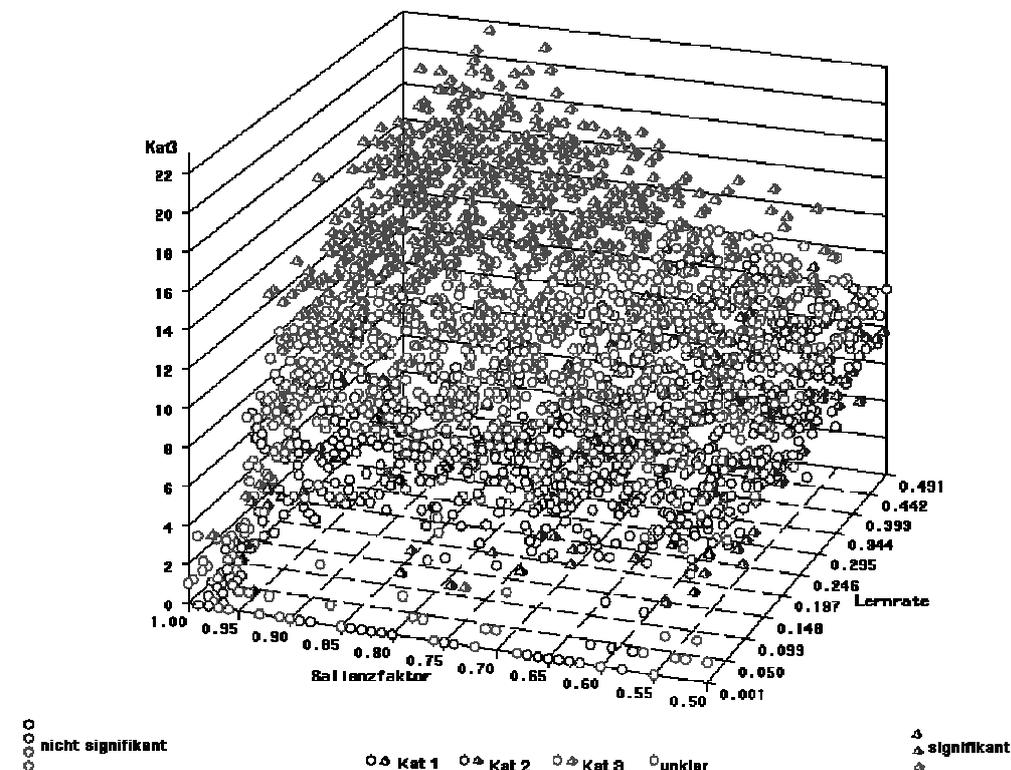


Abb. 2.12: Bevorzugung der Kategorien aufgetragen in Farbe im durch die Anzahl der Wahlen von Kategorie 3 (Kat3), den Saliencyfaktor und die Lernrate aufgespannten Raum.

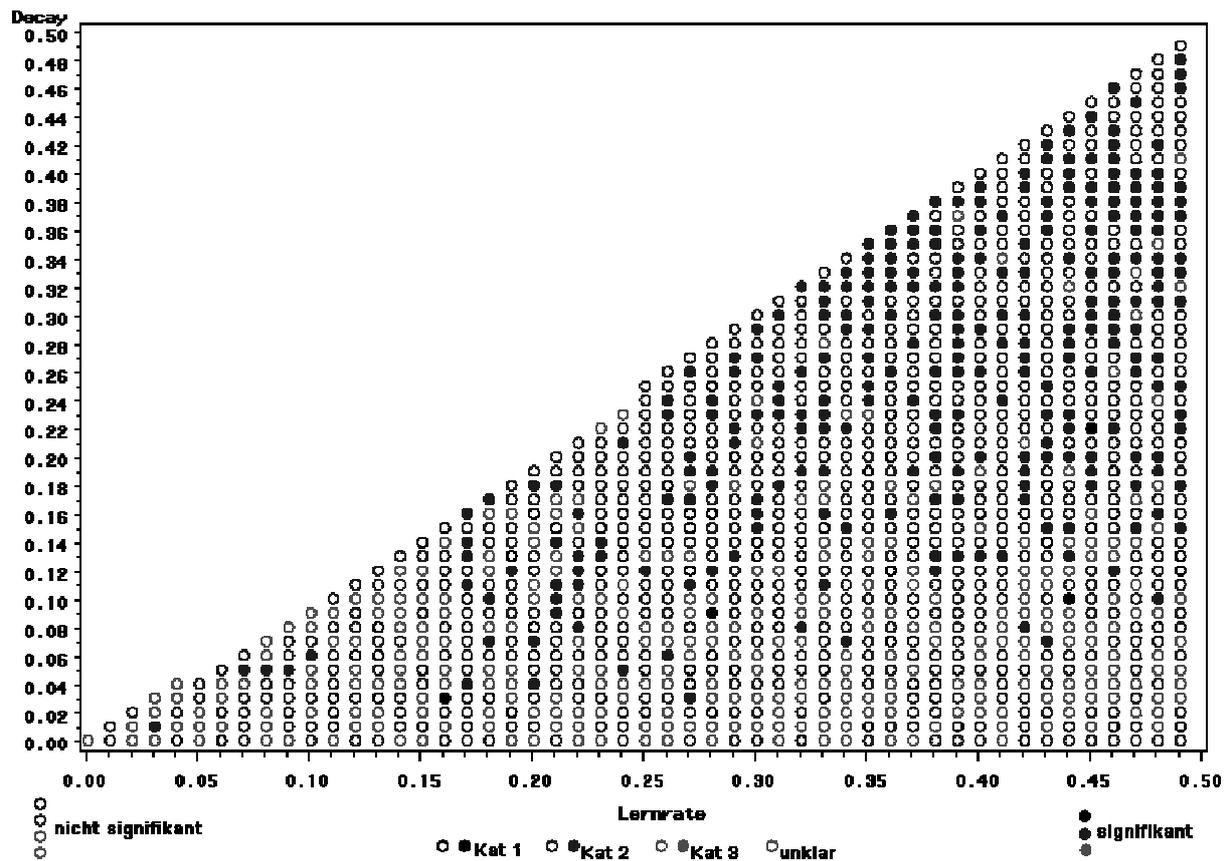


Abb. 2.13: Verteilung der Kategorienwahlen in Simulationen nach dem Modell von Rumelhart & Zipser (1985).

Bei 149 simulierten Versuchspersonen zeigte ein Chi-Quadrat³¹ Test eine signifikante Abweichung von der Gleichverteilung. Der Erwartungswert unter Annahme der Gleichverteilung hierfür beträgt $1255 \cdot 0.05 = 62.75$. Die Anzahl an 149 signifikanten Abweichungen wich damit signifikant von der Anzahl an Abweichungen ab, die an zufälliger Signifikanz bei 1255 Tests zu erwarten ist. ($B_{1255;0.95}(1106) < 0.001$). Abbildung 2.13 zeigt die Verteilung der Kategorienwahlen.

Die erhöhte Anzahl „signifikanter“ Simulationsdurchläufe ist darauf zurückzuführen, dass im Experiment zufällig Icons zwischen den einzelnen Prototypen ausgetauscht wurden. Aufgrund zufälliger Schwankungen wurden dabei etwas häufiger Icons des Buchstabens O (aus Sicht der Gruppe 1) ausgetauscht. Dies hat zwar nur eine minimale Auswirkung auf die Gewichte, da der Entscheidungsprozess aber ohne Fehler modelliert wurde, führte auch der kleinste Unterschied zu einer Präferenz der ersten 2 Kategorien. Da die Präsentationsfiles für jede Parameterkombination wiederholt wurden, wirkte sich diese leichte Asymmetrie systematisch auf die Kategorienpräferenz aus. Diese Auswirkung wird im Delta-Netzwerk

³¹ Dies dient nur einer groben Abschätzung, da die Erwartungswerte häufig kleiner als 5 waren.

durch die Abnahme der Salienz verhindert³². Ebenso vermindert die geringe Differenz zur erreichbaren Ausgabe-Aktivität, welche die Hemmung nicht-aktiver Gewichte bestimmt, den Effekt der Präsentationsreihenfolge auf das Delta-Netzwerk. Obwohl die Ergebnisse des Rumelhart und Zipser Modells den empirischen Daten widersprachen, wurde untersucht, ob das Modell mit leichten Veränderungen nicht doch die Versuchspersonendaten erklären könnte. Im Modell wurde daraufhin eine Fehlerüberlagerung der Entscheidung durch die ratio rule:

$$p(\text{„Kategorie } x\text{“}) = \frac{e^{o_x \Theta}}{\sum_{j=1}^v e^{o_j \Theta}}$$

(z.B. Gluck & Bower, 1988a) implementiert.

Der ratio rule zufolge hängt die Wahrscheinlichkeit einer Zuordnung zu einer Kategorie x von dem Verhältnis der Aktivierungsstärke des Ausgabeknoten o_x zu den anderen Ausgabeknoten o_j ab. Der Parameter Θ spiegelt dabei die Stärke des „zufälligen Rauschens“ wider. Bei hohen Werten erfolgt eine nahezu deterministische Entscheidung für den am stärksten aktivierten Ausgabeknoten. Für Werte nahe 0 besteht nahezu kein Zusammenhang mehr zwischen den Aktivierungsstärken und der Wahrscheinlichkeit. Nimmt Θ den Wert 0 an, wird jede Kategorie mit gleicher Wahrscheinlichkeit, unabhängig von den Aktivierungsstärken gewählt. Für das Rumelhart und Zipser Netzwerk sollte ein Rauschen gewählt werden, das minimale Unterschiede in den Aktivierungsstärken, entstanden durch eine unterschiedliche Anzahl vertauschter Icons der einzelnen Kategorien, ausgleicht. Gleichzeitig sollte aber nicht durch das Rauschen selbst die Gleichverteilung erzeugt werden. Daher wurde Θ auf 20 gesetzt. Dieser Wert liegt weit über den von Gluck und Bower (1988a) verwendeten Werten.

In 1255 Simulationen wurde die Lernrate im Bereich 0.001-0.5 in Schrittweiten von 0.01 variiert. Der Decay-Parameter (vgl. Anhang 10, Konnektionistisches Rumelhart & Zipser-Modell I) wurde ebenfalls in 0.01 Schritten von 0.001 bis zur jeweiligen Lernrate variiert. Die Berechnung des Chi-Quadrat Wertes für die einzelnen Simulationen ergab 47 signifikante Abweichungen von der Gleichverteilung.

Der Erwartungswert an zufällig signifikanten Ergebnissen lag erneut bei 62.75. Ein anschließender Binomial-Test ergab eine signifikant unterdurchschnittliche Anzahl an signifikanten Durchgängen ($B_{1255;0.05}(47) < 0.05$).

³² Die abnehmende Salienz verhinderte im Delta-Netzwerk, dass ein im Laufe der Präsentationen erreichter „optimaler“ Zustand nicht mehr aufgrund seltener Iconvertauschungen verlassen wurde.

Das signifikante Ergebnis verdeutlicht erneut, dass die signifikanten Durchläufe eine systematische Komponente des Modells darstellen und die Annahme der zufälligen Signifikanzen, begründet allein durch eine Vielzahl an Testungen, verletzt ist. Die „erfolgreichen“ Parameterkombinationen bildeten keinen zusammenhängenden Bereich wie aus der Abbildung in Anhang 10b) zu ersehen ist. Das Delta-Modell zeigte bei dieser Entscheidungsregel keinen qualitativen Unterschied zur Modellierung ohne ratio rule (vgl. Abbildung in Anhang 10c). Simulationen des Rumelhart & Zipser Modells, welche ebenfalls die beiden für das Delta-Netzwerk getroffenen Annahmen implementierten (vgl. Anhang 7, „Konnektionistisches Rumelhart & Zipser-Modell II“) zeigten keine Präferenz für die 3. Kategorie (siehe Anhang 10d). Wie man in Anhang 10d an den signifikanten Bevorzungen einer Kategorie bei einem hohen im Gegensatz zu bei einem niedrigem Salienzfaktor erkennt, reicht die Abnahme des Salienzfaktors bereits aus, um zufällige minimale Schwankungen nicht als einen systematischen Effekt des Modells überzubewerten. Simulationen mit der Oja-Regel (vgl. Abschnitt 1.2.7, Kompetitives Lernen) brachten keine qualitativen Veränderungen.

Ob die numerische wenn auch eindeutig nicht signifikante Bevorzugung der Kategorie 1 verglichen mit Kategorie 2 für Stimulus EJO in den empirischen Daten erklärt werden muss, hängt von der zugrundeliegenden Fehlervorstellung ab. Die Bevorzugung der 1. Kategorie gegenüber der 2. Kategorie durch eine Person mehr kann durch die Annahme erklärt werden, dass Versuchspersonen, die nicht den Blocking-Effekt durch Zuweisung des Stimulus EJO zu Kategorie 3 zeigten, andere Strategien verfolgten und die Ähnlichkeit nicht anhand der gemeinsamen Icons beurteilten. In diesem Fall wären assoziative Erklärungen für die Versuchspersonen ungeeignet, welche keinen Blocking-Effekt zeigten. Selbst unter der Annahme, dass Versuchspersonen, welche EJO nicht in Kategorie 3 einordneten, prinzipiell die gleichen assoziativen Prozesse durchlaufen haben, wäre das empirische Ergebnis mit der Vorstellung über den Blockierungseffekt vereinbar. In diesem Fall könnten die Entscheidungen durch ein fehlerhaftes „Ablesen“ der Aktivierung für die Repräsentanten der einzelnen Kategorien verfälscht worden sein. Für diese Erklärung müsste lediglich der „Fehleranteil“ an der Entscheidung als relativ groß im Verhältnis zu den Unterschieden in der Aktivierungsstärke zwischen Kategorie 1 und Kategorie 2 oder als nicht linear additiv angenommen werden.

Nachtrag: Diskussion der Reaktionszeit auf Stimulus AHK

Die empirisch bestätigte Vorhersage des ursprünglichen Modells, dass auf den Teststimulus AHK langsamer reagiert werden sollte als auf ein gewöhnliches Item, lässt sich auch mit der gerade ausgeführten Modellvorstellung vereinbaren. Da bei einem gewöhnlichen Testitem nur wenige Icons vertauscht wurden, sollte auch höchstens ein geringe Tendenz entstehen den Stimulus unterschiedlichen Kategorien zuzuweisen. Für Testitem EJO sollte nur dann eine stärkere Tendenz den Stimulus unterschiedlichen Kategorien zuzuordnen und damit ein größerer Konflikt auftreten, wenn die Blockierung und Salienz Wirkung nicht vollständig wirksam waren. Ein sehr starker Effekt, verglichen mit EJO sollte bei Teststimulus AHK auftreten, da hier immer noch die Buchstaben aus Kategorie 1 und 2 bei einer sehr ähnlichen Verbindungsstärke von 33% enden sollten. Die etwas stärkere assoziative Verbindung von A zu Kategorie 1 als von H zu Kategorie 2 sollte durch die schwächere Salienz von A ausgeglichen werden.

Kategorienzuordnung von Stimulus AHK

Da im Stimulus AHK sowohl A wie auch K in beiden Phasen präsentiert wurden und dies keinen signifikanten Einfluss auf die Zuordnung von AHK ausübte, kann davon ausgegangen werden, dass die Anzahl an Präsentationen per se über ein bestimmtes Limit hinaus die Assoziation nicht mehr weiter steigert. Dies steht in klarem Widerspruch zu hebbischen Lerntheorien³³ ohne Beschränkung der maximalen Verknüpfungsstärke. Eine direkte Anwendung des Delta-Modells ohne Salienzparameter (Habitationshypothese) und Annahmen über die Neubildung eines Prototyps (Wiedererkennungshypothese) würde folgendes für Gruppe 1 (Gruppe 2 und 3 analog) vorhersagen:

1. K sollte am stärksten aktiviert sein. Der Grund hierfür ist, dass K in Phase 2, aufgrund des relativ großen Fehlers durch N und O, seine Assoziationsstärke aufbauen und gleichzeitig aus Phase 1 bereits eine relativ hohe Assoziationsstärke besitzen sollte.
2. H-Icons sollten, da sie nur in Phase 2 auftraten, die kleinste Verbindungsstärke erreichen.
3. A sollte eine mittlere Verbindungsstärke erreichen, da A über die 2 Phasen hinweg häufiger auftrat als H, aber in der 2. Phase die Vorhersagefehler relativ klein waren und somit nur eine kleine Stärkenveränderung als für K erfolgte.

³³ Allerdings nicht zu einem kompetitiven Netzwerk, da hier in gewisser Form auch eine Gewichtslimitierung durch den Decay-Parameter, welcher mit der Gewichtsstärke multipliziert wird, erfolgt. Allerdings ist diese Gewichtslimitierung, im Gegensatz zur Limitierung im Delta-Modell, unabhängig von anderen gleichzeitig vorhandenen Informationen in Form von Reizen.

Der Chi-Quadrat Test zeigt zwar keinen signifikanten Unterschied in den empirischen Daten zwischen den Zuordnungen von AHK zu den einzelnen Kategorien, aber numerisch widersprechen die erhaltenen Daten diesen Vorhersagen. Unter den oben getroffenen Annahmen zur Salienz und Wiedererkennung und der zusätzlichen Annahme, dass der Salienzfaktor einen Einfluss auf die Entscheidung ausübt (vgl. Abschnitt 1.2.1, Problem (e) des Rescorla-Wagner Modells aus Sicht von Miller, Barnet und Grahame, 1995), lassen sich die numerischen Verhältnisse erklären:

Da A und K am häufigsten präsentiert wurden, ist ihre Salienz am Ende der 2. Phasen geringer als H (Annahme 2, Habituationshypothese). K hat eine niedrigere Verbindungsstärke zu AHK als H, da K in Phase 2 auch mit einer Verbindungsstärke von 0 anfangen musste (Wiedererkennungshypothese) aber gleichzeitig eine niedrigere Salienz als H besaß (Habituationshypothese). Die Verbindungsstärke von K beträgt am Ende von Phase 2, aufgrund der gleichen Ausgangsposition wie die konkurrierenden Icons der Buchstaben NO (bzw. NJ für Gruppe 3) aber geringerer Salienz etwas unter 0.33. H endet, aufgrund gleichstarker „Mitstreiter“ (IJ, IE, IO für Gruppe 1, 2 und 3) etwa bei einer Verbindungsstärke von 0.33. Komplizierter ist der Fall für A. A endet etwas über 0.33, abhängig davon, wie sehr seine niedrigere Salienz am Anfang der Phase 2 sein weiteres Anwachsen verhinderte. Unter der Annahme eines Performanzmodells, welches die Entscheidung sowohl von der Salienz als auch von der Verbindungsstärke abhängig macht (vgl. z.B. Mackintosh, 1975; Kruschke & Johansen, 1999), kann durch geeignete Wahl des Salienzabnahmeparameters, die Bevorzugung von H gegenüber A erklärt werden. Dazu wird für die Kategorienwahl die Aktivität jedes Eingangneurons nicht nur mit der Gewichtsstärke sondern auch dem Salienzfaktor multipliziert. Da die qualitative Aussage aber von der Größe des Salienzparameters abhängt, muss, um die Einordnung des AHK Stimulus zu erklären, ein Parameter geschätzt werden. Die Zuordnung des Stimulus AHK wich in der empirischen Studie nicht signifikant von der Gleichverteilung ab. Daher beruht die Bevorzugung der 2. Kategorie und die Stärke dieser Bevorzugung für den Stimulus AHK auf theoretischen Spekulationen. Um zu überprüfen, für welche Parameterbereiche das Netzwerkmodell die empirischen Daten qualitativ widerspiegelt, wurde deshalb als Kriterium gewählt, dass AHK häufiger der Kategorie 2 als den beiden anderen Kategorien zugewiesen werden sollte. Die Simulationen zeigten, dass über einen weiten Parameterbereich sowohl signifikante Präferenzen für Kategorie 3 als Reaktion auf Teststimulus EJO als auch gleichzeitig Präferenzen für Kategorie 2 als Antwort auf Teststimulus AHK auftraten. Damit konnte das Netzwerk über einen weiten Parameterbereich die empirischen Daten qualitativ repro-

duzieren. Abbildung 2.14 zeigt die „erfolgreichen“ Simulationsdurchgänge. Auch die Anzahl an signifikanten Präferenzen für Kategorie 3 bei Teststimulus EJO, ohne Beachtung des Ergebnisses für AHK, nahm durch die Aufnahme des Salienzparameters in die Entscheidungsregel zu.

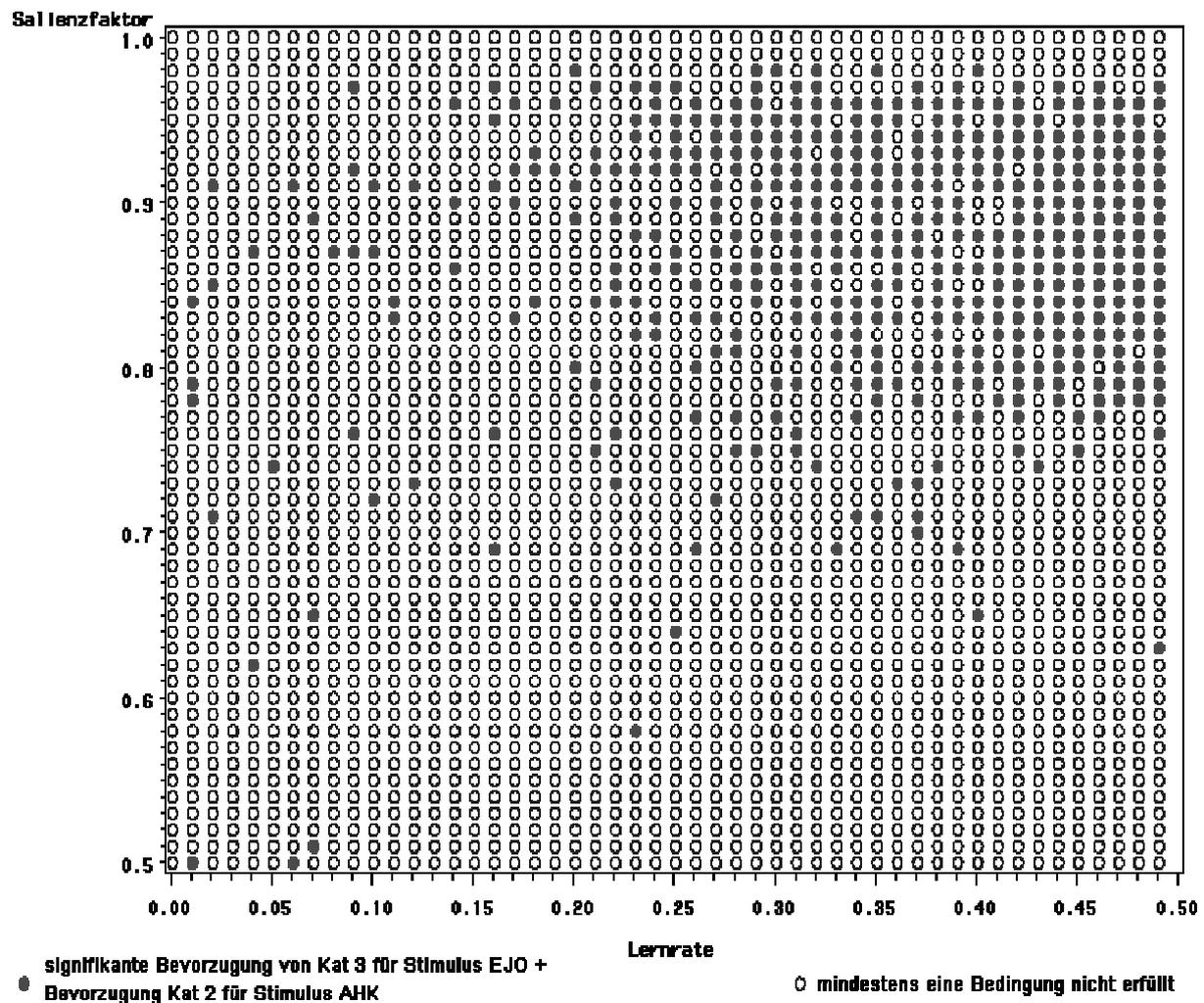


Abb. 2.14: Qualitative Erfüllung der empirischen Daten durch Netzwerksimulationen mit Delta-Algorithmus II und Salienzparameter in der Entscheidungsfunktion.

Da aber unklar ist, ob die empirische Bevorzugung der 2. Kategorie als Reaktion auf AHK ein systematisches Verhalten von Versuchspersonen oder „Zufallsrauschen“ darstellt, rechtfertigt eine Erklärung dieser „leichten“ Bevorzugung vermutlich nicht eine weitere Annahme.

Mit einem Netzwerkmodell konnten somit relativ parameterunabhängig die empirischen Daten bezüglich der Kategorienentscheidung qualitativ reproduziert werden. Gleichzeitig war der gewählte Algorithmus konsistent mit den Daten von Zwickel und Wills (2002).

Bekanntheitseinschätzung

Entgegen den Erwartungen zeigte sich kein signifikanter Unterschied in der Einschätzung des Bekanntheitsgrades von Icons aus unterschiedlichen Kategorien. Dass dennoch bei der Stimuluszuordnung konsistent eine Kategorie bevorzugt wurde, legt den Schluss nahe, dass die kurze Darbietungszeit während der Lernphase eine tiefere Verarbeitung der Stimuli in Form von expliziter Erinnerungsabspeicherung nicht ermöglichte. In der Literatur finden sich verschiedene Belege dafür, dass eine Dissoziation zwischen explizitem Wissen über Bekanntheit eines Reizes und impliziter Kategorisierung besteht. Normalerweise wird diese Dissoziation bei Menschen mit Gehirnverletzungen beobachtet (z.B. Knowlton & Squire, 1993). Knowlton und Squire (1993) oder auch Reed, Squire, Patalano, Smith und Jonides (1999) führen diese Dissoziation auf unterschiedliche Gehirnbereiche für explizites und implizites Lernen zurück. Unter Umständen verhinderte die kurze Darbietungszeit und der Druck, sofort den nächsten Stimulus hinsichtlich der Kategorie zu verarbeiten, diese „expliziten“ Prozesse auch bei gesunden Probanden.

Zaki und Nosofsky (2001) hingegen erklären die Befunde mittels *eines* Systems, welches durch die verschiedenen Aufgaben unterschiedlich stark belastet wird. Während bei Kategorisierungsaufgaben einzelne Merkmale bereits für die Kategorisierung suffizient sein können, muss für eine Wiedererkennung eine größere Anzahl an Merkmalen integriert werden. So kann eine Versuchsperson, welche bemerkt, dass Icons aus einer Kategorie eher runde Formen darstellen, einen neuen Stimulus, der ebenfalls runde Formen enthält, recht leicht derselben Kategorie zuordnen. Für eine Wiedererkennung wäre hingegen eine Unterscheidung zwischen den einzelnen möglichen Icons einer Kategorie notwendig. Die größere Anforderung bei der Wiedererkennungsaufgabe, die eine Unterscheidung zwischen ähnlichen Stimuli einer Kategorie erfordert, während für eine Kategorisierungsaufgabe eine größere Ähnlichkeitseinstufung ausreicht, machen die Erklärung, dass unter Zeitdruck nur die Kategorisierung erlernt wird, plausibel.

Eine alternative Erklärung ist, dass die kurzen Reaktionszeiten in Experiment 1 zu einer eher konfiguralen Analyse der Stimuli (Smith & Kehler Nelson, 1984) führt. In diesem Fall bestände eine geringere Ähnlichkeit zwischen den 4 für die Bekanntheitseinschätzung dargestellten einzelnen Icons und den vorher gesehenen konfiguralen Stimuli.

Alle 3 Erklärungen sind mit den empirischen Befunden und der Blocking-Modellvorstellung vereinbar.

Auskünfte der Teilnehmer

Die Löser schienen seltener den Farbeindruck als Klassifizierungsgrundlage zu nutzen als die Nicht-Löser. Gleichzeitig schienen die Nicht-Löser eher mehrere Icons zu integrieren. Beide Strategien scheinen daher die Lösung des Klassifizierungsproblems zu erschweren. Gleichzeitig konnten aber auch viele Löser mit diesen Strategien das Problem lösen, was darauf schließen lässt, dass auch mit diesen Klassifizierungsgrundlagen eine Lösung möglich war, diese also nur erschwert wurde.

Von den 6 Versuchspersonen, welche Stimulus EJO Kategorie 3 zuordneten *und* deren Aussage bezüglich der Anzahl verwendeter Icons von beiden Ratern gleich interpretiert wurden, schien allerdings nur eine Versuchsperson eine Einzel-Icon-Strategie zu verwenden. Dies lag signifikant über der Wahrscheinlichkeit von 0.25 für das Auftreten der Mehr-Iconstrategie bei allen anderen Lösern ($B_{6;0.75}(1) < 0.01$). Der höhere Anteil an Mehr-Iconlösern bei den Versuchspersonen, welche durch eine Zuweisung des Teststimulus EJO zu Kategorie 3 einen Blocking-Effekt zeigten, deutet darauf hin, dass für die erfolgreiche Beobachtung des Blockingphänomens eine Mehr-Iconstrategie förderlich ist. Welche Faktoren aber diese Mehr-Iconstrategie begünstigen, kann aufgrund der vorliegenden Daten nicht erschlossen werden.

Die Beschreibungen nach Farben zu kategorisieren und mehrere Icons zu beachten bzw. nach Form und nur ein Icon zu beachten hingen sehr stark zusammen³⁴. In die Auswertung gingen nur Versuchspersonendaten ein, für die eine Übereinstimmung zwischen den beiden Ratern bestand. Es zeigte sich ein signifikant erhöhter Zusammenhang zwischen der Farbkategorisierung und dem Beachten mehrerer Icons sowie der Formkategorisierung und der Beachtung nur eines Icons bei den Lösern (Fisher-Exact-Test, $p < 0.05$) nicht aber bei den Nicht-Lösern (Fisher-Exact-Test, $p > 0.10$). Die Insignifikanz bei den Nicht-Lösern kam durch die starke Besetzungszahl nur einer Zelle zustande. Obwohl 16 der 17 aufgrund der Raterübereinstimmung auswertbaren Versuchspersonendaten eine Mehr-Iconstrategie in Kombination mit Farbbeurteilungen zu benutzen schienen, fehlte für den Nachweis des Zusammenhangs das gemeinsame Auftreten von Form- und Einzelkategorisierung.

³⁴ Da dies eine post-hoc datengetriebene Untersuchung darstellt, wird der statistische Test erst in der Diskussion berichtet. Auf die Interpretation des Ergebnis als Anzeichen einer Ressourcenlimitierung wird im Rahmen der Diskussion von Experiment 2 eingegangen.

Eventuell stellt die Einzel-Icon Interpretation bei den „Form-Lösern“ aber auch ein Artefakt (vgl. Diskussion von Experiment 2 zur Anzahl verwendeter Icons) der Schwierigkeiten, welche die Versuchspersonen damit hatten ihre visuellen Eindrücke der komplexen Formen schriftlich zu fixieren, dar. Die vermutete größere Komplexität bei Verwendung des Form-Kriteriums als bei der Verwendung des Farbkriteriums könnte aufgrund der kognitiven Überlastung dazu führen, dass keine *Erinnerungsspuren* für mehrere Icons gebildet werden konnten.

Dass die Anzahl an Versuchspersonen, deren Berichte die Verwendung einer Restkategorie nahe legten, sich nicht deutlich zwischen Lösern und Nicht-Lösern unterschied kann so gedeutet werden, dass es auch den Nicht-Löser gelang eine Kategorienvorstellung zu entwickeln, welche erlaubte die Stimuli in 3 Kategorien einzuteilen. Ihren „Status“ als Nicht-Löser verdanken sie dieser Interpretation zu Folge lediglich der Tatsache, dass sie eine andere Einteilung als nach der Anzahl gemeinsamer Icons vornahmen.

Fraglich ist auch, ob die Versuchspersonen überhaupt in der Lage waren, ihre Entscheidungsstrategie zu beschreiben. Eventuell führte die kurze Darbietungszeit und die komplexe Aufgabe dazu, dass eher implizite Mechanismen die Entscheidungen leiteten. Dieser Vorstellung zufolge versuchten die Versuchspersonen erst im Nachhinein eine „rationale“, verbalisierbare Erklärung zu generieren. Berry und Broadbent (1984) konnten zeigen, dass Versuchspersonen, welche ein durch eine mathematische Gleichung definiertes Problem lösen konnten, ihre Lösungsstrategie nicht besser verbalisieren konnten als schlechtere Problemlöser. Dass dies kein Artefakt einer unterschiedlichen Sensitivität der Tests darstellt, konnten die Autoren dadurch zeigen, dass die Problemlösefähigkeit sogar signifikant negativ mit der späteren verbalen Beschreibung korrelierte. Nisbett und Wilson kommen bei der Zusammenfassung verschiedener Studien zu dem Schluss: „In studies where the data are available, no association is found between degree of verbal report change and degree of behavior change in experimental groups.“ (1977, p. 235). Nisbett und Wilson gehen davon aus, dass Versuchspersonen keinen direkten Zugang zu ihrem gezeigten Verhalten besitzen, sondern vielmehr versuchen plausible Gründe für ihr Verhalten im Nachhinein zu konstruieren. Wenn die Aufgabe nicht durch ein typisches Ursache-Wirkungsgefüge gelöst wurde, so resultiert hieraus eine Diskrepanz zwischen den tatsächlichen und den genannten Gründen. Diese Mechanismen sind nach Nisbett und Wilson relativ unabhängig von der jeweiligen Aufgabe.

2.2.4 Zusammenfassung Experiment 1

Mit den oben diskutierten Annahmen zeigte Experiment 1 damit auch bei Klassifizierung ohne Rückmeldung den Blocking-Effekt. Die dazu notwendigen post-hoc Annahmen schränken die Validität der Erklärung ein. Durch eine simple Replikation der Studie würde zwar oberflächlich das Kriterium der Bestätigung einer Vorhersage durch ein Experiment vermutlich erfüllt - von einem ähnlichen Ergebnis wäre aufgrund des erhaltenen Signifikanzniveaus bei einer Replikationsstudie auszugehen - doch würden damit keine unabhängigen Belege für die oben getroffenen Annahmen erhalten werden. Im Rahmen dieser Diplomarbeit sollte deshalb untersucht werden, inwieweit der Effekt auch bei unterschiedlichen Präsentationszeiten auftritt. Es wurde erwartet, dass der Blocking-Effekt bei kürzeren Zeiten stärker ausgeprägt sein sollte, da bei ausreichender Präsentationszeit die assoziativen Prozesse um andere kognitive Verarbeitungsprozesse ergänzt werden können und eventuell vollständiges Memorieren aller Icons erfolgen könnte. Diese Erklärung reduziert den Blocking-Effekt aber nicht auf eine simple Reduktion der Informationen unter Zeitdruck, da die Reduktion nicht zufällig über die Icons verteilt, sondern abhängig von bereits erworbenem Wissen erfolgt. Bei längeren Darbietungszeiten und den damit verbundenen Möglichkeiten über die Aufgabe zu reflektieren wäre eine stärkere Verwendung von auf einzelnen Icons beruhenden Regeln zu erwarten. Hingegen sollte die Integration der Informationen verschiedener Icons zurückgehen (vgl. Abschnitt 1.4; vgl. auch Ward, 1983; Smith & Kemler Nelson, 1984). In einem 2. Experiment wurden deshalb die Stimuli für jeweils 10 Sekunden präsentiert, was eine ausreichende Zeit für die Verarbeitung und Entscheidung sein sollte.

2.3 Experiment 2

Das Design entsprach dem von Experiment 1. Allerdings wurde die Stimuluspräsentationszeit auf 10 Sekunden heraufgesetzt, erneut beendet durch eine 1200ms dargebotene graue Maske. Die maximale Blockanzahl wurde auf 5 verkürzt, da in Experiment 1 nur 1 Löser mehr als 5 Blöcke benötigte. Die Reduktion der maximalen Blockanzahl sollte den Zeitaufwand auch für Nicht-Löser, trotz der längeren Präsentationszeit der Stimuli, auf ein vertretbares Maß beschränken. Bei der 2. Erhebungsstufe (vgl. Abschnitt Design von Experiment 1) wurde aus Versehen eine Präsentationsstruktur doppelt erhoben (Vpn 7 bei den Lösern sowie Vpn 5 und 7 bei den Nicht-Lösern sahen eine identische Präsentation).

2.3.1 Ergebnisse

17 der 36 Versuchspersonen lernten die Stimuli in der erwarteten Weise zu klassifizieren. Das Verhältnis von Lösern zu Nicht-Lösern war in beiden Erhebungsstufen in etwa gleich. In der 1. Erhebungsstufe lösten 11 Versuchspersonen die Klassifizierungsaufgabe und 13 teilten die Stimuli nicht nach der erwarteten Struktur auf. In der zweiten Erhebungsstufe war das Verhältnis 6 zu 6.

Anzahl benötigter Blöcke

Die Anzahl benötigter Blöcke unterschied sich nicht signifikant zwischen der 1. und 2. Phase ($t(16)=0.32$, $p>0.10$ siehe Abbildung 2.15). Auch korreliert die Anzahl benötigter Blöcke in der 1. und 2. Phase nur minimal (Pearson, $r=0.06$, $p>0.10$).

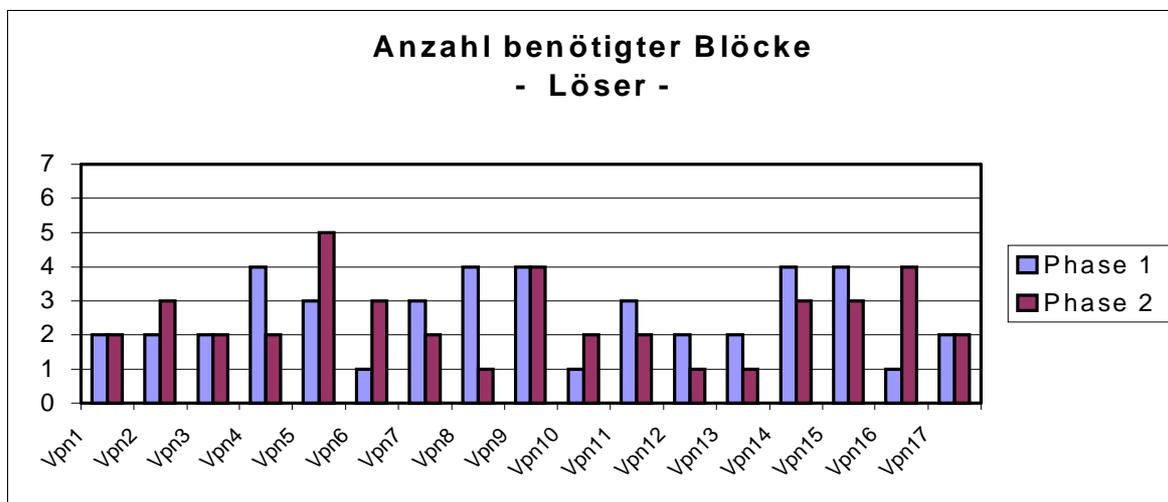


Abb. 2.15: Anzahl benötigter Blöcke.

Reaktionszeiten

Die mittlere Reaktionszeit für die einzelnen Blöcke blieb weit unter dem Limit von 11.2 Sekunden. Dennoch sind die Reaktionszeiten in den jeweils ersten Blöcken bzw. den beiden Testitems signifikant länger als in Experiment 1 (vgl. Tabelle 2.1).

Tabelle 2.1: Unterschied in der mittleren Reaktionszeit zwischen Experiment 1 und 2 für die jeweils ersten Blöcke der beiden Phasen und für die Testitems, ermittelt durch Mann-Whitney U-Test.

Vergleich	Phase 1, Block 1	Phase 2, Block 1	Testitem EJO	Testitem AHK
Exp. 1 – Exp. 2	Z(U)=-5.36, $p<0.001$	Z(U)=-6.08, $p<0.001$	Z(U)=-3.34, $p<0.001$	Z(U)=-3.46, $p<0.001$

In beiden Phasen von Experiment 2 war die mittlere Reaktionszeit des 1. Blockes signifikant höher als die des letzten Blockes ($t(16)=4.06$, $p<0.01$; $t(16)=3.50$, $p<0.01$). Es bestand hingegen nur ein marginal signifikanter Unterschied zwischen der mittleren Reaktionszeit des letzten Blockes der 1. Phase und dem 1. Block der 2. Phase ($t(16)=-1.89$, $p<0.10$).

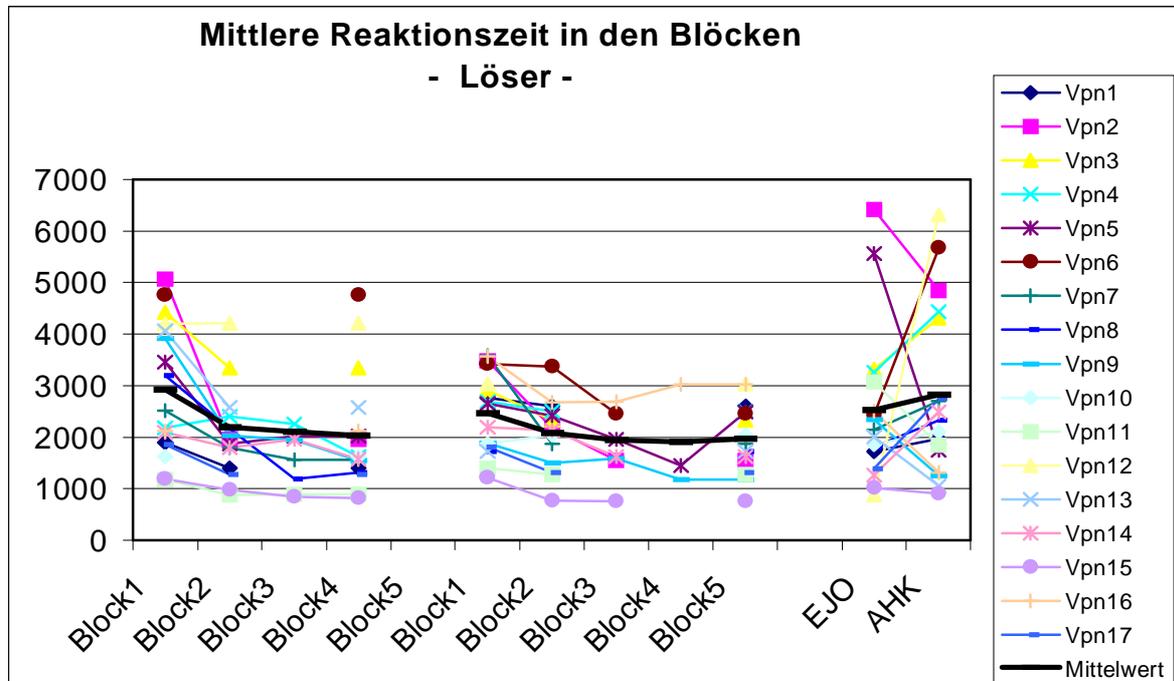


Abb. 2.16: Mittlere Reaktionszeiten über die Blöcke. Der Mittelwert ist in schwarz hervorgehoben.

Alle Löser blieben in ihrer Reaktionszeit auf die Testitems unterhalb des Zeitlimits der Experimentalstimuli. Ein t-Test zwischen den Reaktionszeiten für das individuell letzte Item der 2. Phase und der Reaktionszeit für das Testitem EJO war marginal signifikant ($t(16)=-1.83$, $p<0.10$, vgl. Abbildung 2.16). Die Reaktionszeit auf das Testitem AHK war signifikant länger als auf das Experimental-Item ($t(16)=-2.60$, $p<0.05$). Diese Ergebnisse stellen erneut kein Artefakt der Aufhebung des Zeitlimits dar, da keine der Versuchspersonen das in den Experimentalitems gesetzte Zeitlimit überschritt. Ein Mann-Whitney U-Test ergab keinen signifikanten Unterschied zwischen dem 1. und 2. Experiment bezüglich der Differenz zwischen dem letzten Experimental-Item und AHK ($Z(U)=-1.15$, $p>0.10$). Ein t-Test unter Annahme ungleicher Varianzen ergab hingegen einen signifikanten Unterschied ($t(17.03)=-2.16$, $p<0.05$).

Gesamt-Konsistenz

Die Gesamt-Konsistenz nahm über die Blöcke hinweg zu (vgl. Abbildung 2.17). Der Unterschied in Konsistenz zwischen den jeweiligen ersten Blöcken der 1. und 2. Phase war signifikant (Wilcoxon $Z=-2.40$, $p<0.05$). Nicht jedoch zwischen den letzten Blöcken beider Phasen (Wilcoxon $Z=-0.79$, $p>0.10$). Innerhalb der beiden Phasen nimmt die Konsistenz vom 1. zum individuell letzten Block ebenfalls signifikant zu (Wilcoxon $Z=-3.31$, $p<0.01$; $Z=-3.13$, $p<0.01$). Ein reliabler Unterschied ist auch zwischen dem letzten Block der 1. Phase und dem 1. Block der 2. Phase festzustellen (Wilcoxon $Z=-3.46$, $p<0.01$).

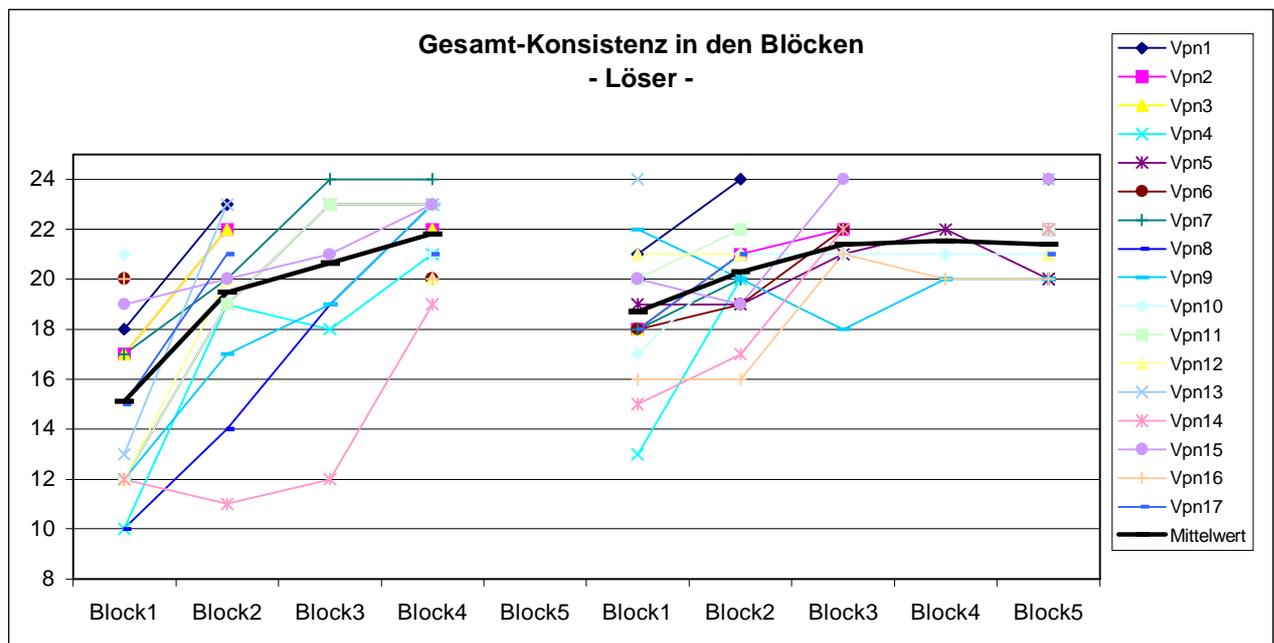


Abb. 2.17: Gesamt-Konsistenz über die Blöcke hinweg.

Tastenzuweisung

Auf den 1. Stimulus reagierten 13 Versuchspersonen mit Taste „1“, 2 mit Taste „2“, 1 mit Taste „3“ und 1 Versuchsperson reagierte überhaupt nicht in der gesetzten Zeit. Es ergab sich eine signifikante Abweichung von einer Gleichverteilung, die jede Taste als gleichwahrscheinliche Antwort auf den 1. Stimulus ansieht ($\chi^2(2)=16.63$, $p<0.001$; $B_{16;0.66}(3)<0.001$; $B_{16;0.33}(2)<0.10$; $B_{16;0.33}(1)<0.05$). In der 3. Phase weicht die Reaktion mit 3 mal Taste „1“, 10 mal Taste „2“ und 4 mal Taste „3“ nur marginal signifikant von einer Gleichverteilung ab ($\chi^2(2)=5.06$, $p<0.10$).

Konsistenz der Tastenzuweisung

Im letzten Block der 2. Phase verwendeten 9 Versuchspersonen die gleiche Taste wie am Ende von Phase 1 als Reaktion auf Stimuli der Kategorie 1 ($B_{17;0.66}(8)<0.10$), 6 dieselbe Taste

auf Stimuli der Kategorie 2 ($B_{17;0.33}(6) > 0.10$) und 7 auf Stimuli der Kategorie 3 ($B_{17;0.33}(3) > 0.10$). Damit unterschied sich die Konsistenz bezüglich Kategorie 1 zwischen Experiment 1 und 2 signifikant (Fisher-Exact-Test: $p < 0.05$). Die Zuordnung der Tasten zu den Kategorien am Ende der 1. und 2. Phase ist in Abbildung 2.18 widergegeben. Aus Abbildung 2.18 zusammen mit Abbildung 2.6 wird ersichtlich, dass deutlich mehr Versuchspersonen in Experiment 2, verglichen mit Experiment 1, die Tasten-zu-Kategorien-Zuordnung wechselten. Die Zuordnungsstruktur veränderte sich in Experiment 2 häufig für alle 3 Kategorien, sichtbar an den vielen diagonalen „Wechseln“ in Abbildung 2.18.

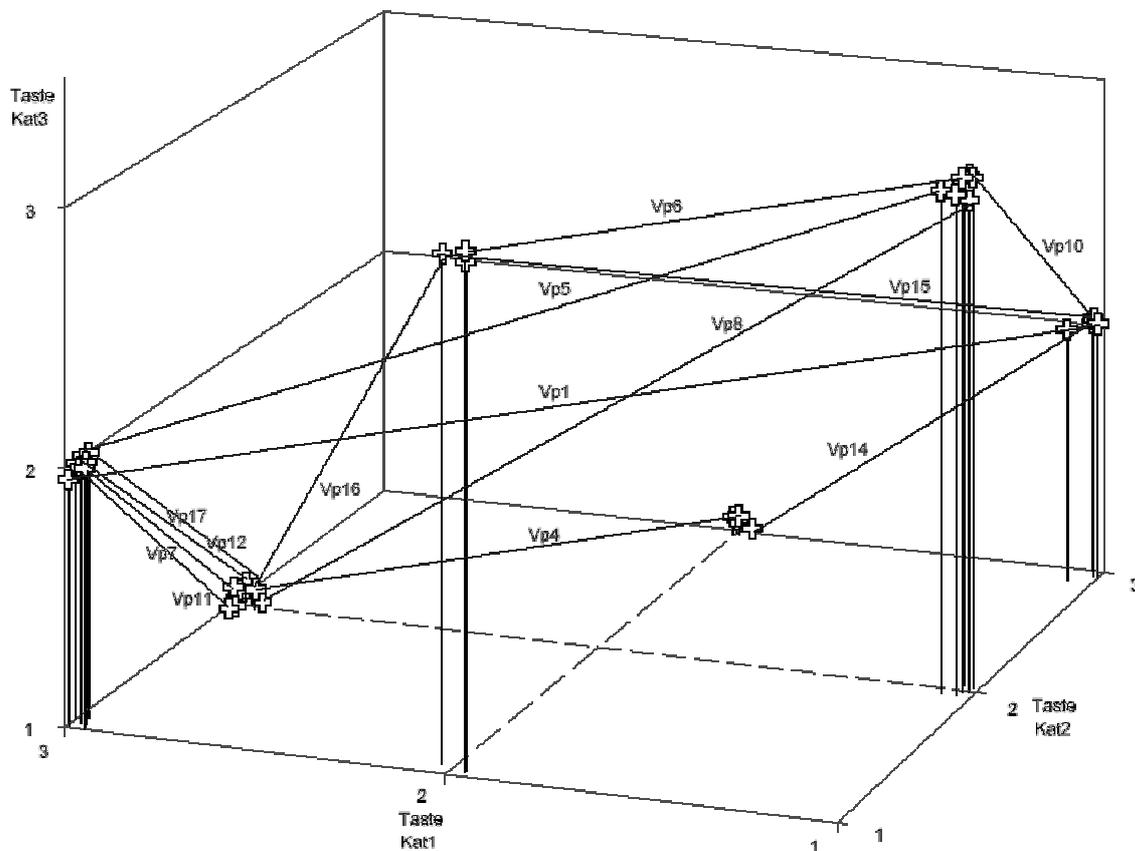


Abb. 2.18: Zuordnung der Tasten zu den Kategorien für Löser des Experimentes 2.

Erneut kann das Ergebnis für die Nicht-Löser nicht klar interpretiert werden, da sie nicht die erwarteten Kategorienstrukturen entwickelt hatten. Hier betrug die Zahlen für die Kategorien-zu-Tasten-Zuordnung 7, 8 und 8.

Kategorienwahl

Der Teststimulus EJO wurde 6 mal der Kategorie 1, 8 mal der Kategorie 2 und 3 mal der Kategorie 3 zugeordnet. Die Zuordnung unterschied sich damit nicht signifikant von einer

zufälligen Zuordnung ($\chi^2(2)=2.24$, $p>0.10$). Der Stimulus AHK wurde den Kategorien im Verhältnis 5:9:3 zugeordnet. Auch hier wurde keine signifikante Tendenz gefunden ($\chi^2(2)=3.29$, $p>0.10$). Abbildung 2.19 enthält auch die Ergebnisse der auswertbaren Nicht-Löser. Ein Vergleich der Löser in Experiment 1 und 2 zeigte einen signifikanten Unterschied für Teststimulus EJO ($\chi^2(2)=8.28$, $p<0.05$), nicht aber für Teststimulus AHK ($\chi^2(2)=0.05$, $p>0.10$). Obwohl bei beiden statistischen Tests die erwartete Zellhäufigkeit teilweise kleiner als 5 war, wird aufgrund der klaren Signifikanz bzw. klaren Insignifikanz der Chi-Quadrat Tests als bedeutungsvoll angesehen.

Zuordnung EJO

	Kategorie 1	Kategorie 2	Kategorie 3
Löser	6	8	3
Analysierbare Nicht-Löser	3	1	1

Zuordnung AHK

	Kategorie 1	Kategorie 2	Kategorie 3
Löser	5	9	3
Analysierbare Nicht-Löser	1	2	2

Abb. 2.19: Zuordnung der Teststimuli EJO und AHK zu den Kategorien. Aufgegliedert in Löser und auswertbare Nicht-Löser.

Bekanntheitsrating

Die geblockten Icons der Kategorie 1 wurden am unbekanntesten, die der ungeblockten Kategorie 2 am bekanntesten eingeschätzt. Die Icons der Kategorie 3 erhielten mittlere Bekanntheitswerte (vgl. Abbildung 2.20). Der Unterschied zwischen der am häufigsten gewählten Kategorie 2 und Kategorie 1 ist signifikant ($t(16)=-2.13$, $p<0.05$), nicht aber zu Kategorie 3 ($t(16)=0.64$, $p>0.10$).

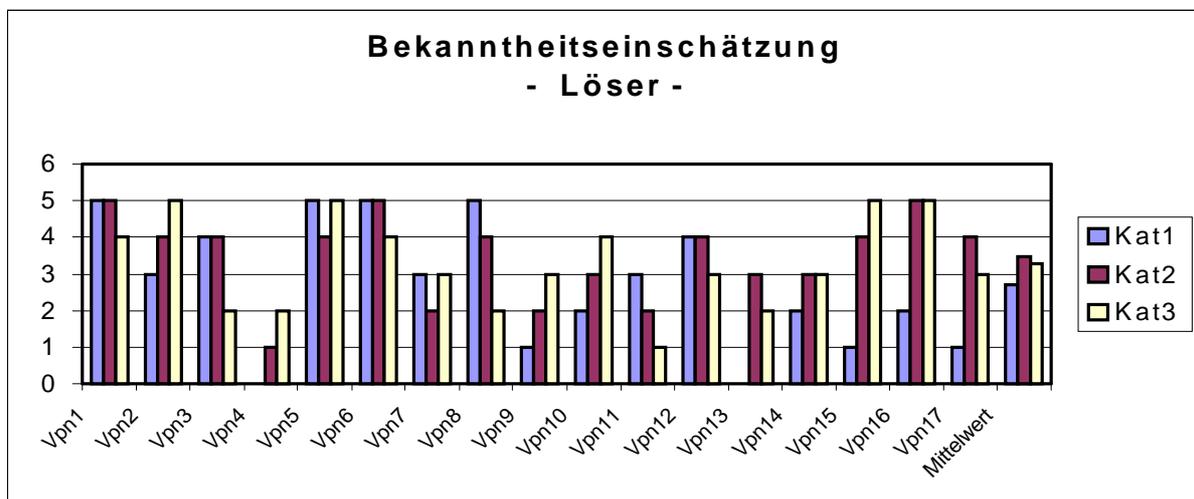


Abb. 2.20: Bekanntheitsgrad der Icons.

Auskünfte der Teilnehmer

Auf den einzelnen Dimensionen ergaben sich für die Löser folgende Nicht-Übereinstimmungen zwischen dem Erst- und Zweit-Rater: Farb-Form-Dimension: 2 von 17 ($B_{17;0.5}(2) < 0.01$), Anzahl-Dimension: 2 von 17 ($B_{17;0.5}(2) < 0.01$), Restkategorien-Dimension: 1 von 17 ($B_{17;0.5}(1) < 0.001$) und Wechsel-Dimension: 0 von 17 ($B_{17;0.5}(0) < 0.001$). Bei den Nicht-Lösern stimmten jeweils 0, 1, 4 und 0 Einstufungen nicht überein bei 18 abgegebenen Kommentaren ($B_{18;0.5}(0) < 0.001$; $B_{18;0.5}(1) < 0.001$; $B_{18;0.5}(4) < 0.05$; $B_{18;0.5}(0) < 0.001$). Da auch hier wieder sehr hohe Übereinstimmungen erzielt werden konnten, wurden die Häufigkeiten, mit denen die Bewerter die einzelnen Kategorien wählten, gemittelt.

Löser

29% der Löser wurden von den Bewertern als Klassifizierer auf Grundlage von Farben, 24% als Nutzer von nur einem Icon und 27% als Benutzer von Restkategorien eingestuft. 12% schienen in den Augen der Rater ihre Strategie im Laufe des Experimentes geändert zu haben. Der Zusammenhang zwischen der Anzahl verwendeter Icons und der Verwendung der Beurteilungsrundlage „Farben“ bzw. „Form“ erwies sich als nicht-signifikant (Fisher-Exact-Test, $p > 0.10$). Eine genauere Analyse ergab, dass bei der Klassifizierungsgrundlage „Form“ anscheinend sowohl einzelne als auch mehrere Icons gleichzeitig beachtet worden waren.

Nicht-Löser

72% der Nicht-Löser schienen eine Strategie gewählt zu haben, bei der vorwiegend die Farben zur Kategorisierung genutzt wurden. 14% benutzten in den Augen der Rater nur 1 Icon zur Bestimmung der Zugehörigkeit eines Stimulus zu einer Klasse. 22% der Aussagen wurden als Einbeziehung einer Restkategorie gewertet und keiner schien einen Wechsel der Bewertungsstrategie vollzogen zu haben. Der Zusammenhang zwischen der Anzahl verwendeter Icons und der Beurteilung aufgrund von Farben oder Form war signifikant (Fisher-Exact-Test, $p < 0.05$).

Vergleich Löser mit Nicht-Löser

Ein Vergleich der erhaltenen Zellohäufigkeiten zwischen den Lösern und Nicht-Lösern ergab für die Beurteilungsdimension, ob Farben die Grundlage der Klassifizierung darstellten einen signifikanten Unterschied ($\chi^2(1) = 6.41$, $p < 0.05$). Keinen hingegen für die Dimension der Anzahl genutzter Icons ($\chi^2(1) = 0.26$, $p > 0.10$). Fisher-Exact-Tests zeigen ebenfalls keine reliablen Unterschiede für die restlichen 2 Beurteilungsdimensionen (Fisher-Exact-Test, $p > 0.10$; Fisher-Exact-Test, $p > 0.10$).

Vergleich Löser Experiment 1 und Löser Experiment 2

Es zeigt sich auf allen 4 Beurteilungsdimensionen kein signifikanter Unterschied zwischen den Lösern des 1. und 2. Experimentes ($\chi(1)=0.95$, $p>0.10$; $\chi(1)=0.90$, $p>0.10$; $\chi(1)=0.33$, $p>0.10$; $\chi(1)=0.19$, $p>0.10$).

2.3.2 Diskussion Experiment 2

Im Wesentlichen zeigte sich dieselbe Zunahme an Konsistenz und Abnahme an Reaktionszeit über die 2 Phasen wie in Experiment 1. Der Anteil an Lösern und Nicht-Lösern ist ebenfalls wieder relativ konstant für die 1. und 2. Erhebungsstufe. Die ordinale Ordnung der Kategorienstrukturen nach ihrer Schwierigkeit dürfte also auch in Experiment 2 stark zwischen den Individuen variieren.

Interessanterweise änderte die längere Entscheidungszeit nicht maßgeblich den Anteil an Lösern zu Nicht-Lösern. Die längere Darbietungszeit schien also nicht zwangsläufig eine Strategie zu induzieren, die nur einzelne Icons verwendet. Ansonsten wäre bei einer ausreichenden Darbietungszeit von einer starken Zunahme an Lösern auszugehen, da ohne Zeitdruck die einzelnen Icons fehlerfrei analysiert werden dürften. Wie aus dem signifikanten Unterschied in der Reaktionszeit zwischen Experimente 1 und 2 ersichtlich, nutzen die Versuchspersonen die längere Zeit auch. Die Versuchspersonen in Experiment 1 dürften daher eine andere „Strategie“ gewählt haben, welche auch unter Zeitdruck zu Erfolg führen kann. Dies könnte erklären, warum die längere Darbietungs- und damit Verarbeitungszeit in Experiment 2 nicht zu einer starken Zunahme des Löser-Anteils führte, bzw. die kürzere Zeit in Experiment 1, die Löser nicht stärker beeinträchtigte. Eine alternative Erklärung wäre, dass die Strategie in Experiment 1 beschleunigt ausgeführt und damit vollständig im Zeitlimit abgeschlossen werden konnte. Bestärkt wird die Vermutung unterschiedlicher Strategien in Experiment 1 und 2 durch Unterschiede in den Reaktionsmaßen. So war z.B. die Konsistenz der Zuordnung der 1. Kategorie zu einer Taste in den letzten Blöcken der beiden Experimentalphasen in Experiment 2 gegenüber Experiment 1 signifikant vermindert.

Die Tastenzuordnung für Teststimulus EJO wich nur in Experiment 2 marginal signifikant von der Gleichverteilung ab. Würde sich dieser Unterschied als reliabel herausstellen, wäre dies ein weiteres Indiz für unterschiedliche Strategien in Experiment 1 und 2. Die Ergebnisse der Tastenzuweisung legen erneut nahe, dass über die Versuchspersonen hinweg bestimmte Tasten eine bestimmte Funktion wie z.B. „Reaktionstaste bei Überraschung“ erfüllten. Eventuell verhinderte die längere Reaktionszeit, ein „versehentliches“ Vertauschen der Kategorien und gewährleistete damit, dass sich die anfangs getroffenen Kategorienzuordnungen bis zu Teststimulus EJO erhalten konnten. Auch

könnte aufgrund der längeren Reaktionszeit eine Restkategorienüberlegung eine Rolle spielen. Würden die Restkategorien interindividuell eher einer bestimmten Taste zugeordnet, so würde sich die Abweichung von der Gleichverteilung für den neuen, „nicht-klassifizierbaren“ Stimulus EJO erklären. Für die interindividuelle Tendenz einer Restkategorie eine bestimmte Taste zuzuordnen spricht auch die signifikante Bevorzugung einer bestimmten Taste als Reaktion auf den 1. Stimulus.

Vorstellbar wäre weiterhin, dass die Versuchspersonen, bedingt durch die längere Darbietungszeit, andere und komplexere Strategien der kognitiven Verarbeitung bemühten. Durch die längere Zeit ist vor allem zu erwarten, dass simple Ähnlichkeitsurteile auf einer oberflächlichen Vergleichsebene durch abstraktere Ähnlichkeitsurteile ersetzt werden. Ein Beispiel für den ersten Fall wäre eine Beurteilung anhand der Anzahl gemeinsamer Icons. Dem zweiten Fall würde eine Beurteilung aufgrund einer Regel wie z.B. „oben rechts ein blaues und unten links ein rotes Icon“ entsprechen. Aber auch Memorier-Strategien wie z.B. alle Icons explizit zu erinnern könnten durch die Zeit zur Reflexion ermöglicht worden sein. Die direkte Ähnlichkeitsbeurteilung aufgrund gemeinsamer Icons würde damit durch abstraktere, von den einzelnen Icons gelöste Eigenschaften ersetzt werden. In diesem Fall würde Blocking nicht mehr auf der Iconebene ansetzen. Dies könnte erklären, warum die Zuweisung des Teststimulus EJO sich nicht mehr signifikant von der Gleichverteilung unterschied.

Kategorienzuordnung

Dass die Veränderung der Darbietungszeit einen Einfluss auf das Verhalten der Versuchspersonen hatte, zeigte sich besonders deutlich bei der Zuordnung des Teststimulus EJO. Bei einem sonst identischen Design verhielten sich die Versuchspersonen von Experiment 1 und 2 bezüglich Teststimulus EJO unterschiedlich ($\chi^2(2)=8.28, p<0.05$). Wäre der signifikante Effekt in Experiment 1 Folge eines nicht bedachten Artefakts des Designs gewesen, hätte sich durch die Veränderung der Präsentationszeit kein Unterschied zeigen sollen.

Eine mögliche Strategie der Versuchspersonen stellt eine 1-Icon-pro-Kategorie Strategie dar:

1. Suche zufällig aus jeder Kategorie 1 Icon aus, und entscheide aufgrund deren Anwesenheit über die Kategorienzugehörigkeit.
2. Falls neue Stimuli auftreten, die keine bekannten Icons enthalten, wähle ein neues Icon aus.
3. Falls mehrere gewählte Icons gleichzeitig präsent sind, wähle zufällig 1 Icon von den bekannten für die Entscheidung aus.

Diese Strategie wurde auch durch informelle Mitteilungen der Versuchspersonen nach dem Experiment nahegelegt.

Unter diesen Annahmen würde sich die Wahrscheinlichkeit für Gruppe 1, 1 Icon des Buchstabens E in der Testphase zu kennen, aus der Wahrscheinlichkeit ergeben, dass in der 1. Phase 1 C-Icon gewählt wurde und in der 2. Phase ein E-Icon: $\frac{1}{3} * \frac{1}{3}$. Für den Buchstaben J wäre die Wahrscheinlichkeit $\frac{1}{3}$ und für O $\frac{2}{3} * \frac{1}{3}$.

Für die Wahl der E-Kategorie als Reaktion auf den Teststimulus beträgt die Wahrscheinlichkeit bei Kenntnis aller drei Buchstaben $\frac{1}{3}$ (Annahme 3). Es wurden alle Kombinationen bestimmt, in denen relevante Icons in Teststimulus EJO auftreten konnten. Diese Kombinationen wurden mit den entsprechenden Wahrscheinlichkeiten des Eintretens gewichtet und mit der Wahlwahrscheinlichkeit der verschiedenen Kategorien bei Vorliegen der einzelnen Kombinationen aus relevanten Icons multipliziert. Daraus ergaben sich die Wahrscheinlichkeiten den Teststimulus EJO in die jeweiligen Kategorien einzuordnen zu: 0.24, 0.43 und 0.33. Bei 17 Versuchspersonen betragen die erwarteten Häufigkeiten damit: 4.02, 7.38 und 5.60. Der Chi-Quadrat Wert dieser Vorhersage, verglichen mit der tatsächlich erhaltenen Verteilung ergab eine gute Modellpassung ($\text{Chi}(2)=2.23$, $p>0.10$). Allerdings ergab die Berechnung der Abweichung der empirischen Werte von der Gleichverteilung eine nur minimal schlechtere Passung ($\text{Chi}(2)=2.24$, $p>0.10$). Für Stimulus AHK ergaben sich nach der gleichen Berechnungsprozedur erwartete Häufigkeiten von 5.28, 4.44 und 7.28. Die Modellpassung der empirischen Daten an diese erwarteten Häufigkeiten ergab eine signifikante Abweichung ($\text{Chi}(2)=7.20$, $p<0.05$). Die Abweichung der empirischen Daten von der Gleichverteilung zeigte hingegen eine relativ gute Modellpassung ($\text{Chi}(2)=3.29$, $p>0.10$). Die Häufigkeitszuordnung des Stimulus AHK widersprach also der vermuteten Strategie, welche annahm, dass jeweils nur ein Icon für jede Kategorie als Klassifizierungsgrundlage genutzt wurde.

Von einem Vertreter der obigen Strategievorstellung könnte natürlich eingewendet werden, dass die Klassifizierungsentscheidungen für Teststimulus AHK eventuell durch weitere Prozesse oder auch versehentliche Fehlentscheidungen überlagert worden waren. Um zu überprüfen, ob die Versuchspersonen eventuell doch die oben diskutierte Strategie verfolgten und jeweils nur 1 Icon pro Kategorie memorierten, wurden als zweites die Durchgänge analysiert in denen ein Icon vertauscht worden war. Wenn das „eingetauschte“ Icon die Kategorienentscheidung für die ursprüngliche Kategorie bestimmte, so sollte in 55% der Fälle das „eingetauschte“ Icon die Klassifizierungsgrundlage auch im neuen Stimulus darstellen (Für Herleitung der Wahrscheinlichkeit, siehe Anhang 11). In diesem Fall sollte der

Stimulus in die *ursprüngliche* Kategorie des Icons eingeordnet werden. Wird ein Icon vertauscht und die Klassifizierung erfolgt nicht aufgrund des Icons, so wird das Icon aus der Sicht seines eigentlichen Prototyps falsch zugeordnet. Stimuli die im letzten Block der beiden Phasen konsistent in dieselbe Kategorie eingeordnet werden, obwohl deren Icons manchmal auch in einer „neuen“ Stimulusumgebung des anderen Prototyps auftauchen, machen das oben diskutierte Modell unwahrscheinlicher. Dies gilt vor allem dann, wenn davon ausgegangen wird, dass die Regel immer befolgt wird. Davon kann aber vermutlich in den letzten „Löserblöcken“, die sich durch eine erfolgreiche Entscheidungsstrategie auszeichnen, ausgegangen werden. Tritt hingegen ein Icon in einer neuen Stimulusumgebung einer „fremden“ Kategorie auf und wird der Stimulus der Kategorie zugewiesen aus der das Icon stammt, so wird die Modellvorstellung, der zufolge Entscheidungen auf jeweils einem Icon beruhen, gestärkt. Im folgenden werden solche Icons als „identifizierte“ Icons bezeichnet. Tabelle 2.3 zeigt die erhaltenen Anzahlen identifizierter Icons für Experiment 1 und 2.

Tabelle 2.3: Anzahl der „identifizierten“ Icons im letzten Block der 1. und 2. Phase. Die erste Spalte gibt jeweils die Anzahl der „identifizierten“ Icons in der 1. bzw. 2. Phase wider. In der jeweiligen zweiten Spalte ist die Anzahl der Versuchspersonen, welche dieses Muster zeigten dargestellt. Markiert sind die 5 Versuchspersonen, welche in Experiment 2 mehr als 1 Icon in mindestens einer Phase zu nutzen schienen.

Experiment 1		Experiment 2	
Anzahl identifizierter Icons	Anzahl Versuchspersonen.	Anzahl identifizierter Icons	Anzahl Versuchspersonen
Phase 1 Phase 2		Phase 1 Phase 2	
0 0	5	0 0	2
0 1	2	0 1	4
1 0	2	1 0	3
1 1	2	1 1	3
0 2	2	2 0	3
2 2	1	2 1	1
0 3	1	0 2	1

Es stellte sich heraus, dass von den 17 Versuchspersonen des 2. Experiments anscheinend nur 5 Versuchspersonen im letzten Block der 1. und 2. Phase mehr als 1 Icon in mindestens einer Phase als Grundlage verwendeten. Nur bei einer Versuchsperson stammten die beiden Icons aus unterschiedlichen Kategorien. Keine Versuchsperson in Experiment 2 schien ihre Kategorisierung auf mehr als 2 Icons zu basieren. In Experiment 1 gab es eine Versuchsperson, welche anscheinend 3 Icons als Grundlage nutzte. Allerdings kamen auch hier alle 3 Icons aus derselben Kategorie. Ebenso stammten bei den restlichen 3 Personen aus Experiment 1, für die 2 Icons „identifiziert“ werden konnten, die Icons aus derselben Kategorie.

Ein Problem bei der Interpretation stellt die Tatsache dar, dass nicht notwendigerweise alle Icons einmal vertauscht wurden und sich deshalb als Klassifizierungsgrundlage erweisen konnten. Die Wahrscheinlichkeit des empirischen Ergebnisses in Experiment 2, dass über alle Versuchspersonen in keinem der beiden Blöcke alle 3 Icons entdeckt wurden, beträgt 19% (für Herleitung siehe Anhang 11).

Die Wahrscheinlichkeit, dass höchstens in einem Fall 2 Icons aus unterschiedlichen Kategorien beobachtet werden konnten, ist wesentlich kleiner als 0.001 (für Herleitung siehe Anhang 11) und bliebe auch noch signifikant für den Fall, dass es sich bei den Icons der 5 Versuchspersonen mit mehr als 1 Icon, um Icons aus unterschiedlichen Kategorien gehandelt hätte. Auch unter der Annahme, dass nur 7 Versuchspersonen diese Strategie verwendeten wäre die Anzahl „identifizierter“ Icons signifikant unter dem zu erwartenden Wert.

Damit kann das Modell, welches die Klassifizierung aufgrund einzelner Icons vorhersagt, jedenfalls in der oben formulierten Schärfe der zufälligen Iconselektion bei Konflikt (Annahme 3) mit relativ großer Sicherheit sowohl für Experiment 1 als auch 2 widerlegt werden. Nicht im Widerspruch sind allerdings diese Ergebnisse zu dem 2-Prozess Modell von Ahn und Medin (1992), welches im Konfliktfall keine zufällige Selektion eines Icons fordert, sondern eine Erweiterung der Klassifizierungsgrundlage erwartet.

Eine andere Möglichkeit die unterschiedlichen Kategorienzuordnungen in Experiment 1 und 2 zu erklären ist, dass die Versuchspersonen in Experiment 2 eine ähnliche Strategie wie die Versuchspersonen in Experiment 1 verfolgten, allerdings aufgrund der längeren Verarbeitungszeit die Kategorie 3 in Phase 2 wiedererkennen konnten. Unter diesen Voraussetzungen sollte sich eine Präferenz für Kategorie 2 entwickeln, da die Wiedererkennungshypothese, wie sie oben formuliert wurde, nicht mehr gilt. Diese Präferenz für Kategorie 2 zeigte sich zwar numerisch, ließ sich aber nicht signifikant absichern. Mögliche Gründe für diesen schwachen Effekt werden im nächsten Abschnitt diskutiert.

Einschätzung der Bekanntheit

In Experiment 2 wurden die Icons aus Kategorie 2 als signifikant bekannter eingeschätzt als Icons aus Kategorie 1 und numerisch bekannter als Icons aus Kategorie 3. Gleichzeitig wurde, wie oben berichtet, auch Kategorie 2 für Teststimulus EJO numerisch am häufigsten gewählt. Dies scheint die in der Diskussion zu Experiment 1 geäußerte Vermutung zu bestätigen, dass die Dissoziation zwischen Kategorienentscheidung und Bekanntheitsgrad der Icons durch genügend Zeit aufgehoben wird. Die Erklärung des Unterschiedes zwischen den Reaktionsmaßen „Kategorienzuordnung“ und „Bekanntheitseinstufung“ durch eine unterschiedliche Sensitivität wird durch die hier gezeigte Umkehr des signifikanten

Reaktionsmaßes in ein nicht signifikantes und des nicht-signifikanten Reaktionsmaßes in ein signifikantes unwahrscheinlich. Auch zeigte eine separate Analyse für Versuchspersonen, welche in der Bekanntheitseinschätzung einen Blocking-Effekt zeigten und Versuchspersonen, die in der Bekanntheitseinschätzung keinen Blocking-Effekt zeigten, keinen wesentlichen Unterschied³⁵ bei der Kategorienzuordnung des Teststimulus EJO. Bestünde der Unterschied zwischen den beiden Reaktionsmaßen in einer unterschiedlichen Sensitivität bei unterschiedlichen Darbietungszeiten wäre zumindest eine Assoziation zwischen Bekanntheit und Kategorienzuordnung zu erwarten. Dass Kategorie 2 am bekanntesten eingestuft und diese Kategorie auch numerisch am häufigsten gewählt wurde, könnte darauf hindeuten, dass bei mehr Zeit doch die Ähnlichkeit zwischen den Stimuli aus der 3. Kategorie der 1. und 2. Phase erkannt werden konnte. Ein Grund für die bessere Erkennungsleistung könnte sein, dass länger über die Stimuli nachgedacht werden konnte. Eventuell konnte auch aufgrund der längeren Verarbeitungszeit eine größere Anzahl an Icons oder komplexere Merkmale als Klassifizierungsgrundlage berücksichtigt und damit die objektiv gegebene Ähnlichkeit besser erkannt werden. Unter Annahme derselben „Blocking-Mechanismen“ wie in Experiment 1 wäre dann die Zuordnung von EJO zur 2. Kategorie zu erwarten. Der schwächere Effekt auf die Kategorienzuordnung als auf die Bekanntheitseinschätzung könnte darauf zurückgeführt werden, dass metakognitive Strategien und deliberate Prozesse die Kategorienzuordnungen überlagerten, während die Erinnerungsprozesse nicht durch diese metakognitiven Strategien beeinflusst wurden. Metakognitive Strategien sind vor allem auch deshalb zu erwarten, weil in der künstlichen Laborsituation die Versuchspersonen explizit auf ihre Aufgabe hingewiesen wurden und sie diese Leistungs-Situation sicherlich reflektierten. Diese Vorstellung der Verwendung von metakognitiven Strategien ist konsistent mit dem Ergebnis, dass die Reaktionszeit auf den Teststimulus AHK signifikant langsamer ausfiel als auf das letzte Trainings-Item. Dies wäre unter der Annahme, dass die Kategorie 3 in Phase 2 aufgrund der längeren Zeit wiedererkannt werden konnte, zu erwarten, da in diesem Fall sowohl A als auch K eine nahezu identische Verbindungsstärke zu AHK aufweisen sollten. Gleichzeitig sollte die Verlangsamung für AHK in Experiment 2 stärker als in Experiment 1 sein. Der Grund hierfür ist, dass in Experiment 1 die mit geringer Salienz *neu* zu lernende Verbindung zwischen K und Kategorie 3 die „Konkurrenz“ verringerte. Dieser Unterschied war in einem t-Test unter der Annahme

³⁵ Für die Analyse wurden die Versuchspersonen in 2 Gruppen eingeteilt. In die 1. Gruppe kamen Probanden, die eine höhere Bekanntheitseinschätzung für Icons aus Kategorie 2 als für Icons aus Kategorie 1 abgegeben hatten. Die restlichen Versuchspersonen wurden der anderen Gruppe zugeteilt. Das Verhältnis der Kategorienwahlen für EJO betrug in der 1. Gruppe 3:4:2 und in der 2. Gruppe 3:4:1.

ungleicher Varianzen signifikant, da sich aber aufgrund des höheren Zeitlimits die Varianz der Reaktionszeiten vergrößerte, sollten auch die Differenzen größer werden. Der konservativere U-test, der weniger durch ungleiche Varianzen verfälscht wird, wurde nicht-signifikant. Daher kann die größere Differenz zwischen der Reaktionszeit auf AHK und dem letzten Item der 2. Phase in Experiment 2 verglichen mit Experiment 1 nicht als valide für einen größeren Konflikt bei Stimulus AHK angesehen werden. Damit stellt die Annahme über die Wiedererkennung der 3. Kategorie in der 2. Phase und der damit einhergehende größere Konflikt auf Teststimulus AHK, auf Basis der gegenwärtigen Daten reine Spekulation dar.

Auskünfte der Teilnehmer

Dass bei den Lösern in Experiment 1 sowie bei den Nicht-Lösern des Experimentes 2 Formbeurteilungen häufig zusammen mit der Beurteilung nur eines Icons auftraten, nicht aber bei den Lösern des Experiment 2, könnte so gedeutet werden, dass die Form-Kategorisierung mehr kognitive Ressourcen beansprucht als die Farbbeurteilung. Da in Experiment 1 (und Experiment 3) das Zeitlimit vermutlich³⁶ eine Ressourcenlimitierung in Form verkürzter Verarbeitungszeit erzeugte, reichte die Prozesskapazität nicht aus, auch bei Formverarbeitung mehrere Icons zu beachten. Hingegen standen aufgrund der längeren Entscheidungszeit in Experiment 2 genügend Ressourcen in Form von z.B. Arbeitsgedächtnis zur Verarbeitung und Abspeicherung mehrerer Icons zur Verfügung. Dies trifft jedoch nicht auf die *Nicht-Löser* des 2. Experimentes zu, welche, aufgrund ihrer Überforderung, auf keine freien Ressourcen mehr zurückgreifen konnten. Ob sich Klassifizierungsvorgänge tatsächlich sinnvoll als abhängig von kognitiven Ressourcen beschreiben lassen und die Ressourcenbeschränkung nicht durch metakognitive Strategien umgangen wird, ist eine nicht einfach zu beantwortende Frage. Die Tatsache, dass auch die Nicht-Löser ihr Zeitbudget nicht ganz ausschöpften, spricht jedenfalls nicht automatisch gegen diese Sichtweise der limitierten kognitiven Ressourcen, da Versuchspersonen nicht nur durch das Zeitlimit sondern auch durch die eigene fehlende Motivation komplexere Verarbeitungsvorgänge auszuführen, Ressourcenbeschränkungen unterliegen können.

Die Auskünfte der Probanden über ihre Klassifizierungsstrategie zeigten hinsichtlich der 4 ausgewerteten Dimensionen keine Unterschiede zwischen Teilnehmern des 1. und 2. Experimentes. Im Gegensatz dazu erfolgten bei vergleichbaren objektiven Stimulusstrukturen unterschiedliche Kategorienentscheidungen in Experiment 1 und 2. Dies bestärkt die Vermutung, dass die Versuchspersonenberichte keinen validen Aufschluss über die

³⁶ Dies muss nicht unbedingt als Folge einer zeitlimitierten Darbietung auftreten. So lässt sich eine Parallelverarbeitung der Icons und eventuell sogar der Stimuli denken.

verwendeten Klassifizierungsstrategien erlauben. Ein weiterer Test bestätigte diese Vermutung: Ein neuer, mit der inhaltlichen Frage nicht betrauter Rater erhielt in zufälliger Reihenfolge die auf Karteikärtchen geschriebenen Auskünfte von 14 Löser aus Experiment 1 und 17 Löser aus Experiment 2. Der Rater sollte die Karten entsprechend der Ähnlichkeit der Beschreibungen in 2 annähernd gleichgroße Gruppen sortieren. Der Anteil an Lösern aus Experiment 1 zu Lösern aus Experiment 2 betrug 5:3 in der einen gebildeten Gruppe und 9:14 in der anderen Gruppe. Ein Fisher-Exact-Test zeigte keine signifikante Tendenz die Löser aus Experiment 1 und 2 in unterschiedliche Gruppen einzuteilen (Fisher-Exact-Test, $p > 0.10$). Hingegen ließen sich Nicht-Löser von den Lösern anhand ihrer verbalen Beschreibungen unterscheiden: Ein weiterer Rater sollte 17 Beschreibungen von Lösern und Nicht-Lösern aus Experiment 2 ebenfalls in 2 annähernd gleichgroße Gruppen aufteilen. In der einen Gruppe ordnete der Rater 10 Löser und 4 Nicht-Löser ein. Die andere Gruppe bestand aus 7 Löser und 13 Nicht-Löser. Ein Chi-Quadrat Test ergab einen signifikanten Unterschied für das Verhältnis Löser zu Nicht-Löser in den beiden Gruppen ($\chi(1) = 4.37$, $p < 0.05$). Diese beiden Tests zeigen somit, dass es zwar anhand der Beschreibungen möglich ist Löser von Nicht-Lösern zu unterscheiden, aber nicht zwischen Lösern aus Experiment 1 und 2.

2.3.3 Zusammenfassung Experiment 2

Das 2. Experiment scheint einen Effekt der längeren Bearbeitungszeit auf die eingesetzten Strategien zu zeigen. Während sich der Blocking-Effekt unter Vorbehalt der zwei post-hoc Annahmen in Experiment 1 bei kurzen Darbietungszeiten in der Kategorienentscheidung zeigen konnte, wurde dieser Effekt in Experiment 2, vermutlich aufgrund der längeren Darbietungszeit und den damit verbundenen Reflexionen, durch weitere Entscheidungsstrategien überlagert. Während die Verhaltensweisen bei kurzen Reaktionszeiten eher denen aus Tierexperimenten entsprechen dürften, legen Experimente mit längeren Reaktionszeiten und eventuell damit verbundenen Reflexionen Vergleiche mit Blocking-Experimenten, wie sie mit Menschen bei kognitiven Aufgaben durchgeführt wurden (z.B. Gluck & Bower, 1988a; Dickinson, Shanks & Evenden, 1984), nahe. Während aber in diesen „kognitiven“ Aufgabendesigns die explizite Entscheidung, redundant erscheinende Informationen zu vernachlässigen, ebenfalls zu einem Blocking-Effekt führen würde, erlaubte das hier verwendete Design unterschiedliche Konzeptualisierungen des Stimulusraumes. Damit besteht die Möglichkeit, dass Blocking-Effekte, welche auf einer anderen Ebene als der erwarteten Iconebene arbeiten, nicht entdeckt werden. „Metakognitive“ Blockingprozesse werden somit nicht fälschlicherweise als assoziative Blocking-Effekte verstanden. Da natürliche Klassifizierungsprozesse vermutlich ebenfalls in einer wenig reflektierenden Form

ablaufen, dürfte durch den Nachweis von Blocking auf assoziativer Ebene in Experiment 1 die externe Validität gemeinsamer Kategorisierungsmodelle für Kategorisierungs- und Klassifizierungsaufgaben gegeben sein. Die Tatsache, dass die Bekanntheitseinstufung der nichtgeblockten Icons signifikant höher ausfiel, der Blocking-Effekt sich aber bei der Kategorienentscheidung nicht signifikant nachweisen ließ, könnte auf die Verwendung komplexerer Ähnlichkeitsmaße und deliberater Entscheidungsstrategien zurückgeführt werden. Hingegen verlief die Abspeicherung der Icons automatisch und unbeeinflusst von abstrakten Regeln oder Erfahrungen über die Vorhersagekraft der Icons. Damit konnte sich der Blocking-Effekt auf Iconebene in Form eines geminderten Bekanntheitsgrad der geblockten Icons zeigen. Vermutlich könnte diese „automatische“ Abspeicherung der Icons dadurch verhindert werden, dass Versuchspersonen explizit aufgefordert werden, die Icons zu erinnern. Unter dieser Instruktion wären auch hier metakognitive Strategien zu erwarten, die vermutlich den Blocking-Effekt überlagern würden.

Aus der Zuweisungshäufigkeiten des Stimulus AHK zu den einzelnen Kategorien sowie aus der Anzahl an „identifizierten“ Icons konnte geschlossen werden, dass die Versuchspersonen nicht die oben diskutierte 1-Icon-pro-Kategorie Strategie verfolgten.

Schließlich wurde in einem weiteren Experiment versucht den Blocking-Effekt bei 2 Kategorien zu belegen, um den Einfluss eventuell nichtbedachter Artefakte der 3-Kategorien Bedingung zu vermindern.

2.4 Experiment 3

Aufgabe in Experiment 3 war es, die dargebotenen Stimuli in eine von 2 Kategorien zu klassifizieren. Die Darbietungszeit wurde auf 800ms begrenzt, gefolgt von einer für 1200ms eingeblendeten grauen Maske. Durch die Verkürzung der Darbietungszeit konnte die maximale Anzahl an Blöcken wieder auf 6 erhöht werden, ohne, dass das Gesamtexperiment länger als 30 Minuten dauerte. 30 Minuten wurden als Obergrenze für die Konzentrationsfähigkeit und Kooperationswilligkeit angesehen. Als Kriterium, um in die nächste Phase zu gelangen, wurde die „richtige“ Klassifikation von mindestens 10 Items pro Kategorie festgelegt. Damit mussten mindestens 83% der Items konsistent einer Kategorie zugeordnet werden. Die Icons vertauschten vor jeder Stimuluspräsentation zufällig ihre Plätze im Rechteck. Die äußere Form glich somit stark dem Experiment von Zwickel und Wills (2002). Die einzige Änderung betraf den Aufbau der Phasen, welcher eher denen von Experiment 1 und 2 entsprach. Abbildung 2.21 zeigt den Aufbau von Zwickel und Wills (2002). Abbildung 2.22 den hier verwendeten Aufbau.

	<u>Phase 1</u>	<u>Phase 2</u>	<u>Phase 3</u>	<u>Phase 4</u>	<u>Test</u>
Gruppe 1	GH IJ	GE IF	AB CD	AE CF	(CE, AF)
Gruppe 2	GH IJ	GA IC	EB FD	AE CF	(EC, AF)

Abb. 2.21: Aufbau von Zwickel & Wills (2002). Die Präsentationshäufigkeiten der geblockten und ungeblockten Testicons sind über die Phasen hinweg gleich.

	<u>Phase 1</u>	<u>Phase 2</u>	<u>Test 1</u>	<u>Test 2</u>	<u>Bekanntheit</u>
Gruppe 1	AB CD	AE GF	EF	AG	(E,F)
Gruppe 2	GB CD	GF AE	EF	AG	(E,F)

Abb. 2.22: Aufbau von Experiment 3. Die geblockten Icons der Gruppe 1 stellen die ungeblockten in Gruppe 2 dar.

Wie aus dem Design ersichtlich, wird das 1. Testitem EF generiert, indem die Hälfte seiner Elemente aus der geblockten Kategorie und die andere Hälfte aus der ungeblockten Kategorie entnommen werden. Der Instruktionstext wurde so abgeändert, dass er auf 2 Kategorien anwendbar war und die Daten mittels der linken und rechten STRG-Taste erhoben. Die Übungsphase entsprach im Wesentlichen den Übungsphasen von Experiment 1 und 2.

2.4.1 Ergebnisse

18 der 36 Versuchspersonen lernten im Verlauf des Experimentes die Stimuli gemäß der erwarteten Kategorienstruktur zu klassifizieren. Wiederum war das Verhältnis von Lösern zu Nicht-Lösern auf beiden Erhebungsstufen gleich (12:12 bzw. 6:6). Zwischen allen 3 Experimenten bestand kein signifikanter Unterschied in dem Verhältnis Löser zu Nicht-Löser ($\chi^2(2)=0.52, p>0.10$).

Anzahl benötigter Blöcke

Die Anzahl der benötigten Blöcke für Phase 1 und 2 unterschieden sich nicht signifikant ($t(17)=1.31, p>0.10$). Auch war nur eine sehr geringe Korrelation festzustellen ($r=0.15, p>0.10$). Nur 2 der 18 Lösern brauchten mehr als 5 Blöcke (vgl. Abbildung 2.23).

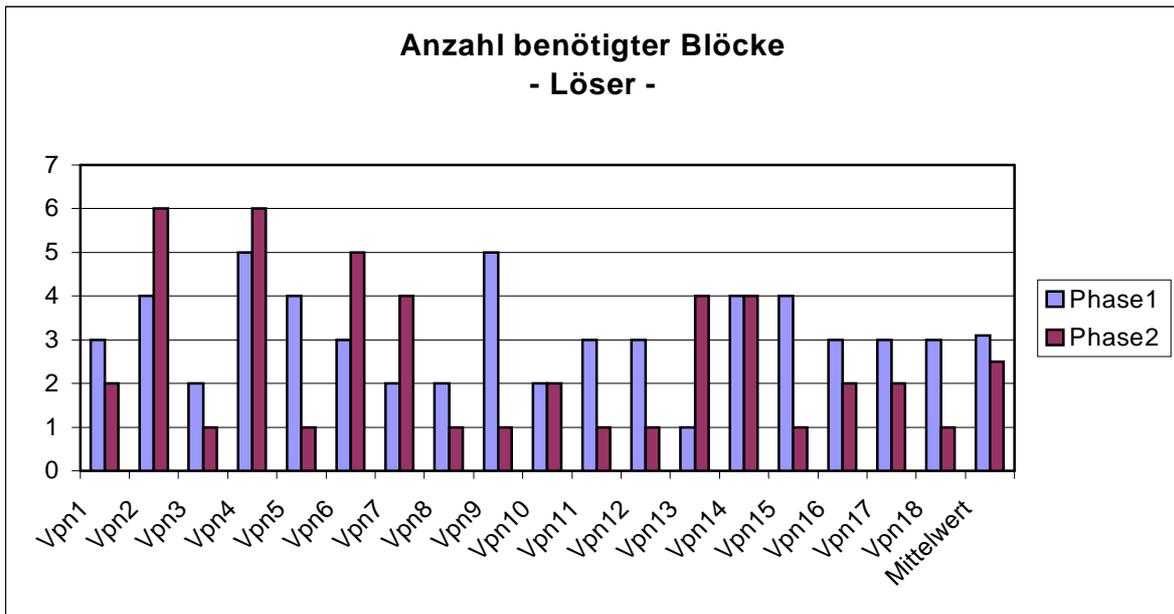


Abb. 2.23: Anzahl benötigter Blöcke bei den Lösern.

Reaktionszeit

Die Reaktionszeit nahm nur noch leicht über die Blöcke hinweg ab. Die mittleren Reaktionszeiten im 1. und individuell letzten Block in Phase 1 und 2 nahmen nicht signifikant ab (vgl. Abbildung 2.24; $t(17)=1.87, p<0.10$; $t(17)=-0.25, p>0.10$). Ebenfalls zeigte sich kein signifikanter Unterschied zwischen der mittleren Reaktionszeit des letzten Blockes der 1. Phase und dem 1. Block der 2. Phase

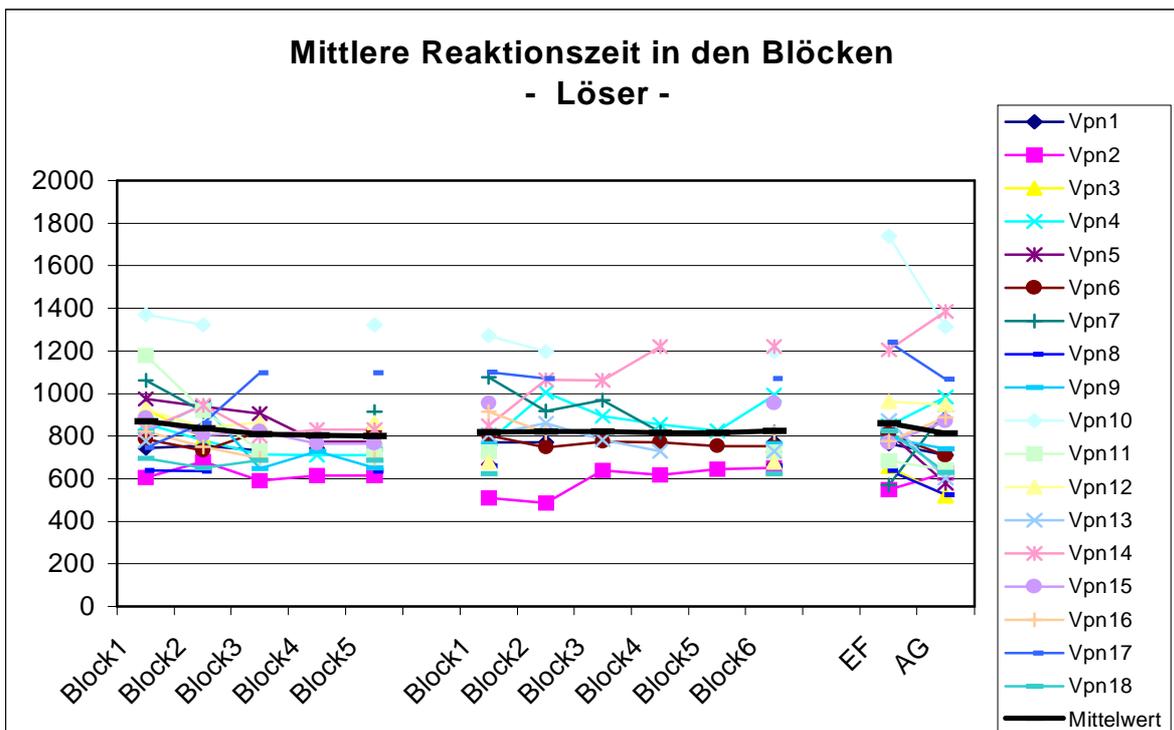


Abb. 2.24: Mittlere Reaktionszeit in den Blöcken.

Die Reaktionszeit auf EF differierte nur schwach von der Reaktionszeit auf das letzte Item ($t(17)=-0.56$, $p>0.10$). Das gleiche galt für die Differenz des Stimulus AG und dem letzten Item ($t(17)=0.10$, $p>0.10$). Erneut überschritt kein Löser in seiner Reaktionszeit auf die Testitems das Zeitlimit für die Lernitems von 2 Sekunden. Über alle Löser in den 3 Experimenten hinweg zeigte sich ein signifikanter Anstieg in der Reaktionszeit auf die 2 Testitems verglichen mit dem individuell letzten Item ($t(49)=-2.14$, $p<0.05$; $t(49)=-2.65$, $p<0.05$). Der Vergleich der Reaktionszeiten auf die beiden Testitems war über alle Experimente hinweg nicht signifikant ($t(49)=-0.45$, $p>0.10$).

Gesamt-Konsistenz

Es zeigte sich erneut eine Zunahme der Konsistenz über die Blöcke hinweg (vgl. Abbildung 2.25). Die ersten Blöcke beider Phasen unterschieden sich signifikant voneinander (Wilcoxon $Z=-3.36$, $p<0.01$). Nicht jedoch die jeweiligen letzten Blöcke beider Phasen (Wilcoxon $Z=-0.03$, $p>0.10$). Die Zunahme war in der 1. Phase zwischen dem 1. und individuell letzten Block stark ausgeprägt (Wilcoxon $Z=-3.63$, $p<0.001$). Der 1. und der letzte Block in der 2. Phase unterschieden sich hingegen nur marginal signifikant voneinander (Wilcoxon $Z=-1.79$, $p<0.10$). Ebenfalls nicht signifikant erwies sich der Unterschied zwischen dem 1. Block der 2. Phase und dem letzten Block der 1. Phase (Wilcoxon $Z=-1.36$, $p>0.10$).

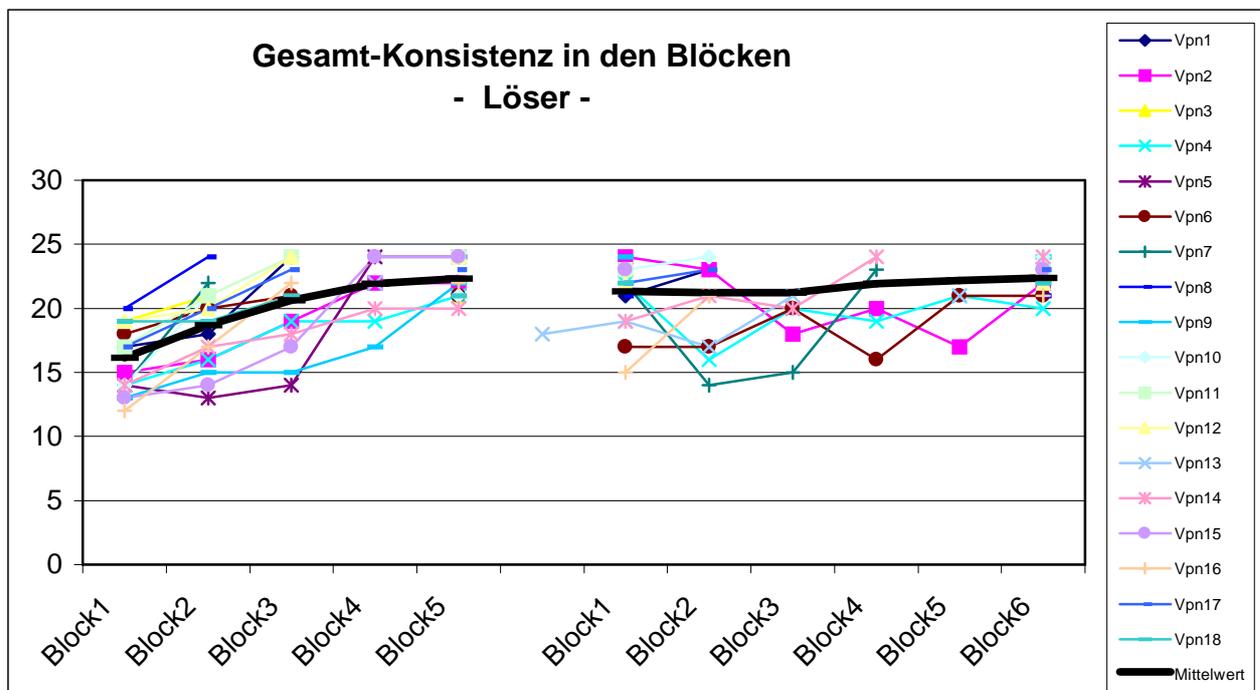


Abb. 2.25: Gesamt-Konsistenz über die Blöcke hinweg.

Tastenzuweisung

Weder die Tastenreaktion auf das 1. Item noch auf das Testitem EF wich signifikant von einer Gleichverteilung ab ($\chi(1)=0.06$, $p>0.10$; $\chi(1)=0.89$, $p>0.10$).

Konsistenz der Tastenzuweisung

14 von den 18 Lösern behielten ihre Kategorienstruktur in Phase 2 bei, 4 wechselten sie ($B_{18;0.5}(4)<0.05$). Dieser Sachverhalt ist in Abbildung 2.26 visualisiert. 8 der 18 Nicht-Löser ordneten die Stimuli der Kategorie 1 im letzten Block der Phase 2 in dieselbe Kategorie wie im letzten Block der Phase 1. Bei 9 Versuchspersonen war dies für Stimuli aus Kategorie 2 der Fall³⁷.

Kategorienwahl

6 der 18 erfolgreichen Löser platzierten den Teststimulus EF in die Kategorie 1, 12 platzierten den Teststimulus wie erwartet in Kategorie 2. Der entsprechende Binomial-Test $B_{18;0.5}(6)$ ergab einen p-Wert von 0.12. Bei der Interpretation dieses Wertes muss aber berücksichtigt

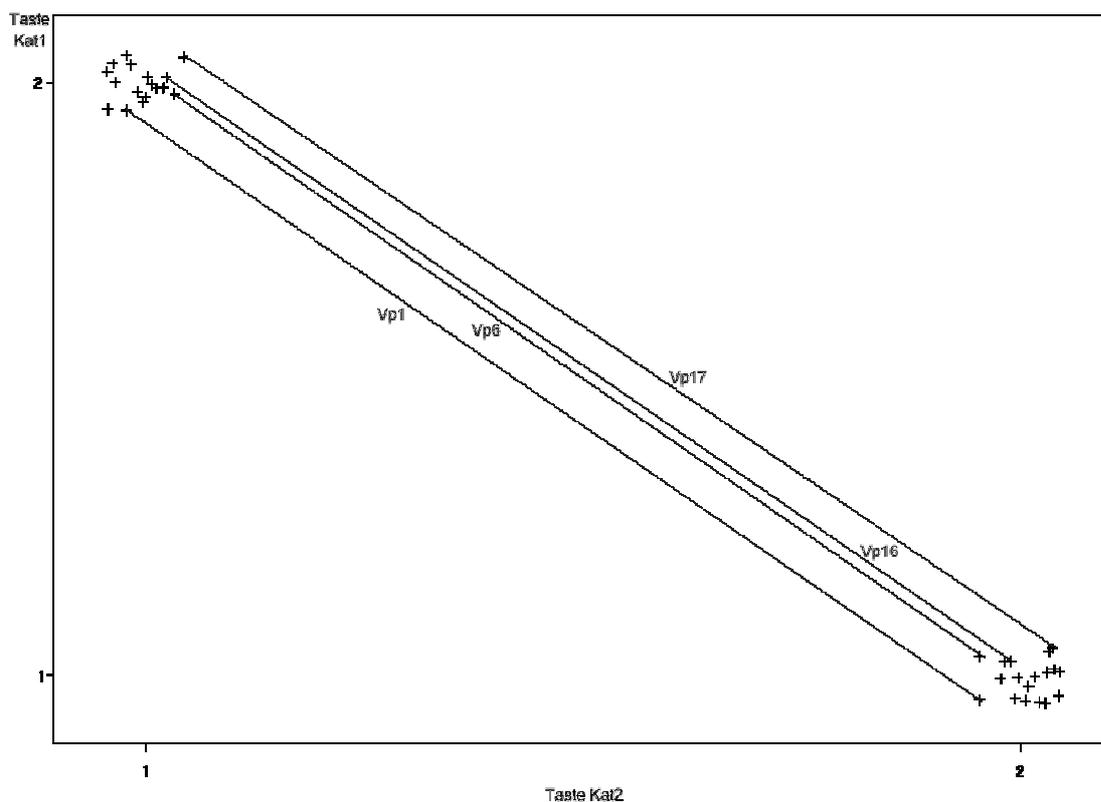


Abb. 2.26: Zuordnung der Tasten zu den Kategorien am Ende der beiden Experimentalphasen für Löser des 3. Experiments.

³⁷ Der Unterschied zwischen Kategorie 1 und 2 bei den Nicht-Lösern beruht darauf, dass mehrere Nicht-Löser Stimuli aus beiden Kategorien derselben Taste zuwiesen (vgl. Abbildung in Anhang 6).

werden, dass die Erwartung war, dass Versuchspersonen überdurchschnittlich häufig Stimulus EF der Kategorie 2 zuordnen sollten. Deshalb konnte ein Design, welches die Kategorie 1 bevorzugte, toleriert werden. So würde z.B. eine Versuchsperson, welche immer dieselbe Taste bei einem „überraschend“ neuen Reiz betätigt, bei dem 1. Stimulus der 2. Phase und bei Testitem EF dieselbe Taste betätigen. Da der 1. Stimulus der 2. Phase immer der Prototyp der Kategorie 1 war, resultierte eine Bevorzugung der Kategorie 1. Auf einen weiteren möglichen Grund für die Bevorzugung der Kategorie 1 hat mich Andy Wills (Email 08.01.2003) hingewiesen. Wenn „schlechte“ Lerner die Prototypen aus beiden Phasen nicht separat repräsentieren, sondern in einer Form der Überlagerung, so ähnelt die Kategorie 1 mit ABE prozentual stärker als CDGF dem Teststimulus EF. Diese Form der Ähnlichkeitsbeurteilung unter Einbeziehung der nicht-gemeinsamen Informationen wird z.B. von Pearce (1994) gefordert. Um diesen Bias in der Testung zu berücksichtigen und um die Anzahl auswertbarer Versuchspersonen zu erhöhen, wurden in den Test der Kategorienzuordnung von Stimulus EF auch Versuchspersonen integriert, welche nicht das Kriterium in Phase 1 oder 2 erreicht hatten. Allerdings mussten die Versuchspersonen, um in die Analyse aufgenommen werden zu können, im letzten Block eine Zuordnungspräferenz der Kategorien zu den 2 Tasten gezeigt haben. Von diesen Versuchspersonen wählten 9 die Kategorie 1 und 4 die Kategorie 2. Abbildung 2.27 zeigt die erhaltenen Zuordnungshäufigkeiten.

Zuordnung EF		
	Kategorie 1	Kategorie 2
Löser	6	12
Analysierbare Nicht-Löser	9	4

Zuordnung AG		
Löser	12	6
Analysierbare Nicht-Löser	5	8

Abb. 2.27: Erhaltene Zuordnungshäufigkeiten der Teststimuli zu den Kategorien.

Ein Chi-Quadrat Test auf Unabhängigkeit war signifikant ($\chi^2(1)=3.89$, $p=0.048$). Da dieser Wert sehr nahe unter dem üblichen Signifikanzniveau lag, wurde ein Fisher-Exact-Test (Siegel 1956) durchgeführt. Der aufsummierte p-Wert betrug damit 0.053. Die genauen p-Werte sind in Tabelle 2.2 aufgeführt.

Tabelle 2.4: p-Werte für die erhaltenen und extremeren Zellhäufigkeiten für Stimulus EF.

p-Wert	Zellhäufigkeiten (a,b,c,d)
0.04416467	9,4,6,12
0.00815348	10,3,5,13
0.00079417	11,2,4,14
3.5296E-05	12,1,3,15
5.0908E-07	13,0,2,16

Siegel (1956) schlägt bei p-Werten, die bei Vernachlässigung des p-Wertes für die tatsächlich erhaltenen Zellhäufigkeiten aufsummiert unterhalb der Signifikanzgrenze liegen, Tocher's Modifikation vor. Tocher's Modifikation berücksichtigt, dass aufgrund der diskreten Werte die Wahrscheinlichkeit für alle noch akzeptierten Verteilungen zusammen weit unterhalb der nominalen Signifikanzgrenze liegt. Der „Anteil“ der beobachteten diskreten Realisierung, der noch unter die Signifikanzgrenze fallen sollte, wird durch ein zweites Losverfahren bestimmt. Dazu werden sämtliche p-Werte für alle Zellhäufigkeiten außer den tatsächlich erhaltenen aufsummiert und von dem gesetzten Signifikanzniveau abgezogen. Der erhaltene Wert wird dann durch den p-Wert der erhaltenen Zellhäufigkeiten dividiert:

$$\frac{\alpha - P_{\text{extremereWerte}}}{P_{\text{beobachteteWerte}}}$$

Eine Korrektur hin zu Kontinuität ist ein übliches Verfahren für Chi-Quadrat Tests (vgl. Bortz, 1979). Falls bei Ziehung einer reellwertigen Zufallszahl zwischen 0 und 1 dieser Wert überschritten wird, wird die erhaltene Zellhäufigkeit als jenseits der Signifikanzgrenze angesehen, andernfalls unter der Signifikanzgrenze. Der so erhaltene Wert betrug für die Daten $\frac{0.05 - 0.009}{0.044} = 0.92$. Eine durch Microsoft Excel gezogene Zufallszahl betrug 0.27 und

blieb damit unter diesem Wert. Es kann also von einer signifikanten Bevorzugung der ungeblockten Kategorie für Stimulus EF ausgegangen werden. Für den Stimulus AG betrug die Anzahl an Zuordnungen zur 1. Kategorie 12 und zur 2. Kategorie 6. Damit wurde wieder derselbe Binomial-Wert wie für Stimulus EF erhalten. Der Chi-Quadrat-Test unter Einbeziehung der auswertbaren Nicht-Löser, von denen 5 AG in die Kategorie 1 einordneten und 8 in die Kategorie 2 ergab keine Auffälligkeit in den Daten, die klar von Zufallsschwankungen abzugrenzen war ($\text{Chi}(1)=2.42$, $p>0.10$). Der Fisher-Exact-Test ergab einen p-Wert von 0.117. Unter Ausschluss der tatsächlich beobachteten Zellhäufigkeit ergab sich die Summe der p-Werte zu 0.027. Die zu ziehende Zufallszahl sollte damit unterhalb von 0.259 liegen. Die Zufallszahl 0.754 blieb darüber. Damit ließ sich der Zuteilungsvorgang zu den Kategorien für Stimulus AG nicht zuverlässig von einer zufälligen Zuweisung unterscheiden.

Anzahl identifizierter Icons

Die Analyse, ob die Klassifizierung aufgrund einzelner Icons erfolgte, ergab 13 Versuchspersonen die anscheinend keinem Icon folgten, 3 die anscheinend ihre Entscheidungen auf 1 Icon basierten, und nur einer Person konnte in einer Phase 2 Icons zugeordnet werden. Die Wahrscheinlichkeit, dass bei Vorliegen der 1-Icon-pro-Kategorie Strategie für höchstens 1 Versuchsperson 2 Icons identifiziert werden konnten war kleiner als 0.05.

Bekanntheitsrating

Die Bekanntheitsratings der geblockten Icons unterschieden sich nicht von den Bekanntheitsratings der ungeblockten Icons ($t(17)=0.00$, $p>0.10$). Abbildung 2.28 zeigt die Aufschlüsselung nach Versuchspersonen.

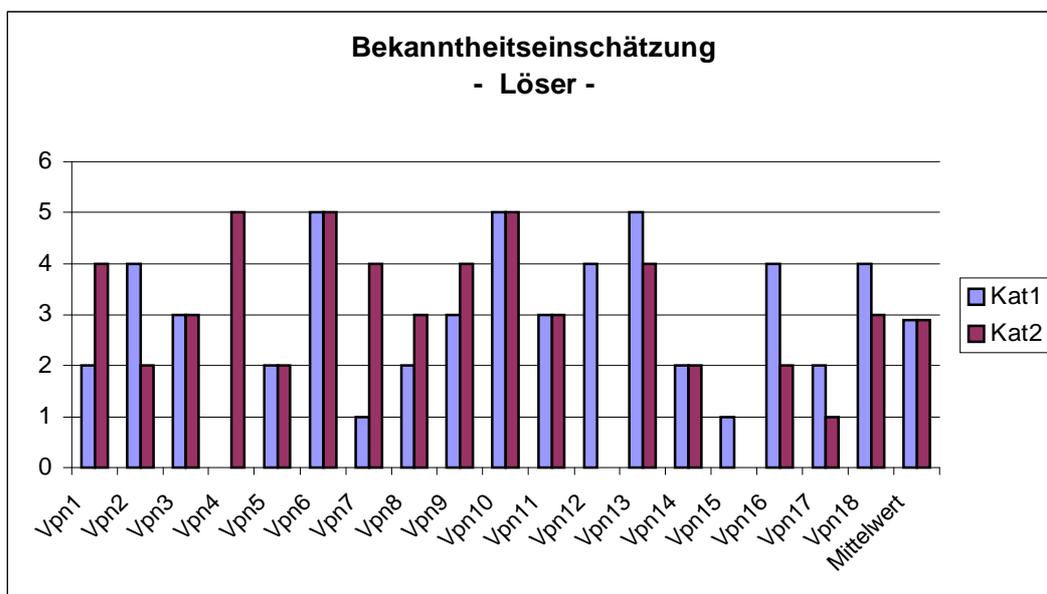


Abb. 2.28: Bekanntheitsrating aufgeschlüsselt nach Versuchspersonen und Kategorienzugehörigkeit der Icons.

Auskünfte der Teilnehmer

Von den Lösern und Nicht-Lösern machten jeweils 3 Teilnehmer keine Angaben zur Klassifizierungsstrategie.

Löser

Die Übereinstimmung zwischen beiden Ratern betrug auf der 1. Beurteilungsdimension 100% ($B_{15;0.5}(0)<0.001$), auf der 2. Dimension gab es 1 unterschiedliche Beurteilung ($B_{15;0.5}(1)<0.001$). Auf den restlichen 2 Dimensionen differierten die Meinungen bei 2 ($B_{15;0.5}(2)<0.01$) bzw. 1 Versuchsperson ($B_{15;0.5}(1)<0.001$).

Nicht-Löser

Bei den Nicht-Lösern unterschieden sich die Rater bei 2 Versuchspersonen ($B_{15;0.5}(2) < 0.01$) bezüglich der 1. und 2. Dimension und bei 0 ($B_{15;0.5}(0) < 0.001$) bzw. 3 ($B_{15;0.5}(3) < 0.05$) Versuchspersonen auf der 3. bzw. 4. Dimension.

Damit zeigten die Rater auch in Experiment 3 eine ausreichende Übereinstimmung um eine Mittelwertbildung ihrer Kategorienhäufigkeiten zu rechtfertigen.

Löser

Den Bemerkungen zufolge hatten 36% der Versuchspersonen die Stimuli anhand von Farben klassifiziert. 46% benutzten nach Einschätzung der Rater 1 einzelnes Icon zur Kategorisierung. 14% der Teilnehmer schienen eine Restkategorie zu bilden und 11% der Versuchspersonen wurden als Strategienwechsler eingestuft. Der Zusammenhang zwischen der Benutzung von Farben als Beurteilungsgrundlage und dem Beachten mehrerer Icons war nicht signifikant (Fisher-Exact-Test, $p > 0.10$).

Nicht-Löser

79% schienen die Stimuli aufgrund der Farben einzuordnen, 21% benutzten den Ratern zufolge nur 1 Icon. Aus keiner der Versuchspersonenbeschreibungen folgerten die Auswerter die Bildung einer Restkategorie. Nur in 11% der Fälle stuften die Rater eine Versuchsperson als Strategienwechsler ein. Der Zusammenhang zwischen den ersten zwei Dimensionen war marginal signifikant (Fisher-Exact-Test, $p < 0.10$).

Vergleich Löser und Nicht-Löser

Ein Unterschied zwischen den Lösern und Nicht-Lösern bezüglich der Häufigkeit mit der sie die dargebotenen Stimuli gemäß der Farben zu gruppieren schienen, ließ sich auf dem 5% Niveau absichern ($\chi^2(1) = 4.82$, $p < 0.05$). Die Löser und Nicht-Löser differierten nicht signifikant hinsichtlich der Nutzung eines einzelnen Icons als Klassifizierungsgrundlage ($\chi^2(1) = 1.89$, $p > 0.10$). Auch auf den restlichen 2 Dimensionen unterschieden sich die Lösern nicht wesentlich von den Nicht-Lösern (Fisher-Exact-Test, $p > 0.10$; Fisher-Exact-Test, $p > 0.10$).

Vergleich Löser Experiment 1-3

Es zeigte sich kein verlässlicher Unterschied zwischen den Lösern der Experimente 1-3 hinsichtlich der Rating-Dimensionen ($\chi^2(2) = 1.03$; $\chi^2(2) = 1.56$; $\chi^2(2) = 0.91$; $\chi^2(2) = 0.31$; für alle: $p > 0.10$; bei den letzten beiden Tests waren die Zellhäufigkeiten < 5).

2.4.2 Diskussion Experiment 3

Um die empirischen Resultate durch den Blocking-Effekt auf Elementebene zu erklären, muss angenommen werden, dass die einzelnen Icons platzunabhängig verarbeitet wurden. Andernfalls müsste das Lernen des Vorhersagewertes eines Icons für unterschiedliche Positionen getrennt erfolgt sein, da die Icons zufällig ihre Plätze tauschten. Experimente von Treisman (1988) machen die Annahme der positionsunabhängigen Verarbeitung aber plausibel. Treisman zeigte, dass zwei verschiedene Merkmale zweier kurz dargebotener Stimuli vertauscht werden können, so dass das Merkmal eines Stimulus mit einem anderen Stimulus assoziiert wird. Erst ein Bindungsprozess scheint also Merkmale an einem Ort zu Objekten zusammenzufassen. Ebenso werden vermutlich die Icons eines Stimulus zuerst unabhängig von ihrer Position verarbeitet.

Ein erstaunliches Ergebnis stellt die Tatsache dar, dass unabhängig vom genauen Versuchsdesign über alle 3 Experimente hinweg ca. 50 % der Teilnehmer die Aufgabe lösen konnten und 50 % nicht. Dies könnte auf zwei Arten in das diskutierte Blocking-Modell integriert werden. Eine Möglichkeit besteht darin, ein Persönlichkeitsmerkmal anzunehmen, welches bei 50% der Teilnehmer vorhanden ist und zu einer Klassifizierung aufgrund gemeinsamer Icons führte. Nur für diese Probanden wäre das oben diskutierte Delta-Modell angemessen. Als alternative Erklärung bietet sich an, dass die erfolbringende Strategie unabhängig von Persönlichkeitsmerkmalen von 50% der Teilnehmer zufällig gewählt wurde. So relevant es auch für ein tieferes Verständnis von Klassifizierungsvorgängen und der dabei auftretenden Probleme sein mag, dieser Frage nachzugehen, ist sie für das Untersuchungsziel dieser Studie irrelevant, da es in dieser Studie darum ging, mit Hilfe des Blocking-Effektes zwischen zwei grundlegenden Modellvorstellungen zu unterscheiden. Der *Nachweis* war in dieser Studie bedingt durch das experimentelle Design nur dann möglich, wenn tatsächlich aufgrund gemeinsamer Icons klassifiziert wurde, die Faktoren, welche zu dieser Klassifizierungsstrategie führten spielten aber für die Fragestellung keine Rolle. Dass sich der Blocking-Effekt in diesen Experimenten nur bei Klassifizierung aufgrund gemeinsamer Icons nachweisen lässt, heißt nicht, dass der Blocking-Effekt für andere Klassifizierungsstrategien keine Bedeutung hat, sondern nur, dass diese Strategien mit einem gesonderten Design untersucht werden sollten.

Tastenzuordnung

Im Gegensatz zu den ersten beiden Experimenten bestand auch für den 1. Stimulus keine Bevorzugung einer Taste über die Versuchspersonen hinweg. Dies mag daran liegen, dass bei 2 Tasten eine kleinere Asymmetrie empfunden wird als bei 3 Tasten. Der zugrundeliegende Mechanismus, welcher einer Taste eine bestimmte Funktion zuweist, verdient in weiteren Studien untersucht zu werden, um eventuelle Zusammenhänge zwischen räumlichen und funktionalen Vorstellungen zu verstehen.

Reaktionszeit

Die Reaktionszeit nahm in Experiment 3 über die Phasen hinweg nicht mehr signifikant ab. Dies zeigt, dass das Zeitlimit den Versuchspersonen nicht mehr erlaubte, ihre Reaktionszeit durch Wahl einer schnelleren Entscheidungsstrategie bzw. durch Synergieeffekte aufgrund Übung, weiter zu senken. Dieses Zeitlimit scheint also ein Mindestmaß für strukturierte Entscheidungen dieser Art darzustellen.

Der Zeitunterschied zwischen Item EJO bzw. EF und dem unmittelbar vorhergehenden Item war für Experiment 1, 2 und 3 einzeln betrachtet nicht signifikant. Dies kann vermutlich auf die sehr geringe Teststärke aufgrund der zu geringen Anzahl an Messungen bei einer so variablen Größe wie der Reaktionszeit zurückgeführt werden. Betrachtet man aber die Reaktionszeiten auf die beiden Items über alle Experimente hinweg, so zeigt sich ein signifikanter Unterschied. Der theoretisch postulierte leichte Konflikt des Stimulus EJO bzw. EF, der widersprüchliche Verbindungen zu 3 bzw. 2 Kategorien besitzt, scheint sich damit empirisch zu bestätigen. Theoretisch wäre zu erwarten, dass Stimulus AHK/AG eine größere Verlangsamung erfahren sollte als EJO/EF. Dies konnte allerdings statistisch nicht abgesichert werden. Ein Problem dieser Testung besteht darin, dass die Verlangsamung auch auf die Neuheit der Teststimuli zurückgeführt werden könnte. Eine Testung gegen den 1. neuen Stimulus in Phase 2 würde aber der allgemeinen Abnahme der Reaktionszeit über die Phase hinweg nicht gerecht. Um zu untersuchen, ob die „Neuheit“ eines Stimulus ausreicht, die Reaktionszeit signifikant zu verlangsamen, wurde die Reaktionszeit auf das letzte Item der 1. Phase gegen die Reaktionszeit auf das 1. Item der 2. Phase getestet. Hier ergab sich für das 1. Experiment eine signifikante Zunahme der Reaktionszeit ($t(14)=-3.42$, $p<0.01$) und eine marginal signifikante Zunahme für Experiment 2 und 3 ($t(16)=-1.77$, $p<0.10$; $t(17)=-2.44$, $p<0.10$) und schließlich über alle Versuchspersonen betrachtet eine signifikante Zunahme ($t(49)=-2.44$, $p<0.05$). Der Vergleich des 1. Items der 2. Phase mit dem letzten Item der 1. Phase stellt aber nur dann ein angemessener Bezug für den Vergleich zwischen den Testitems und dem letzten Item der 2. Phase dar, falls das höhere

Lernniveau am Ende der 2. Phase nicht zu einem geringeren Anstieg der Reaktionszeit auf neue Stimuli führt und falls die Teststimuli als so neuartig wie der 1. Stimulus der 2. Phase angesehen werden können. Dagegen spricht, dass bei dem 1. Stimulus der 2. Phase *neue* Icons auftreten, während die Teststimuli nur *bekannte* Icons, wenn auch in einer neuen Zusammenstellung, enthalten. Die Frage des Einflusses der Neuheit der Teststimuli auf die verlängerte Reaktionszeit kann daher auf der gegenwärtigen Datenbasis nicht geklärt werden. Damit ist die Rückführung der längeren Reaktionszeit auf einen größeren Konflikt zwischen den widersprüchlichen Informationen bei den Teststimuli fragwürdig.

Konsistenz

Im Gegensatz zu den ersten beiden Experimenten fällt die Zunahme an Konsistenz in der 2. Phase nicht mehr signifikant aus. Augenscheinlich gelingt es den Versuchspersonen ihr Wissen über die Kategorien mit in die 2. Phase zu übernehmen, so dass kein weiterer Anstieg mehr zu messen ist.

Kategorienzuordnung

Die Kategorienpräferenz für Teststimulus EF entsprach der theoretisch erwarteten. Netzwerksimulationen mittels des oben beschriebenen Delta-Netzes (siehe Anhang 9, Delta-Netzwerk II) mit und ohne Einfluss der Salienz auf die Performance bestätigten die Erwartung, dass das Delta-Modell für weite Parameterbereiche EJO bevorzugt der nicht-geblockten Kategorie zuordnen sollte. Abbildung 2.29 zeigt die erhaltenen Simulationen für das Delta-Netzwerk II. Der Rumelhart und Zipser Algorithmus konnte hingegen die empirischen Daten nicht erklären (vgl. Abbildung 2.30).

Erneut konnte das Delta-Modell sowohl die empirisch erhaltene Bevorzugung der 1. Kategorie als Reaktion auf Stimulus AG als auch die signifikante Präferenz Stimulus EF in Kategorie 2 einzuordnen simulieren (Abbildung 2.31). Für das Delta-Modell mit dem Salienzfaktor in der Entscheidungsregel zeigte sich kein qualitativer Unterschied (siehe Anhang 10e).

Die Bevorzugung der Kategorie 1 bei Präsentation von Stimulus AG entstand dadurch, dass die vorhergehende Präsentation des Stimulus EF die Assoziation zwischen G und Kategorie 2 abschwächte. Der Grund hierfür war, dass EF der Kategorie 2 zugeordnet wurde und damit alle Verbindungen der nicht aktiven Icons zu Kategorie 2, also auch die Verbindungen der G-Icons, vermindert wurden. Da E in Gruppe 1 (bzw. F in Gruppe 2) keine starke Assoziation zu Kategorie 2 hatte, entstand eine große „Überraschung“ und damit eine große Gewichtsänderung, bzw. Verminderung der Verbindungen von G zu Kategorie 2.

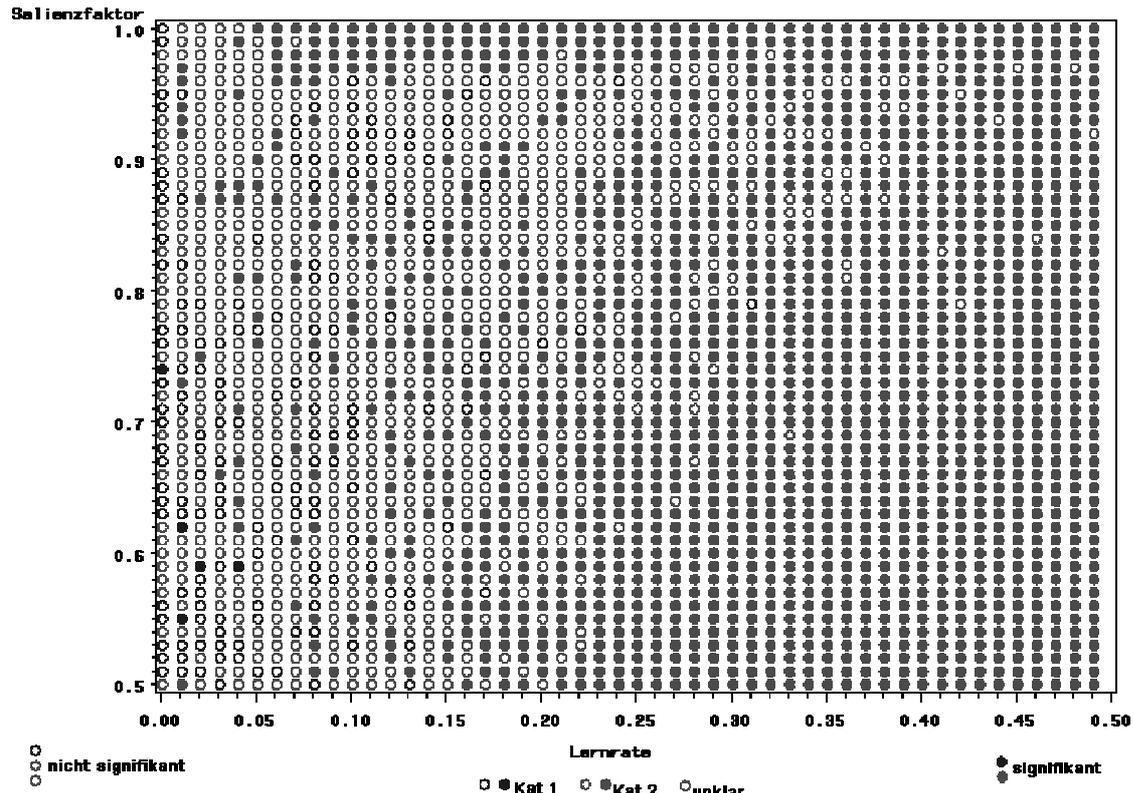


Abb. 2.29. Delta-Algorithmus ohne Einfluss der Salienz auf die Performanz.

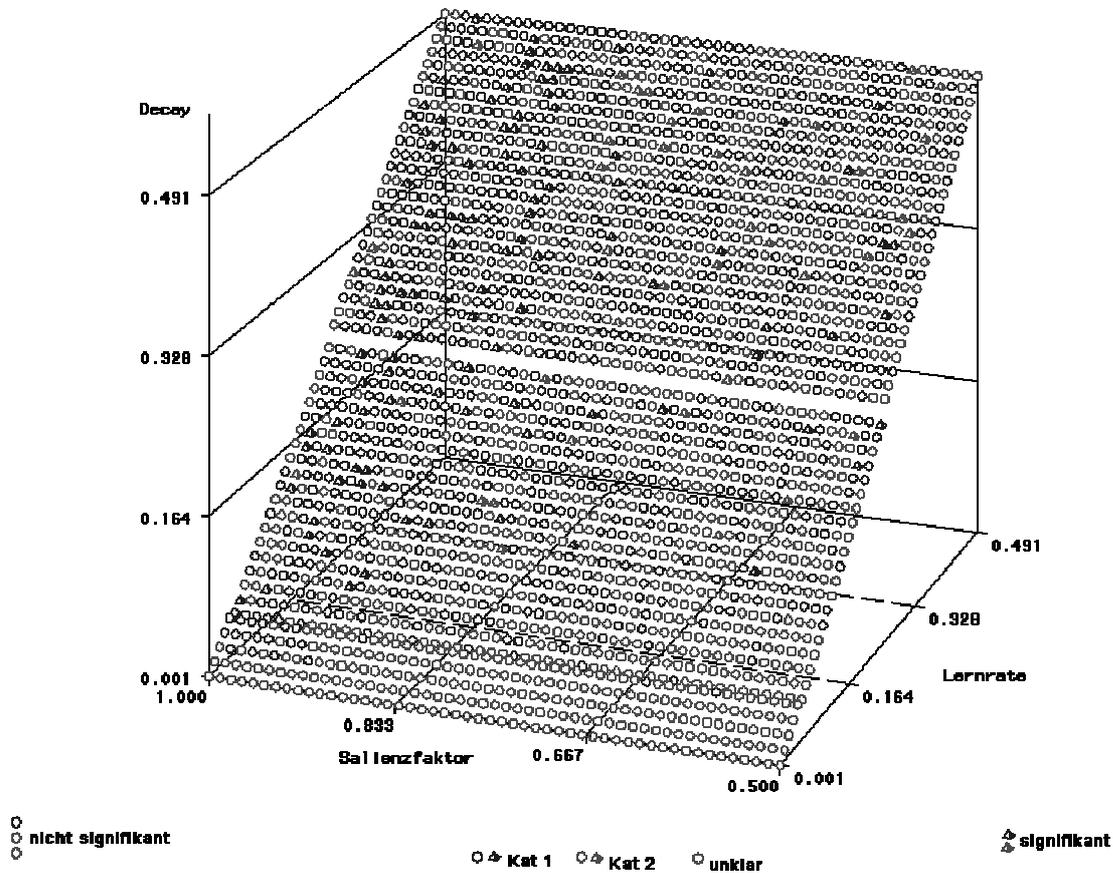


Abb. 2.30: Rummelhart & Zipser Algorithmus mit Salienzparameter. Die Entscheidung für eine Kategorie erfolgte durch die ratio rule mit Parameter $\Theta=20$.

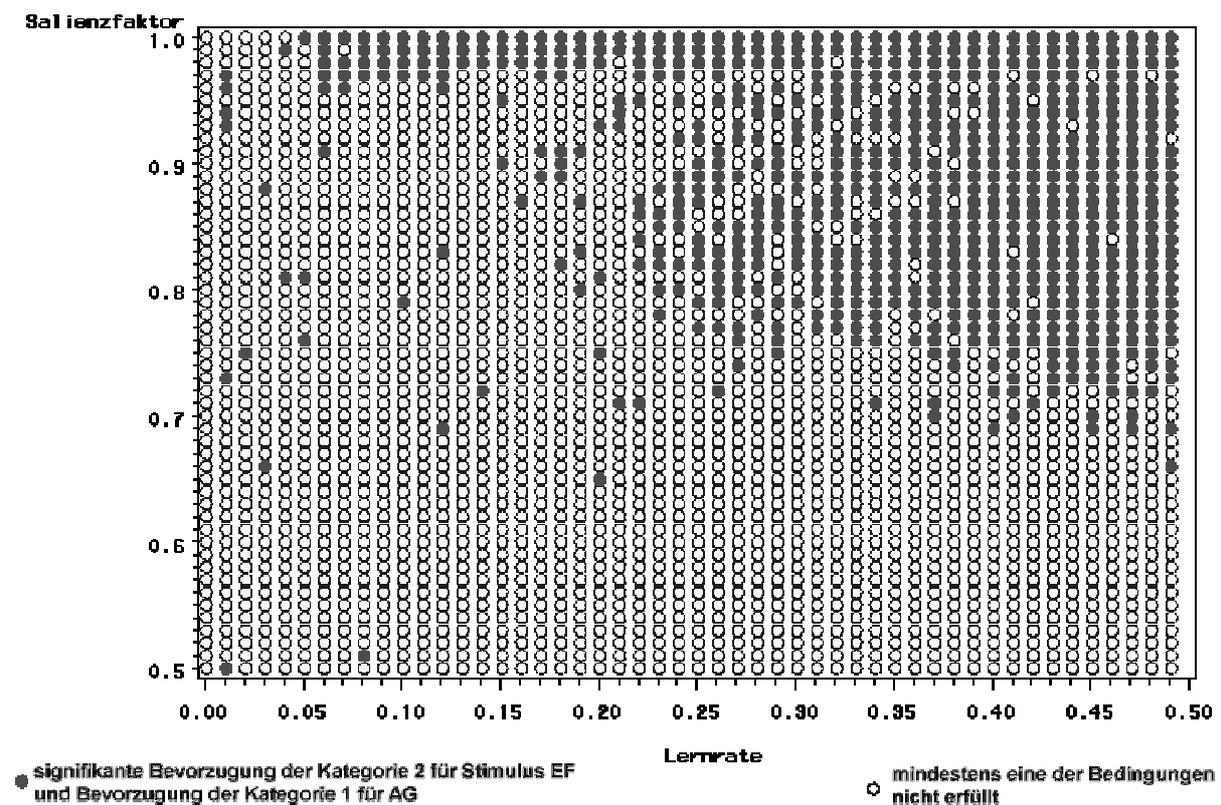


Abb. 2.31. Delta-Netzwerk II. Gefüllte Punkte symbolisieren eine signifikante Bevorzugung der Kategorie 2 für Stimulus EF und eine Bevorzugung der Kategorie 1 für AG.

Da ansonsten A und G relativ³⁸ gleiche Verbindungsstärken besaßen, reichte diese kleine Asymmetrie aus, Kategorie 1 zu bevorzugen. In Experiment 1 und 2 war die Differenz zum Zielwert, im Sinne der asymptotisch zu erreichenden maximalen Verbindungsstärke, geringer. Außerdem wurden nur die Verbindungen zu Kategorie 3 geschwächt. Da A und H nicht Kategorie 3 zugeordnet waren, wurde ihr Einfluss bei Teststimulus AHK nicht vermindert. Dadurch wurde lediglich Kategorie 3 als Reaktion auf Teststimulus AHK „abgelehnt“.

Erneut zeigte sich eine Dissoziation zwischen dem Blocking-Effekt gemessen durch die Kategorienzuordnung und gemessen durch die Bekanntheit der einzelnen Icons. Dies ist erneut konsistent mit der Vermutung, dass die Dissoziation durch kürzere Reaktionszeiten verstärkt wird. Der geringere Unterschied in der Bekanntheitseinschätzung verglichen mit Experiment 2 ist auch mit der Hypothese vereinbar, dass unter Zeitdruck eine eher konfigurale Analyse vorherrscht (Smith & Kehler Nelson, 1984) und deshalb einzelne Icons eine geringere Ähnlichkeit zu den Stimuli der Kategorien besitzen.

³⁸ A konnte auch in Phase 2 eine stärkere Assoziation als G aufbauen, da A bereits zu Beginn von Phase 2 eine Verbindungsstärke besaß. Allerdings konnte dieser Effekt durch eine starke Salienzabnahme ausgeschaltet werden.

Eine spekulativere Erklärung ist, dass für die Einschätzung der Bekanntheit die Icons hinreichend lange präsentiert werden mussten, um die Bildung von Gedächtnisspuren zu ermöglichen. In diesem Fall wäre im kürzesten Darbietungsmodus (Exp. 3) der geringste Unterschied in der Einschätzung der Bekanntheit zu erwarten, da alle Gedächtnisspuren relativ schlecht ausgebildet sind. In Experiment 1 wäre ein nur marginal größerer und in Experiment 2 aufgrund der langen Darbietungszeit der größte Unterschied zu erwarten.

Die Analyse der Anzahl identifizierbarer Icons ergab, dass weniger Teilnehmer 1 Icon "folgten" als dies in Experiment 1 und 2 der Fall war. Dieses Ergebnis bestärkt die Erwartung, dass sich die Urteile unter Zeitrestriktion an globaleren Merkmalen als einzelnen Icons orientieren (Smith & Kemler Nelson, 1984).

Auskünfte der Teilnehmer

Die Befragung erbrachte ein zu den anderen Experimenten sehr ähnliches Ergebnis. Erneut zeigte sich die scheinbare Insensitivität der Befragung, da hinsichtlich der 4 ausgewerteten Dimensionen keine Unterschiede zwischen den Lösern der 3 Experimente zu finden waren, gleichzeitig sich die Kategorienzuweisungen für den 1. Teststimulus aber unterschieden. Keine berichtete Strategie konnte die empirische Bevorzugung einzelner Kategorien in dem ausbalancierten Design begründen. Es zeigte sich in allen 3 Experimenten, dass der Anteil an Versuchspersonen, deren Äußerungen als Beschreibungen einer Klassifizierung aufgrund von Farben gewertet wurden, bei den Nicht-Lösern wesentlich höher als bei den Lösern war. Dies ist auch unmittelbar einsichtig, da die Beurteilung nach Farben vermutlich nur in wenigen Fällen die Prototypen voneinander trennen kann. Aufgrund der zufälligen Zusammensetzung der Prototypen sollten die Farben in allen Prototypen annähernd gleich verteilt sein. Eventuell stellt die relativ willkürliche Entscheidung, die Klassifizierung auf Farben oder Formen zu basieren das Merkmal dar, welches die Teilnehmer der Studien in Löser und Nicht-Löser aufteilte.

Konsistent mit der Studie von Zwickel und Wills (2002) zeigte sich das durch den Blocking-Effekt theoretisch vorhergesagte Ergebnis. Der schwächere Effekt verglichen mit der Studie von Zwickel und Wills (2002) lag vermutlich an kleinen Unterschieden im Versuchsdesign. Zwickel und Wills „blockten“ die betreffenden Buchstaben in 2 Blöcken durch unterschiedliche Buchstaben. Gleichzeitig entfiel in diesen hier berichteten 3 Experimenten jegliche Form der Rückmeldung, was die Unsicherheit bei den Versuchspersonen beträchtlich erhöht haben dürfte.

2.5 Allgemeine Diskussion

Wie die Studien gezeigt haben sollten, scheint das Delta Modell das sichtbare Klassifizierungsverhalten von Menschen unter Zeitdruck mit wenigen Parametern gut zu beschreiben. Offen bleibt allerdings, inwieweit diese einfachen Assoziationsprozesse kognitive Vorgänge wiedergeben. Die Ergebnisvorhersage ist zwar richtig, aber ob das Delta-Modell auch ein gutes Prozessmodell für den Ablauf der kognitiven Vorgänge darstellt, ist unklar. Während simple assoziative Prozesse Tierverhalten gut widerspiegeln und vermutlich auch für die Modellierung menschlichen Verhaltens unter Zeitdruck anwendbar sind, bezweifle ich, dass assoziative Modelle deliberate Entscheidungen gut beschreiben. Bei dilibierten Entscheidungen dürfte nicht so sehr die oberflächliche Ähnlichkeit eines Stimulus zu einem Prototypen, sondern eine „abstraktere“ Ähnlichkeit die Reaktion bestimmen. So wäre eine Entscheidungsregel „wenn oben links eine Trompete zu sehen ist“ eine sehr oberflächliche Ähnlichkeitsbestimmung. Abstrakter wäre die Formulierung „Falls der Stimulus eine Trompete enthält“. Beliebige Steigerungen sind dabei denkbar. Als Operationalisierung dieser Abstraktheit ließen sich Maße für die Veränderungen, welche angenommen werden müssen um aus den Netzhautdaten das Ähnlichkeitsmaß zu erhalten, denken. Natürlich sind Netzwerkmodelle auch für solche abstraktere Ähnlichkeitsmodelle, die von den objektiv gegebenen physikalischen Ähnlichkeiten abstrahieren, möglich. Allerdings erfordern diese dann bereits eine abstraktere Eingabekodierung, welche die relevanten Ähnlichkeitsmerkmale ähnlichkeitsstrukturertretend auf die Eingabeschicht des Netzwerkes abbildet. Dies spricht für einen deutlich größeren Rahmen in den das Delta-Netzwerk eingebettet werden sollte.

Metakognitive Prozesse wie z.B. die deliberate *Entscheidung* nur ein Icon zu beachten, können eventuell besser mittels Regeln beschrieben werden, die einzelne Icons herausgreifen und nicht die Strategie auf alle Icons anwenden. In diesem Fall wäre die Selektion der Klassifizierungsgrundlage aber nicht mehr durch einfache auf alle Icons wirkende Sequenzeffekte bedingt, sondern durch eine Regel, deren Ursprung unklar ist, spezifiziert. Auch bei Regeln können Blocking-Effekte eine Rolle spielen, indem festgelegt wird, auf welchen Icons eine Regel operiert. So könnte eine Regel „bei einer Mehrzahl eckiger Icons Taste 1“ nur auf bereits bekannte Icons angewendet werden. Damit würde neue Information ebenfalls geblockt. Eine andere Form des Blocking-Effektes bei Regeln könnte darin bestehen, dass eine gefundene Regel so lange die Suche nach weiteren Regeln verhindert, wie sie die konsistente Klassifizierung erlaubt.

Rieskamp und Hoffrage (1999) sind der Ansicht, dass unter Zeitdruck eine stärkere Selektion der Informationen erfolgt. Dies ist konsistent mit der hier vertretenen Vorstellung, dass der Blocking-Effekt unter Zeitdruck stärker ausgeprägt sein sollte. Der hier vertretenen Ansicht zufolge tritt der Blocking-Effekt unter Zeitdruck vor allem auf einer sehr niedrigen Repräsentationsebene (in diesem Fall die Repräsentationsebene der Icons) auf. Wird der Zeitdruck aufgehoben, kann sich der Blocking-Effekt auch auf einer höheren Repräsentationsebene wie z.B. bei Hypothesen zeigen. Allerdings wird dann bei einer längeren Verarbeitungszeit der Reduktion der Information keine so hohe Priorität mehr eingeräumt und der Blocking-Effekt durch weitere deliberate Prozesse vermindert.

Lernprozesse im ursprünglichen Delta-Modell erfolgen relativ unabhängig von den spezifischen Merkmalen wie z.B. der Präsentationshäufigkeit eines Reizes. Es scheinen Aspekte wie Gewöhnung und Ermüdung zu fehlen. Dies wurde versucht durch die Aufnahme eines Salienzfaktors zu verbessern. Die Aufnahme von Aufmerksamkeit bzw. Salienz in das Modell scheint den Übergang von vorwiegend *stimulus*getriebenen Tiermodellen hin zu *verarbeitung*getriebenen (kognitiven) Modellen darzustellen. Allerdings fehlt immer noch ein theoretischer Rahmen für eventuell auftretende metakognitive Strategien. Die Dissoziationen zwischen Bekanntheitseinschätzung und Klassifizierungsverhalten konnte mit diesem Modell deshalb nicht repliziert werden, weil die notwendigen Unterscheidungen in der Wissensrepräsentation bzw. im Antwortverhalten fehlen. Sowohl Bekanntheitseinschätzungen als auch Kategorienwahlen basieren im vorliegenden Modell ausschließlich auf den assoziativen Verbindungsstärken und einem eventuell einzuführenden Salienzfaktor. Es lassen sich natürlich komplexere Modelle vorstellen, die dieses Netzwerk integrieren und damit diese Mängel beseitigen könnten.

Diese Arbeit diene vor allem dem Ziel auch für Klassifizierungslernen ausgearbeitetere Modelle zu entwickeln und dabei auf den Fundus von Kategorisierungsmodellen zurückzugreifen. Die in dieser Arbeit getroffenen Modellvorstellungen sind mit dem Modell von Pearce und Hall (1980), in dem die Salienz genutzter Informationen abnimmt, vereinbar. Allerdings müsste das Modell noch genauer hinsichtlich mehrerer Kategorien ausformuliert werden. Die Ausformulierung für die Unterscheidung nur zweier Zustände stellt ein generelles Problem der oben diskutierten Lernmodelle aus Tierstudien und kein spezielles des Pearce und Hall Modells dar. Die geforderten Mechanismen sollten auch in das Modell von Kruschke und Johansen (1999) integrierbar sein, da auch hier davon ausgegangen wird, dass die Aufmerksamkeit für bekannte bzw. bereits genutzte Informationen abnimmt. Der hinter dieser Modellvorstellung stehende Gedanke ist, dass bei Auftreten

einer neuen Kategorie der Fehler durch Nutzung von Informationen, welche mit einer anderen Kategorie verbunden sind, vergrößert werden würde. Gleichzeitig verhindert die Abnahme der Aufmerksamkeit für bereits erfolgreiche Reize, dass durch Umlernen Wissen verloren geht.

Die Idee der abnehmenden Salienz widerspricht hingegen der Grundidee des Mackintosh Modells (1975), welches eine Zunahme der Salienz für genutzte Informationen vorhersagt³⁹. Damit lässt sich die Klasse der mit den hier vorgebrachten Hypothesen konsistenten Modelle einschränken. Generell sollten sich alle Modelle, welche eine Form der Fehlerrückmeldung besitzen, und die eine Abnahme der Salienz für gesehene Informationen postulieren, für eine Implementation der hier vorgestellten Ideen eignen. Die gefundenen Ergebnisse sind hingegen problematisch für regelbasierte Systeme. Zwar können regelbasierte Kategorisierungsmodelle generell Blocking-Effekte erklären, vermutlich aber nicht in diesem Versuchsdesign. Die Bevorzugung einer Kategorie in diesem Design beruht auf der Salienz einzelner Icons, welche in regelbasierten Systemen normalerweise nicht modelliert wird. So dürfte es schwierig sein, eine Regel zu finden, welche zwischen den Kategorien hinsichtlich der Stärke des Blocking-Effektes unterscheiden kann. Diese Regel müsste die Salienz einzelner Reize und in welchem Ausmaß bereits Wissen vorhanden ist berücksichtigen. Diese quantitative Modellierung des Vorwissens dürfte sich aber eher in Modellvorstellungen konnektionistischer Art als in regelbasierte Systeme integrieren lassen.

Gewisse Unsicherheit besteht noch im Wissen über die Repräsentation der relevanten Informationen bei Blocking-Experimenten. Laut Williams, Sagness und McPhee (1994) tritt der Blocking-Effekt nur dann in Erscheinung, wenn die Informationen separat „verarbeitet“ werden. Im Falle einer konfiguralen Analyse wird jeder Stimulus als ideographisch wahrgenommen. Deshalb kann erworbenes Wissen bei konfiguralen Analysen nicht den Erwerb weiteren Wissens blockieren. Da in den hier vorgestellten Experimenten ein selektiver Lerneffekt beobachtet werden konnte, müssen die Versuchspersonen Williams, Sagness und McPhee (1994) zufolge die Stimuli in für Blocking relevante Teile aufgegliedert haben. Dass die kleinsten Einheiten aber nicht einzelne Icons darstellten, legt die 1-Icon-pro-Kategorie Analyse in der Diskussion von Experiment 2 in Abschnitt 2.3.2 nahe. Eine mögliche Integration könnte durch die Annahme entstehen, dass sich Lernvorgänge und

³⁹ Kruschke (2001) konnte zwar die starke Ähnlichkeit zwischen den Modellen von Kruschke und Johansen (1999) und Mackintosh (1975) zeigen, was dafür sprechen würde, dass obige Modellvorstellungen auch mit Mackintosh (1975) vereinbar wären. Allerdings zeigte Kruschke (2001) diese Ähnlichkeit nur für einen Ausgabeknoten. Bei mehreren Ausgabeknoten wären weitere Annahmen nötig wie z.B., dass die Salienz eines Reizes auch durch dessen „Geschichte“ mit einem anderen Ausgabeknoten bestimmt wird, was der „lokalen“ Intention der Salienzregel im Mackintosh (1975) Modell widersprechen dürfte.

Entscheidungsprozesse hinsichtlich ihrer verwendeten Repräsentationsebene unterscheiden können. Dieser Modellvorstellung folgend, *zeigte* sich das Lernen bei Williams, Sagness und McPhee nicht auf elementarer Ebene, weil in ihren Studien die Versuchspersonen genügend Zeit für die Entscheidung hatten. Daher basierten die Versuchspersonen ihre Entscheidungen nicht so sehr auf die assoziativen Verknüpfungen, welche auch bei als konfigural *wahrgenommenen* Stimuli die Lernphase dominieren sollten. Vielmehr entschieden sich die Versuchspersonen aufgrund ihrer Erinnerung an einzelne konfigurale Exemplare und aufgrund weitere deliberater Strategien. Falls die Annahme stimmt, dass auf einer frühen Verarbeitungsstufe auch konfigurale Elemente elementar repräsentiert und gelernt werden, könnte dies den Blocking-Effekt auch bei eigentlich als konfigural wahrgenommenen Stimuli unter Zeitdruck erklären. Bekräftigt wird diese Annahme durch die Studie von Treisman (1988), in der unter Zeitdruck bzw. bei kurzen Präsentationszeiten elementare Repräsentationen von Objekten vertauscht wurden.

In Zukunft sollten weitere Fragen der Generalität des Blocking-Effektes untersucht werden. So stellt z.B. die Blockierung redundanter Information dann einen ökonomischen und damit Selektions-Vorteil dar, wenn die Umwelt als relativ statisch betrachtet werden kann. In diesem Fall tritt die geblockte Information tatsächlich sehr häufig mit der genutzten Information auf. In einer variableren Umwelt hingegen könnte sich der Blocking-Effekt schnell zu einem Nachteil entwickeln, da wichtige neue Informationen nicht mehr gelernt werden, sondern an veralteten Entscheidungsstrategien festgehalten wird.

Bezüglich der Generalisierbarkeit stellt sich die Frage, ob für die Entscheidung welche Informationen geblockt werden, vererbtes Wissen oder Lernerfahrungen eine Rolle spielen. Seligman (1970) konnte zeigen, dass gewisse Assoziationen besser oder schlechter in Abhängigkeit von der Art des Stimulus gelernt werden. So attribuieren Ratten den Grund für Übelkeit nicht auf alle früheren Ereignisse, sondern besitzen die Fähigkeit nach plausiblen Ursachen zu suchen. So könnte es sein, dass ontogenetisch oder phylogenetisch gelernt wurde, welche Informationsart häufig redundant ist und deshalb ignoriert werden sollte um Ressourcen zu sparen. Auch Pearce und Hall (1980) diskutieren biologische Dispositionen als Einflussvariablen auf die Konditionierbarkeit.

Die 1-Icon-pro-Kategorie Analyse konnte nicht zwingend widerlegen, dass einzelne Icons die Klassifizierungsgrundlage darstellen. Wenn das 2-Stufen Modelle von Ahn und Medin (1992) auch nicht der Einordnung der Teststimuli erklären kann, so ist es doch mit den Daten der 1-Icon-pro-Kategorie Analyse konsistent. Aus diesem Grund sollte in weiteren Studien abgesichert werden, dass die Entscheidung nicht doch auf einzelnen Icons im Sinne

des Modells von Ahn und Medin basiert. Ein möglicher Test bestände darin, nach erfolgreichem Lernen dreier Kategorien systematisch ein Icon in einen vollkommen neuen Stimulus einzusetzen. Da hierbei keine widersprüchliche Information auftritt, sollte das einzelne Icon die Kategorisierung bestimmen.

Eine unabhängige Validierung des gefundenen Blocking-Effektes bei Klassifizierungsaufgaben könnte durch Einsatz moderner, weniger reaktiver Verfahren geschehen. So würde es sich anbieten mittels Eye-tracking zu untersuchen, ob Fixationen seltener bei als geblockt eingestuften Icons auftreten als dies bei theoretisch ungeblockten Icons der Fall ist.

Der Nachweis des Blocking-Effektes hat auch Auswirkungen auf andere psychologische Forschungsbereiche. In klinischen Fragestellungen könnte der Blocking-Effekt z.B. gut erklären, warum Lernerfahrungen in der Kindheit einen größeren Einfluss auf das Individuum ausüben als spätere Lernerfahrungen. In diesem Fall würden die in der Kindheit gelernten Gründe und Verhaltensweisen ausreichen, um den Erwerb neuen Wissens zu blockieren.

Chomsky (1959) begründete seine Forderung nach angeborenem grammatikalischem Wissen damit, dass die Umwelt allein zuwenig Information enthalte, um bestimmte Lernaufgaben zu lösen. Elman (1990) hingegen konnte zeigen, dass in einfachen natürlichsprachlichen Sätzen genügend Informationen vorhanden sind, um die syntaktischen Kategorien nachfolgender Wörter aus statistischen Regularitäten früherer Sätze vorhersagen zu können. Diese Diplomarbeit gliedert sich ein in den Versuch nachzuweisen, dass die Rechtfertigung von angeborenem Wissen damit, dass die Umwelt zuwenig Informationen enthält, die Informationen, welche aus der Umwelt vorliegen, unterschätzt. Das vorgestellte Modell versucht sowohl vorhandenes Wissen, in Form von Rückmeldungen aus der Umwelt, zu integrieren als auch bei Fehlen externer Informationen eine Entscheidung zu treffen, die konsistent mit früheren Entscheidungen ist. Dies erlaubt nicht nur eine optimale Nutzung des Wissens sondern hat auch aus darwinistischer Sicht vermutlich einen Selektionswert. Ein Organismus, der zweimal eine Entscheidung fällen kann, kann bei der ersten Entscheidung keinen allzu großen Fehler gemacht haben. Die Arbeit ebnete hoffentlich den Weg, zu untersuchen, inwieweit ein hoher Informationsgehalt auch bei Kategorienbildung ohne Lehrer vorhanden ist und von Menschen genutzt wird.

3 Zusammenfassung

Das spezielle Design der oben besprochenen Experimente erlaubte den Blocking-Effekt auch in Paradigmen ohne Rückmeldung nachzuweisen. Der Blocking-Effekt konnte sowohl in einem Paradigma mit Zeitdruck durch die Kategorienzuordnung eines Teststimulus als auch in einem Paradigma ohne Zeitdruck durch Einschätzung der Bekanntheit einzelner Icons nachgewiesen werden. Die Studie kräftigte die Hypothese, dass sowohl Kategorisierungsmodelle mit Rückmeldung als auch Klassifizierungsmodelle ohne Rückmeldung Lernen in Abhängigkeit von der bereits erreichten Konsistenz modellieren sollten. Im ersten Fall besteht diese Konsistenz zwischen der Handlung und ihrer Rückmeldung, im zweiten Fall in der Konsistenz des Verhaltens auf ähnliche Ereignisse, ausgedrückt durch den Versuch die Verbindungsstärken in einem Delta-Modell zu maximieren. Die hier geäußerten Modellvorstellungen lassen sich leicht mit den Modellvorstellungen von Pearce und Hall (1980) oder auch Kruschke und Johansen (1999) vereinbaren, schwieriger jedoch mit dem Modell von Mackintosh (1975) oder regelbasierten Modellen (z.B. Anderson & Betz, 2001).

Eingereicht

Kruschke, J. K. Cue competition in function learning: Blocking and highlighting. Geladen von "www.indiana.edu/~kruschke/articles/ CueCompFuncLearn.pdf", Stand 02.2003.

Love, B. C. Comparing supervised and unsupervised category learning. Geladen von „<http://love.psy.utexas.edu/~love/concepts/Love2003.pdf>“, Stand 02.2003.

Love, B. C, Medin, D. L. & Gureckis, T. M. Sustain: A network model of category learning. Geladen von "http://love.psy.utexas.edu/~love/concepts/Love_et al2003.pdf", Stand 02.2003.

Literatur

Ahn, W-K. & Medin, D. L. (1992). A two-stage model of category construction, *Cognitive Science*, 16, 81-121.

Allen, S. W. & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General*, 120, 3-19.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, 98, 409-429.

Anderson, J. R. & Lebiere, C. (1998). *The atomic components of thought*. Mahwah: Erlbaum.

Anderson, J. R., & Betz, J. (2001). A hybrid model of categorization. *Psychonomic Bulletin and Review*, 8, 629-647. Geladen von http://act-r.psy.cmu.edu/papers/104/Anderson_Betz.pdf, Stand 02.2003.

Bortz, J. (1979). *Lehrbuch der Statistik für Sozialwissenschaftler*. Berlin: Springer-Verlag.

Berry, D. C. & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *The Quarterly Journal of Experimental Psychology*, 36A, 209-231.

Bösche, W. (2002). *Kodierung und Kapazität in adaptiven Netzwerk-Modellen des Klassifikationslernens*. Aachen: Shaker Verlag.

Chapman, G. B. , & Robbins, S. J. (1990). Cue interaction in human contingency judgement. *Memory & Cognition*, 18, 537-545.

Chater, N. (1999). The search for simplicity: A fundamental cognitive principle? *The Quarterly Journal of Experimental Psychology*, 52A, 273-302.

Chomsky, N. (1959). A review of B. F. Skinner's verbal behavior. *Language*, 35, 26-58

Dickinson, A. Shanks, D. & Evenden, J. (1984). Judgement of act-outcome contingency: The role of selective attribution. *Quarterly Journal of Experimental Psychology*, 36A, 29-50.

Elman, J.L. (1990). Finding structure in time. *Cognitive Science*, 14, 179-211.

- Fodor, J. & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition* 28, 3-71.
- Forster, K. I., & Forster, J. C. (2002). DMDX version 2.9. Geladen im September 2002 von <http://www.u.arizona.edu/~jforster/dmdx.htm>.
- Gigerenzer, G. & Todd, P. M. (1999). Fast and frugal heuristics. In G. Gigerenzer, P. M. Todd, and the ABC Research Group (Eds.). *Simple Heuristics That Make Us Smart*. Oxford: University Press.
- Gluck, M. A. & Bower, G. H. (1988a). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, 117, 227-247.
- Gluck, M. A. & Bower, G. H. (1988b). Evaluating an adaptive network model of human learning. *Journal of Memory and Language*, 27, 166-195.
- Gluck, M. A. (1991). Stimulus generalization and representation in adaptive network models of category learning. *Psychological Science*, 2, 50-55.
- Gluck, M. A. (1992). Stimulus sampling and distributed representations in adaptive network theories of learning. In A. F. Healy, S. M. Kosslyn & R.M. Shiffrin (Eds.). *From Learning Theory to Connectionist Theory*, pp. 169-199. Hillsdale, NJ: Erlbaum.
- Gluck, M. A., Allen, M. T., Myers, C. E. & Thompson, R. F. (2001). Cerebellar substrates for error correction in motor conditioning. *Neurobiology of Learning and Memory*, 76, 314-341.
- Goldstone, R. L. & Barsalou, L.W. (1998). Reuniting perception and conception. *Cognition*, 65, 231-262.
- Goldstone, R. L., Steyvers, M., Spencer-Smith, J. & Kersten, A. (2000). Interactions between perceptual and conceptual learning. In E. Diettrich & A. B. Markman (Eds.). *Cognitive Dynamics: Conceptual Change in Humans and Machines*, pp. 191-228. Lawrence Erlbaum and Associates.
- Hays, W. (1994). *Statistics*. Fifth Edition, Forth Worth: Harcourt Brace College Publishers.
- Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. New York: Wiley.
- Hecht-Nielsen, R. (1987). Kolmogorov's mapping neural network existence theorem. *ICNN'87*, 11-14, IEEE San Diego, CA: CS Press.
- Hertz, J. A., Palmer, R. G. & Krogh, A. S. (1991). *Introduction to the theory of neural computation*. Redwood City, CA: Addison-Wesley.
- Homa, D. & Cultice, J. (1984). Role of feedback, category size, and stimulus distortion on the acquisition and utilization of ill-defined categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 83-94.

- Homa, D. , Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory*, 7, 418-439.
- Hull, C. (1943). *Principles of behavior*. New York: Appleton-Century-Crofts.
- Jones F. W., Wills A. J., McLaren I. P. (1998). Perceptual categorization: Connectionist modelling and decision rules. *Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology*. 51(1). 33-58.
- Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. Campbell & R. Church (Eds.). *Punishment and Aversive Behavior*. New York:Appleton.
- Kant, I. (1995). *Kritik der reinen Vernunft*. Werkausgabe Band III. 13. Auflage Frankfurt am Main: Suhrkamp.
- Kaplan, A. S. & Murphy, G. L.(1999). The acquisition of category structure in unsupervised learning. *Memory and Cognition*, 27, 699-712.
- Knowlton, B. J. & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, 262, 1747-1749.
- Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 25, 1083-1119.
- Kruschke, J. K. & Blair, N. J. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin & Review*, 7, 636-645.
- Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, 45, 812-863.
- Lachnit, H., Kimmel, H., Bevill, M., Martin, I., Levey, A. & Hamm, A. (1990). Classical conditioning with human subjects. In P. J. D. Drenth, Pieter & J. A. Sergeant (Eds.), *European Perspectives in Psychology, Vol. 1: Theoretical, Psychometrics, Personality, Developmental, Educational, Cognitive, Gerontological*. pp. 353-368. Essex: Tiptree.
- Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, 82, 276-298.
- Mackintosh, N. J. (1978). Cognitive or associative theories of conditioning: Implications of an analysis of blocking. In S. H. Hulse, H. Fowler, & W. K. Honig (Eds.), *Cognitive Processes in Animal Behavior* (pp. 177-209). Hillsdale, NJ:Erlbaum.
- Marr, D. (1982). Vision: The philosophy and the approach. In A. M. Aitkenhead & J. M. Slack (Eds.), *Issues of Cognitive Modelling: A Reader*, pp. 103-126. London: Erlbaum.
- McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and a specific information. *Journal Of Experimental Psychology: General*, 114, 159-188.

- McLaren, I. P. L. & Mackintosh, N. J. (2000). An elemental model of associative learning: I. Latent inhibition and perceptual learning. *Animal Learning and Behavior*, 28, 211-246.
- Medin, D. L. , Wattenmaker, W.D. & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology*, 19, 242-279.
- Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.
- Miller, R., Barnet, R. & Grahame, N. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin*, 117, 363-386.
- Minsky, M. L. & Papert, S. (1969). *Perceptrons: An Introduction to computational geometry*. Cambridge, Massachusetts: MIT-Press.
- Nisbett, R. E. & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259.
- Nosofsky, R. M. (1986). Attention, similarity and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115, 39-57.
- Oja, E. (1982). A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology* 15, 267-273.
- Pearce, J. M. & Hall, G. (1980). A model for pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, 87, 532-552.
- Pearce, J. M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review*, 101, 587-607.
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., and Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi- regular domains. *Psychological Review*, 103, 56-115.
- Popper, Karl (1935). *Logik der Forschung*. Wien: Springer.
- Pothos, E. M. & Chater, N. (2001). Category learning without labels – A simplicity approach. In *Proceedings of the twenty-third annual conference of the cognitive science society*, 774-779.
- Preusser, D. & Handel, S. (1970). The free classification of hierarchically and categorically related stimuli. *Journal of Verbal Learning and Verbal Behavior*, 9, 222-231.
- Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, 3, 382-407.
- Reed, J. M., Squire, L. R., Patalano, A. L., Smith, E. E. & Jonides, J. (1999). Learning about categories that are defined by object-like stimuli despite impaired declarative memory. *Behavioral Neuroscience*, 113, 411-419.

- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical Conditioning II: Current Research* (pp. 64 - 99). New York: Appleton-Century-Crofts.
- Regehr, G. & Brooks, L. R. (1995). Category organization in free classification: The organizing effect of an array of stimuli. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21, 347-363.
- Rieskamp, J. & Hoffrage, U. (1999). When do people use simple heuristics, and how can we tell? In G. Gigerenzer, P. M. Todd, and the ABC Research Group (Eds.). *Simple Heuristics That Make Us Smart*. Oxford: University Press.
- Rodrigo, T., Chamizo, V. D., McLaren, I. P. L. & Mackintosh, N. J. (1997). Blocking in the spatial domain. *Journal of Experimental Psychology, Animal Behavior Processes*, 23, 110-118.
- Roediger, H. L. & McDermott, K. B. (2000). Distortions of memory. In E. Tulving (Ed.). *The Oxford Handbook of Memory*. Oxford: Oxford University Press.
- Rolls E. T. and Treves A. (1998). *Neural networks and brain function*. Oxford: University Press.
- Rosenblatt, F. (1962). *Principles of neurodynamics*. Washington: Spartan Books.
- Rumelhart, D. E., J. L. McClelland (1986). On learning the past tenses of english verbs. In D. E. Rumelhart, J. L. McClelland and the PDP Research Group (Eds.) *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models*, pp. 216-271. Cambridge, London: MIT Press.
- Rumelhart, D. E., Hinton, G. E. & J. L. McClelland (1986). A general framework for parallel distributed processing. In D. E. Rumelhart, J. L. McClelland and the PDP Research Group (Eds.) *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Volume 1: Foundations*, pp. 45-76. Cambridge, London: MIT Press.
- Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland and the PDP Research Group (Eds.) *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Volume 1: Foundations*, pp. 318-362. Cambridge, London: MIT Press.
- Rumelhart, D. E. & Zipser, D. (1986). Feature discovery by competitive learning. In D. E. Rumelhart, J. L. McClelland and the PDP Research Group (Eds.) *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Volume 1: Foundations*, pp. 151-193. Cambridge, London: MIT Press.

- Scheffers, M. K. & M. G. H. Coles (2000). Performance monitoring in a confusing world: Error-related brain activity, judgments of response accuracy, and types of errors. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 141-151.
- Seligman, M. E. (1970). On the generality of the laws of learning. *Psychological Review*, 77, 406-418.
- Shanks, D. R. (1995). *The psychology of associative learning*. Cambridge: University Press.
- Shanks, D. R. & Darby, R. J. (1998). Feature- and rule-based generalization in human associative learning. *Journal of Experimental Psychology: Animal Behavior Processes*, 24, 405-415.
- Shiffrin, R. M. & Schneider, W. (1977). Controlled and automatic human information processing, II: Perceptual learning, automatic attending and a general theory. *Psychological Review*, 84, 127-190.
- Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York, London: McGraw-Hill.
- Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.
- Smith, J. D. & Kemler Nelson, D. G. (1984). Overall similarity in adults' classification: The child in all of us. *Journal of Experimental Psychology: General*, 113, 137-159.
- Stone, G. O. (1986). An analysis of the delta rule and the learning of statistical associations. In D. E. Rumelhart, J. L. McClelland and the PDP Research Group (Eds.) *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Volume 1: Foundations*, pp. 444-459. Cambridge, London: MIT Press.
- Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, 88, 135-170.
- Sutton, G. G. & Reggia, J. A. (1994). Effects of normalization constraints on competitive learning. *IEEE Transactions on Neural Networks*, 5, 502-504.
- Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology*, 40A, 201-237.
- Turvey, M. T. (1973). On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review*, 80, 1-52.
- Walkenbach, J., Haddad, N. F. (1980). The Rescorla-Wagner theory of conditioning: A review of the literature. *Psychological Record*, 30, 497-509.

- Wagner, A. R. & Rescorla, R. A. (1972). Inhibition in Pavlovian conditioning: Application of a theory. In R. A. Boakes & M. S. Halliday (Eds.), *Inhibition and Learning* (pp. 301-336). New York: Academic Press.
- Wagner, A. R. (1978). Expectancies and the priming of STM. In S. H. Hulse, H. Fowler, & W. K. Honig (Eds.), *Cognitive Processes in Animal Behavior*. Hillsdale, NJ: Erlbaum.
- Ward, T. B. (1983). Response tempo and separable-integral responding: Evidence for an integral-to-separable processing sequence in visual perception. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 103-112.
- Williams, D. A., Sagness, K. E. & McPhee, J. E. (1994). Configural and elemental strategies in predictive learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20, 694-709.
- Wills, A. J. (2002). Adapting to a response deadline in categorization. *Proceedings of the 24th Annual Conference of the Cognitive Science Society*. 938-943.
- Wills, A. J. & McLaren, I. P. L. (1997). Generalization in human category learning: A connectionist account of differences in gradient after discriminative and non-discriminative training. *The Quarterly Journal of Experimental Psychology*, 50A, 607-630.
- Wills A.J., Reimers S., Stewart N., Suret M., McLaren I. P. (2000). Tests of the ratio rule in categorization. *Quarterly Journal of Experimental Psychology*, 53A, 983- 1011.
- Zaki, S. R. & Nosofsky, R. M. (2001). A single-system interpretation of dissociations between recognition and categorization in a task involving object-like stimuli. *Cognitive, Affective, and Behavioral Neuroscience*, 1, 344-359.
- Zwicker, J. & Wills, A. J. (2002). Is competitive learning an adequate account of free classification? *Proceedings of the 24th Annual Conference of the Cognitive Science Society*. 982-987.

Anhang 1:

Instruktionstext I

In diesem Experiment geht es darum, wie du verschiedene Kategorien bildest, ohne dass du dabei Rückmeldung bekommst. Dies ist für viele Teilnehmer der Studie sehr verwirrend. Lass dich aber davon nicht entmutigen. Alles was von dir verlangt wird ist, dass du die Bilder die dir dargeboten werden versuchst in 3 verschiedene Gruppen einzuteilen. Die Einteilung soll dabei aufgrund von Ähnlichkeiten erfolgen. Am Anfang musst du natürlich raten welches Bild du in welche Kategorie stecken sollst. Nachdem du aber mehr und mehr zugeordnet hast, wirst du einen klareren Eindruck davon erhalten welche Bilder in eine Kategorie gehören. Um anzuzeigen, in welche Kategorie ein Bild eingeordnet werden soll, stehen dir 3 Tasten zur Verfügung: „1“, „2“ und „3“ auf dem Pfeilblock. Du wirst im Laufe der Zeit eventuell unterschiedliche Eindrücke der 3 Kategorien entwickeln. Wichtig ist dabei nur, dass du nicht willkürlich die Kategorien wechselst, also nach einiger Zeit alle Bilder die du Kategorie 1 zugeordnet hast, der Kategorie 2 zuordnest oder Ähnliches. Natürlich kann es aber passieren, dass ein Bild, welches du am Anfang noch Kategorie 2 zugeordnet hättest, jetzt in die Kategorie 1 zu passen scheint.

Mit der Leertaste geht es weiter...

Instruktionstext II

Bitte versuche dich über die nächsten 20 Minuten gleichmäßig zu konzentrieren und lasse dich nicht davon entmutigen, dass du überhaupt keine Rückmeldung erhältst. Schon mal jetzt vielen Dank!

Mit der Leertaste geht es weiter...

Anhang 2:

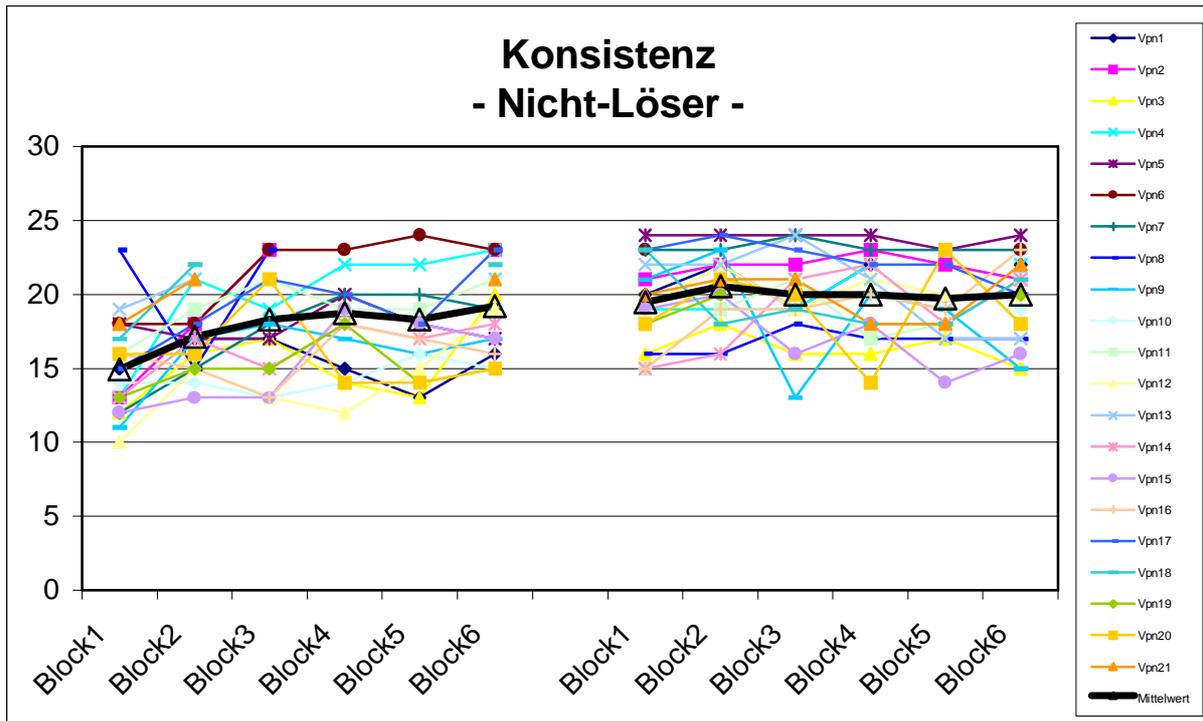
Verwendete Icons.



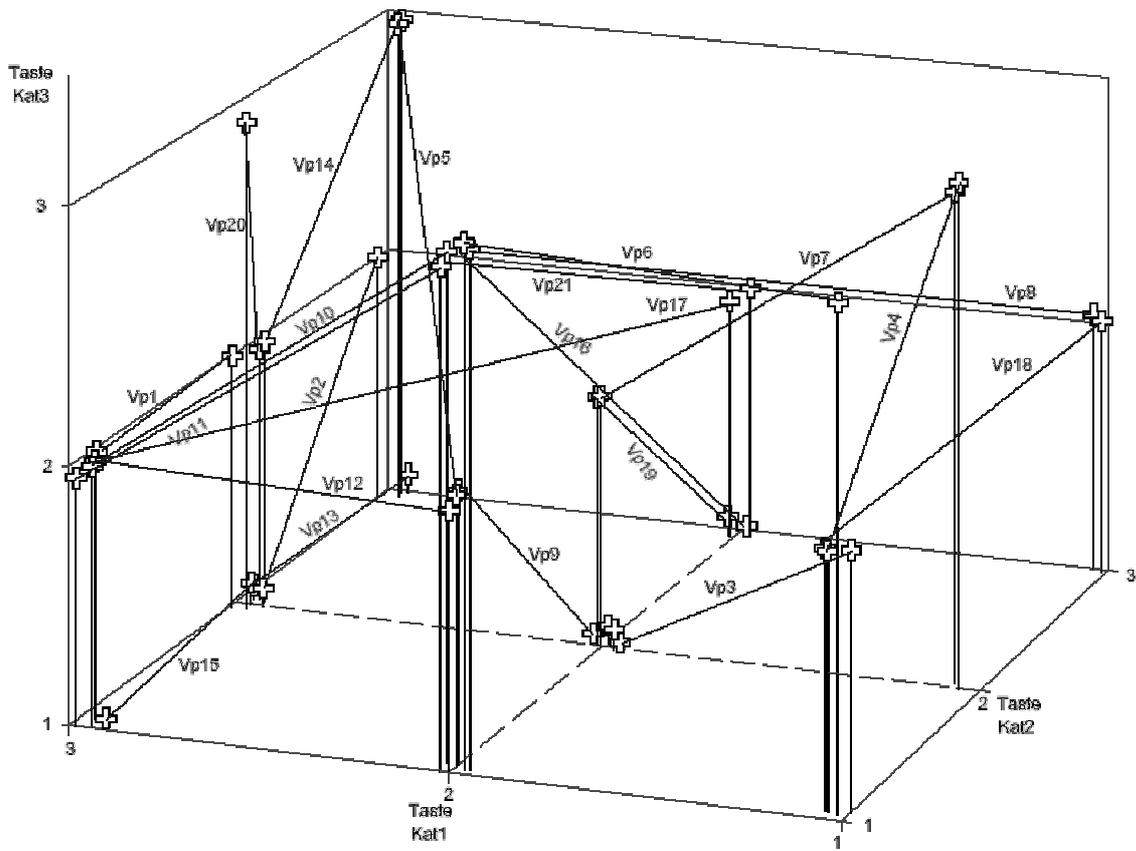
Anhang 3:

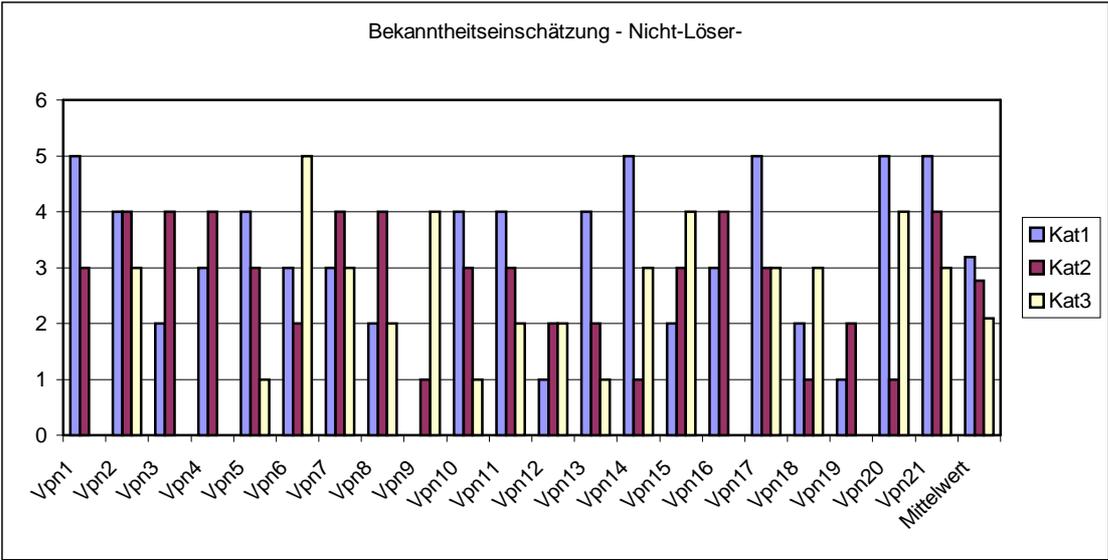
Herleitung der Wahrscheinlichkeit, durch Zufall das Kriterium zu erreichen

Die analytische Lösung für die Wahrscheinlichkeit, dass eine Versuchsperson das Kriterium in einem Block durch zufällige Auswahl der Tasten erreicht, lässt sich nur dann relativ leicht finden, wenn jede Taste unabhängig von der anderen gewählt werden kann. Dies würde aber auch die extrem unrealistischen Fälle mit einschließen, in denen nur eine Taste für alle Kategorien benutzt wurde. Um eine realistischere Abschätzung zu erhalten wurde die Randbedingung hinzugefügt, dass alle 3 Tasten über den Block hinweg gleich häufig gedrückt werden sollen. In diesem Fall sind die Wahrscheinlichkeiten eine bestimmte Taste zu drücken nicht mehr unabhängig voneinander, was eine analytische Lösung erschwert. Aus diesem Grunde wurde eine Computersimulation eingesetzt um diese Wahrscheinlichkeit zu schätzen. In der Computersimulation wurde eine „Urne“ mit jeweils 8 Repräsentanten der Kategorien 1, 2 und 3 gefüllt. Danach wurde zufällig jeweils ein Repräsentant ausgewählt und zufällig einem von 3 „Tasten“-Eimern zugewiesen. Der Repräsentant wurde danach aus der Urne entfernt. Die Eimer konnten jeweils höchstens 8 Tasten beinhalten. Wenn ein Eimer bereits gefüllt war wurde die „Kategorie“ zufällig auf die anderen noch freien Eimer verteilt. Nachdem die Urne leer war wurde überprüft, ob das Kriterium, das darin bestand, dass in allen 3 Eimern mehr als 5 Repräsentanten der gleichen Kategorie vorhanden sein sollen, erfüllt war. In 10^8 Durchgängen wurden auf diese Weise 19432 Fälle erzielt, bei denen das Kriterium erfüllt war. Bei 15 Kontrollsimulationen mit anderen Ausgangswerten für den Zufallszahlengenerator wurden Werte im Bereich zwischen 19372 und 19776 erhalten. Damit lässt sich mit großer Sicherheit die Wahrscheinlichkeit im Experiment in einem Block das Kriterium zufällig zu erreichen durch 0.0002 nach oben abschätzen. Da jede Versuchsperson, höchstens 9 Blöcke in jeder Phase zu absolvieren hatte, war die Wahrscheinlichkeit die nächste Phase bei zufälliger Betätigung der Tasten *nicht* zu erreichen $(1 - 0.0002)^9 = 0.998$. Die Wahrscheinlichkeit in *einer* der Phasen zufällig das Kriterium zu erreichen betrug damit $1 - 0.998^2 = 0.004$.



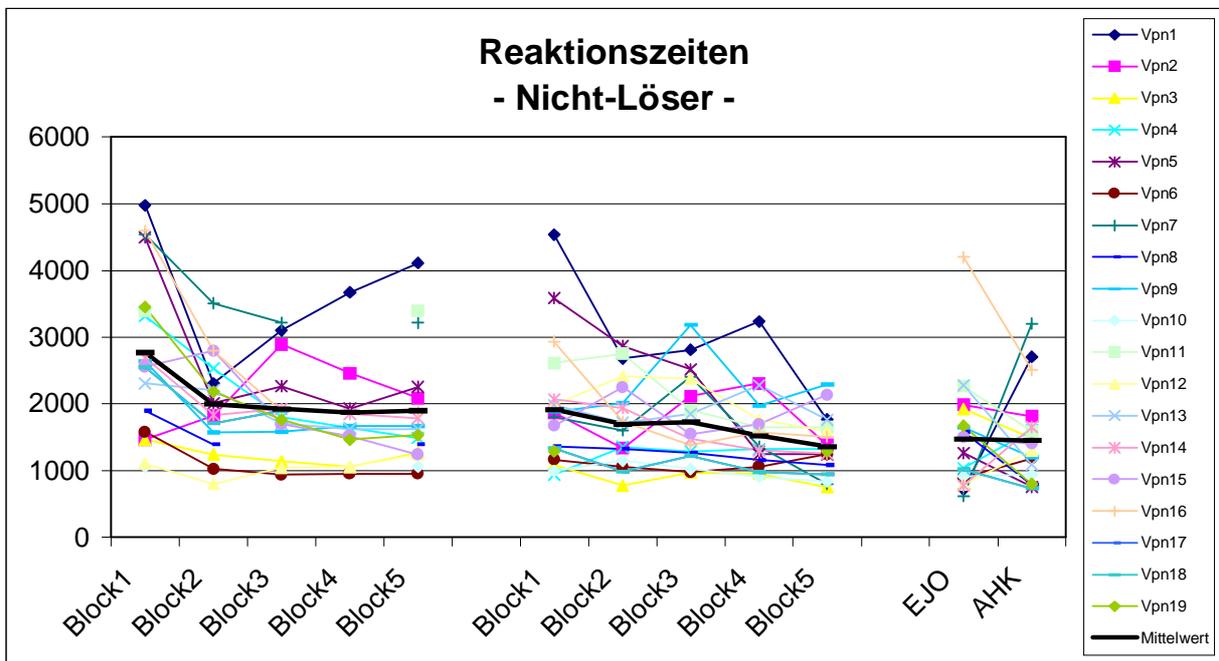
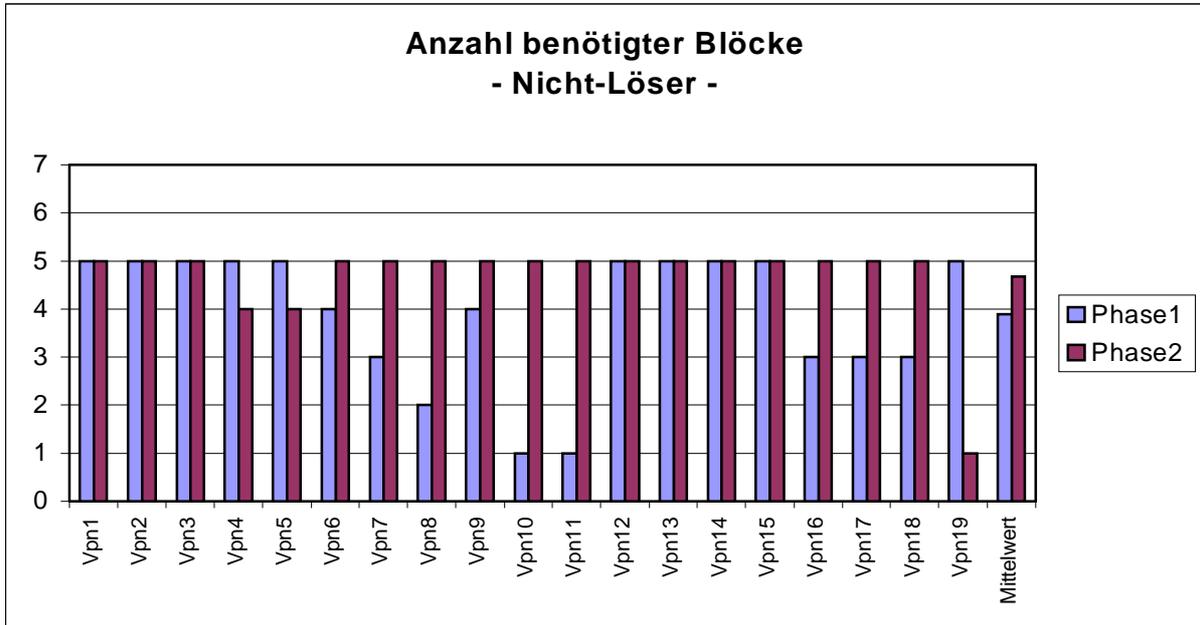
Konsistenz Tasten zu Kategoriezuordnung.

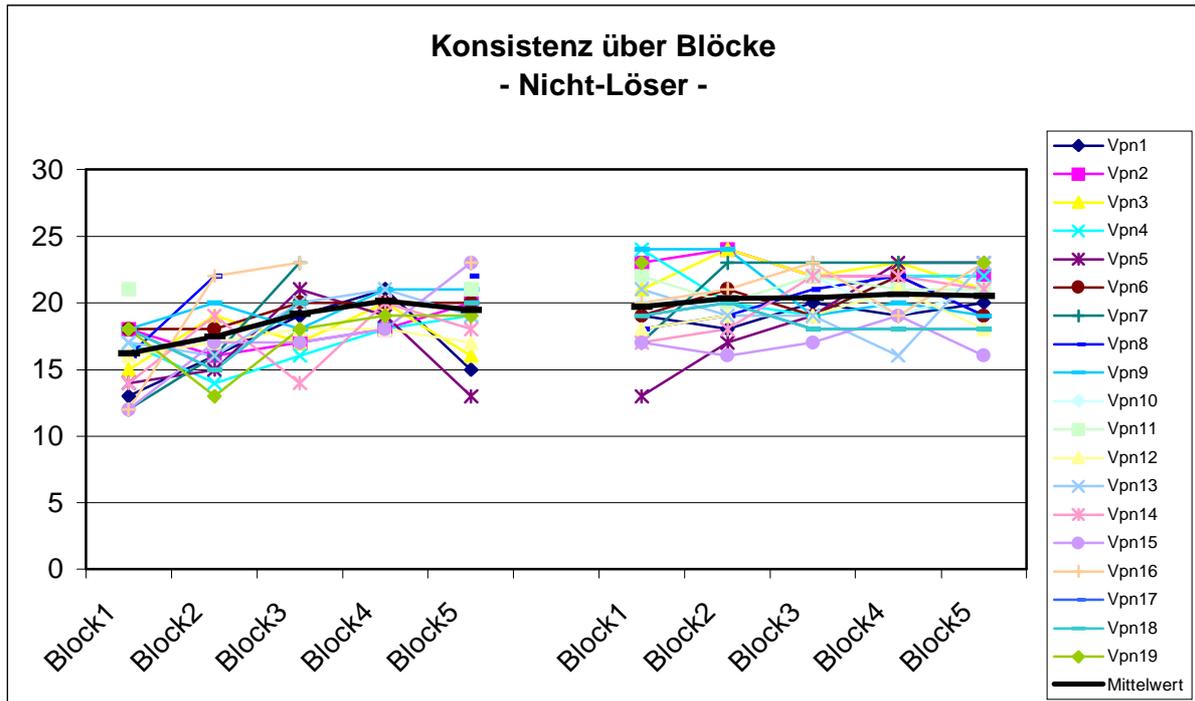




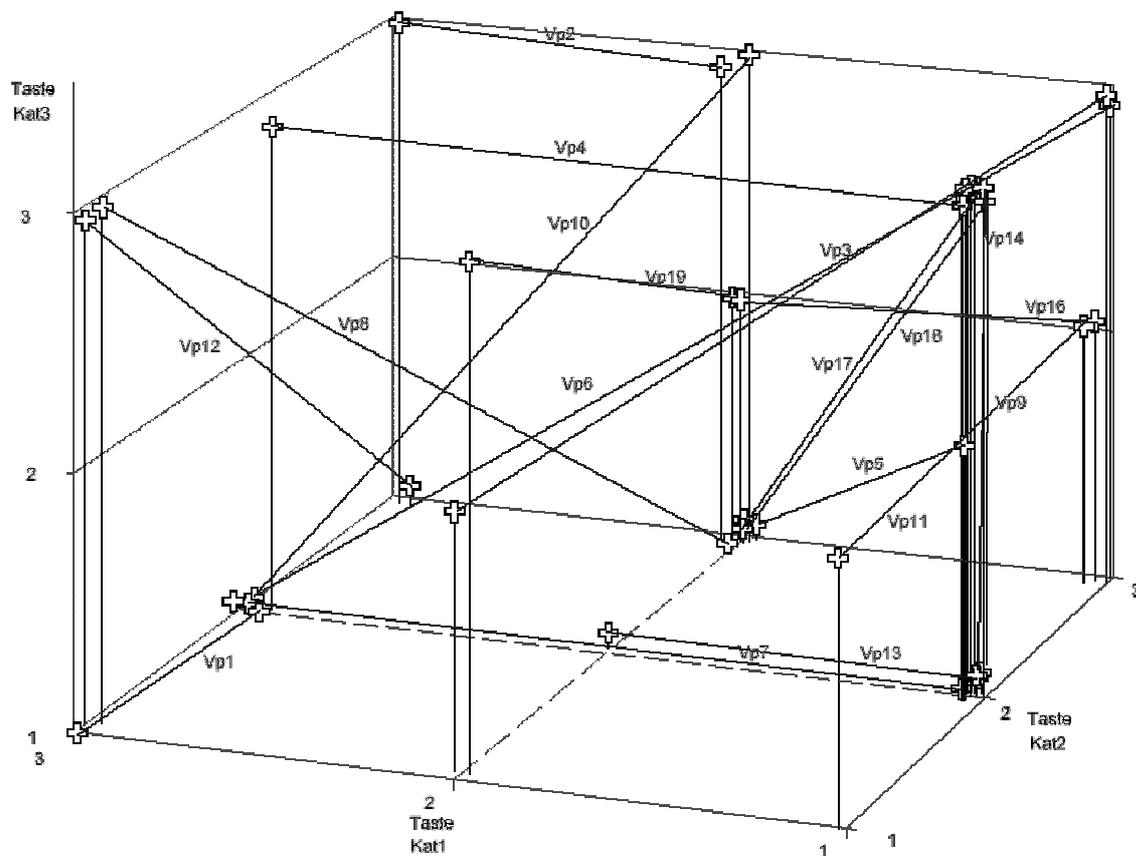
Anhang 5:

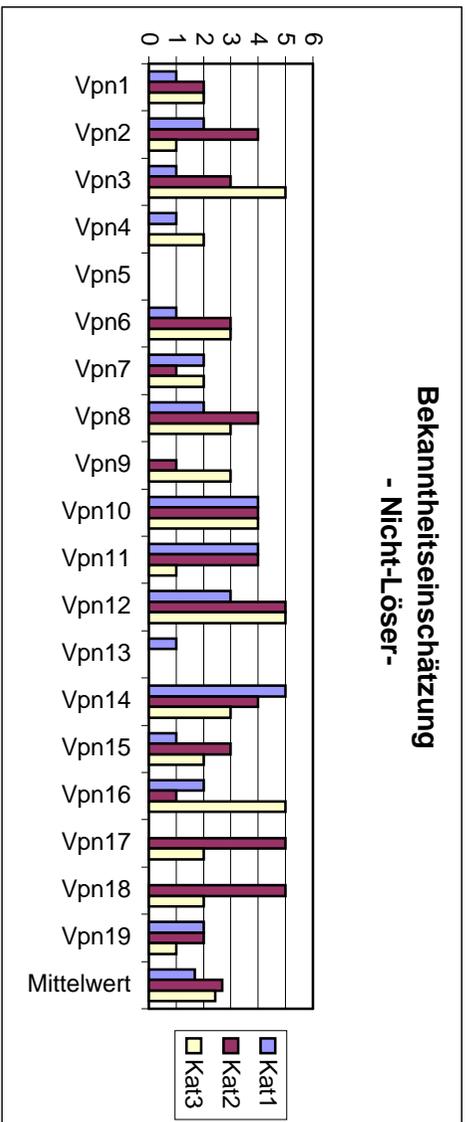
Experiment 2, Ergebnisse für Nicht-Löser

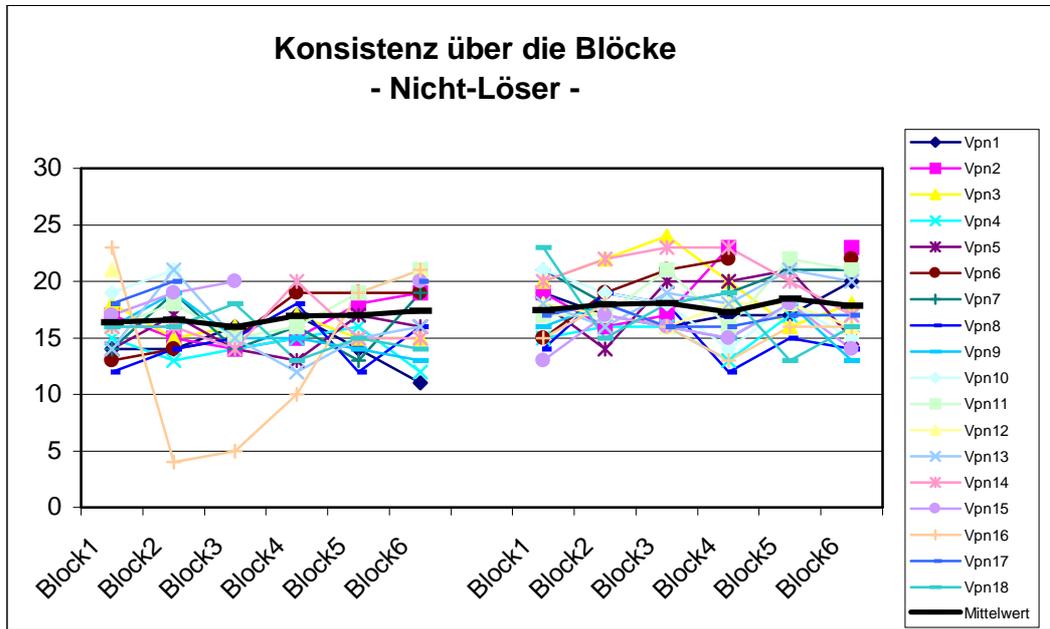




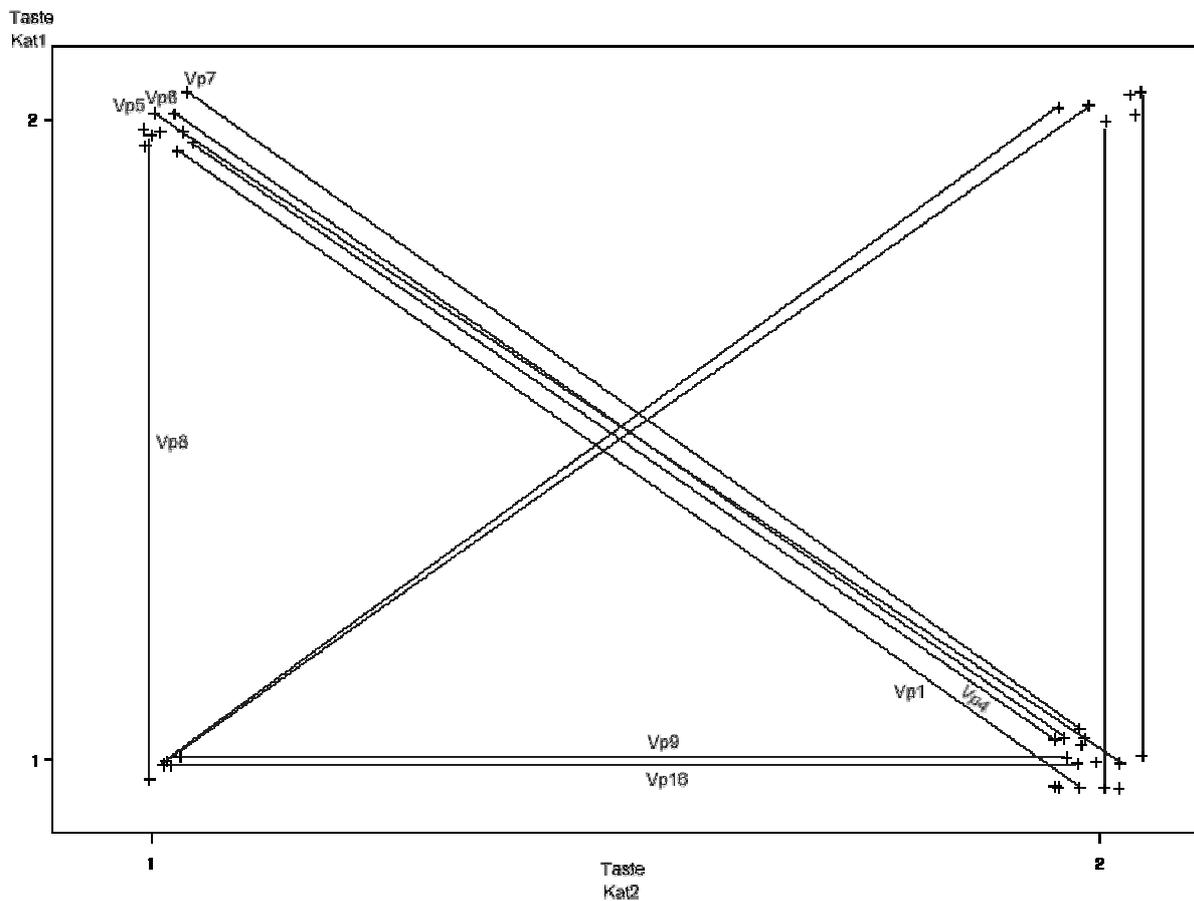
Konsistenz der Tasten zu Kategorie Zuordnungen

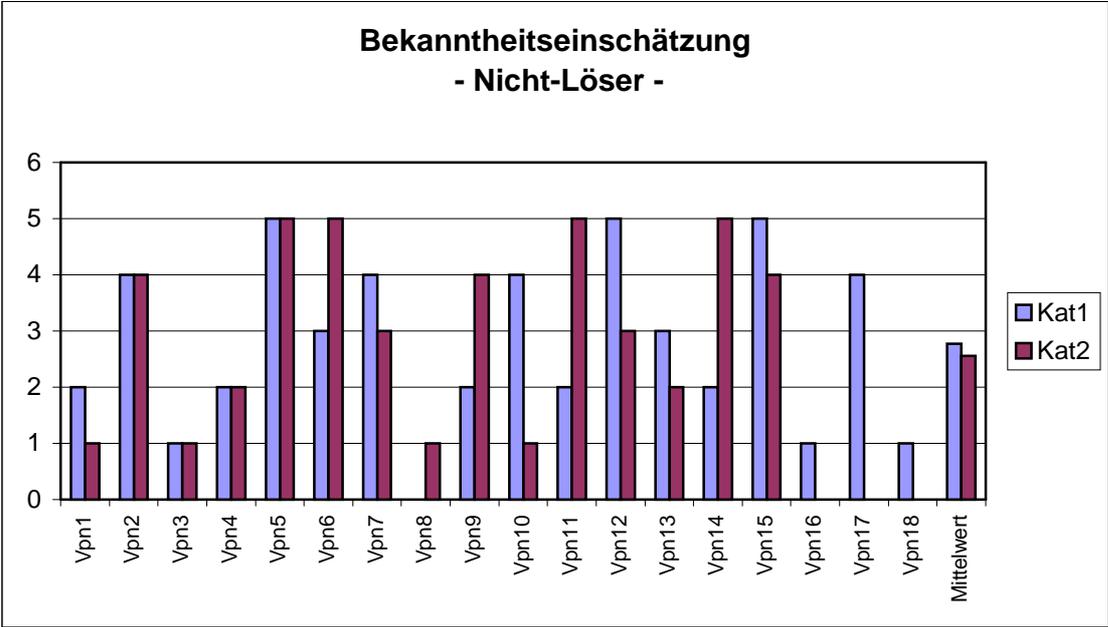






Tasten zu Kategorienzuordnung





Anhang 7:

Instruktion zur Analyse der Beschreibungen über die Kategorienzuordnung

Entscheide zuerst, ob die Versuchsperson eher auf die Form und Ausprägung eines Icons geachtet hat oder auf die Farbe. Entscheide dich für genau eine dieser Möglichkeiten. Auch wenn du der Meinung bist, dass beide Entscheidungsgrundlagen in den Text interpretiert werden können, triff eine Entscheidung danach, was dir die vorherrschende Tendenz zu sein scheint. Bei Aussagen wie z.B. „allgemeine Ähnlichkeit“ überlege dir, ob eine Entscheidung, die auf allgemeiner Ähnlichkeit basiert, eher durch Farben oder durch Formen realisierbar ist. Verfahre als nächstes genauso mit der Unterscheidung, ob für die Kategorienwahl jeweils mehr als 1 Icon eine Rolle spielt. Das heißt, ob die Versuchsperson schildert, dass sie auf mehr als auf 3 Icons, für jede Kategorie eines, geachtet hat. Auch hier musst du wieder eine Entscheidung treffen. Schließlich gib an, ob der Text nahe legt, dass eine Restkategorie genutzt wurde, in die jeweils alle Stimuli platziert wurden, welche in keine andere Kategorie zu passen schienen. Schließlich, entscheide, ob die Versuchsperson ihre Strategie gewechselt hat. Der Wechsel von einem Icon zu einem anderen zählt dabei nicht als Strategiewechsel, der Wechsel von Farb- zu Formbeurteilung jedoch schon.

Anhang 8

Erstrater

Zweitrat

Versuchspersonenberichte Experiment 1. Unterschiede zwischen den Ratern sind farblich hervorgehoben

Erfolgreiche Löser

		Farben	Iconform	einzelnes Icon	mehrere Icons	Restkategorie	Wechsel	Farben	Iconform	einzelnes Icon	mehrere Icons	Restkategorie	Wechsel
Vpn1	sortieren nach Farben welche wo am Häufigsten vorkommt Eindruck	x			x			x			x		
Vpn2	Ich hab mir die Muster angesehen und hab dann entschieden sie nach den überwiegenden Farben zu gliedern. In Gruppe 2 waren aber alle Muster die nirgendwohin passten. In Gruppe 1 waren die rötlichen und in Gruppe 3 die grünen. Später wurde Gruppe 2 die bläulichen und die Gruppe 1 die die nirgendwohin passen.	x			x	x		x			x	x	
Vpn3	keine Angaben												
Vpn4	Bei dem Übungsdurchgang war ich noch völlig ratlos, in welche Kategorien ich einteilen sollte. Nach einiger Zeit meinte ich dann zu bemerken, daß die Zusammenstellung häufig die gleiche war, merkte mir zwei der Symbole und ihren Standort (insbesondere unten rechts und in der Mitte links) und versuchte dann, die gleichen Muster in die gleiche Kategorie zu stecken, vertat mich aber häufiger. Aufgefallen ist mir auch, daß ca. nach der Hälfte der Zeit drei andere Muster immer wieder auftauchten, während die von vorher gar nicht mehr auftraten. Ich teilte dann drei neue Kategorien ein und ging nach dem gleichen System vor, vergaß aber auch hier häufiger die Einteilung oder vertippte mich versehentlich.		x		x				x		x		
Vpn5	1 war eher rot 2 eher blau 3 eher grün lastig	x			x			x			x		
Vpn6	zum Teil habe ich mich an einzelnen Symbolen orientiert, die meistens an der gleichen Stelle des Kastens wieder aufgetaucht sind. Zum Teil bin ich auch nach dem Gesamteindruck vorgegangen, z.B. wenn die Farbe rot/grün überwog oder die Figuren stark kantig wirkten.		x	x					x			x	
Vpn7	jeweils ein Icon		x	x					x	x			
Vpn8	Farben olivgrüne Viererkette Formen die plakativ waren, "rote Trompete" "Schnörkel D", "Blitz" noch mehr Farben.	x			x			x			x		
Vpn9	Habe mich an drei markanten, sich wiederholenden Merkmalen orientiert und sie dann zugeordnet. Dabei gab es Formen, die mir ins Auge fielen und Farb-Kombinationen, die sich von vorangegangenen Bildern abhoben (z.B. türkisfarbenes Zeichen mit roter Umrandung).		x	x					x	x			
Vpn10	Kategorien gebildet nach den beiden ersten Zeichen. Vor allem farbliche Zusammenspiele. 3 war immer die gleiche Farbe bei beiden ersten Zeichen. 1 war immer gelb und rot oder auch andere grelle Farbunterschiede. 2 war dann der Rest bzw. meist blau und eine andere Farbe.	x			x			x			x	x	
Vpn11	Ich habe die Kategorisierung nach Form des Items bestimmt. Die erste Kategorie waren zackige Item, wobei ich mich auf das erste Item beschränkt habe. In die zweite Kategorie habe ich rundere Formen eingeordnet, während ich in die dritte Kategorie ein kreuzförmiges Item eingeordnet habe, und alle Items, die mir unbekannt schienen.		x	x		x			x		x	x	
Vpn12	Ich habe mir immer das erste Zeichen in der oberen linken Ecke angeschaut und festgestellt, dass sie sich wiederholen, das habe ich als Kriterium für die Kategorie 1,2 oder 3 genommen.	x		x					x	x			
Vpn13	zuerst nur nach Menge der Farben, 1-blau,2- gelb, 3-rot, dann nach Häufigkeit der Farben (entsprechend oben), wurde zunehmend schwierig, ging aber bis zum Ende. Beim 2. Teil hatte ich echte Probleme, da ich die Icons nicht so sehr beachtet habe.	x			x			x			x		x
Vpn14	einzelnes Icon		x	x					x	x			

Versuchspersonenberichte Experiment 2.
Unterschiede zwischen den Ratern sind farblich
hervorgehoben

Erstrater

Zweitratr

Erfolgreiche Löser

		Farben	Iconform	einzelnes Icon	mehrere Icons	Restkategorie	Wechsel	Farben	Iconform	einzelnes Icon	mehrere Icons	Restkategorie	Wechsel
Vpn1	Kathogorien habe ich zunächst nicht wirklich gebildet, da mir nicht ersichtlich war, wonach ich die Symbole ordnen könnte. Ich habe mir dann einzelne Symbole herausgesucht, welche Ähnlichkeit zu einem mir mehr oder weniger bekanntem Gegenstand hatten. Nach diesen habe ich dann kathegorisiert und bei öfterem Betrachten auch nach den anderen Symbolen.		x	x					x				
Vpn2	unter 1 kamen die Bilder, in denen mehrere fischähnliche Formen drin waren unter 2 kamen die Bilder, in denen Schleifen, Achter und Ähnliches dominierten unter 3 der Rest, wobei meist rechts oben ein hellblaues Symbol war		x		x	x			x		x	x	
Vpn3	Am Anfang bin ich ziemlich stark nach meinem Gefühl gegangen oder auch ein bisschen nach Farben (z.B. in der unteren Reihe ist alles blau). Danach habe ich mir jedoch markante Formen gemerkt, die immer am gleichen Platz auftauchen (z.B Kategorie 2: Schleife rechts oben). Auch hatte ich ein bisschen in Erinnerung, wie die anderen Formen dazu aussehen, was mich bestätigt hat. Dies hat meines Erachtens auch ganz gut geklappt, bis die Figuren durchgemischt wurden. Ganz am Schluss, wenn in einem Bild Formen waren, die zuvor markante Formen aus verschiedenen Bildern waren, habe ich eher wieder nach Gefühl gehandelt.		x		x		x		x		x		x
Vpn4	Nach bestimmten markanten oder auffallenden Zeichen, dem spontanen Gesamteindruck und den überwiegend vorkommenden Farben!		x		x				x		x		
Vpn5	Überwiegend gleiche Symbole - gleiche Kategorie. Farbliche Trennung	x			x				x		x		
Vpn6	1 für eher dunkler Gesamteindruck 2 für eher heller Gesamteindruck 3 wenn auffällig viele eckige Figuren vorkamen (hab mich am Schluss dabei erwischt, 3 für hell/dunkel-Gemisch zu nehmen)	x			x			x			x		x
Vpn7	Kriterien zur Kategorisierung: - Icon im linken oberen Eck - blau/gelb oder rot/grün überwiegend -> Anzahl der Stimuli pro Farbkombi - Gesamteindruck der Farbkombi		x	x				x			x		
Vpn8	die formen der bilder (spitz,eckig,rund,...) haben mir geholfen die bilder auch nur irgendwie einzuteilen. manchmal hab ich nicht das ganze bild angeschaut, sonder nur einen teil (zwei gleiche zeichen nebeneinander), und so wiedererkannt. ich bin mir sicher,dass ich ab und zu das selbe bild unter verschiedene nummern verteilt habe.		x		x				x		x		
Vpn9	Farbe	x			x			x			x		
Vpn10	Auf Formähnlichkeiten beziehungsweise Familiarität geachtet. Einzelne Bestandteile zugeordnet.		x		x				x		x		
Vpn11	Anhand eines häufig vorkommenden icons die jeweilige Kathogorie bestimmt. Wenn jew. Icon erkannt, jew. Zahl gedrückt, Rest ignoriert; wenn keines der bekannten icons dabei war, meistens 2 gedrückt		x	x		x			x	x		x	

Vpn12	Kategorie 1: große, zackige Gebilde Kategorie 2: gemischt, mittel, hyrogiefenartige Gebilde Kategorie 3: kleine, runde, flächige Gebilde		x		x				x		x		
Vpn13	Kategorie 1: Symbol x vorhanden Kategorie 2: weder Symbol x noch y vorhanden Kategorie 3: Symbol y vorhanden		x	x			x		x	x			x
Vpn14	Die Art wie die Kategorien gebildet habe hat sich im Laufe des Experiments geändert. Zuerst habe ich die Formen in rund, eckig und gemischt unterteilt. Das war aber sehr schwierig. Deshalb habe ich danach immer an einer bestimmten Figur orientiert. Wenn ich die entdeckte, wusste ich, um welche Kategorie es sich handelte.		x	x			x		x	x			x
Vpn15	1: Klein & rund 2: Weder 1 noch 3; eckig und rund 3: eckige Formen		x		x	x			x		x	x	
Vpn16	Ich habe hauptsächlich auf die Farben geachtet, ob sie hell grell oder eher dunkel waren. Zudem habe ich mir die Anordnung der ersten Reihe (waagrecht u. senkrecht) versucht zu merken und anzuordnen.	x			x			x			x		
Vpn17	Kategorienauswahl nach Intensität einer Farbe, d.h. war ein Gesamtbild eher im Ton rot/gelb oder hellblau. Ausserdem nach bekannten Objektformen, die mir mit der Zeit immer bekannter wurden.	x			x			x			x		
Summe		5	12	5	12	4	2	5	12	3	14	5	2

Erstrater

Zweitratr

Farben	Iconform	einzelnes Icon	mehrere Icons	Restkategorie	Wechsel	Farben	Iconform	einzelnes Icon	mehrere Icons	Restkategorie	Wechsel
--------	----------	----------------	---------------	---------------	---------	--------	----------	----------------	---------------	---------------	---------

Nicht-erfolgreiche Löser

Vpn1	nach anfänglichen Schwierigkeiten : Vor allem nach Farbhäufigkeit	x			x				x			x	
Vpn2	Die Kategorien habe ich nach der Anzahl der Farbe blau gebildet. Nicht so sehr nach den Icons, sondern wie oft die Farbe blau im ganzen Gebilde vorkam. Wenn die Farbe bis 3 mal vorkam, war es Kategorie 1, bei 4-5mal die Kat. 2 und bei mehr als 6mal die Kat.3.	x			x				x				
Vpn3	Ich habe nach den Farben sortiert, ob es rot, blau oder braun dominiert.	x			x				x			x	
Vpn4	kategorie 1 : grüner Anteil hat überwiegt kategorie 2 : blauer Anteil hat überwiegt kategorie 3 : gemischte Farben	x			x				x			x	x
Vpn5	Kategorie 1: § drin Kategorie 2: & drin Kategorie 3: Nichts von beiden oder beide drin		x	x					x	x			x
Vpn6	farben	x			x				x			x	
Vpn7	kein wirkliches System immer den ersten in der zweiten Reihe den verschiedenen Kategorien zugeordnet. System nicht immer durchgehalten		x	x					x	x			
Vpn8	nach farben "kalt" (mehr blau) vs. "warm" (mehr rot) bzw "wärmer" (noch mehr rot +goldrand)	x			x				x			x	x
Vpn9	mehrer Icons		x		x				x			x	
Vpn10	Ich habe nach Farbe kategorisiert. Je nachdem welche farbe auf dem Bild überwog, habe ich das Bild in verschiedene Kategorien gesteckt. In der ersten Kategorie war es überwiegend die Farbe Rot, in der zweiten die Farbe Grün und in der dritten die Farbe Blau.	x			x				x			x	

Vpn11	In der ersten Kategorie alles, was nirgends richtig dazupasst, in der zweiten alle Bilder mit fischartigem roten Symbol und in der dritten alle Bilder mit auffälligem verschlungenen hellblauen Symbol. Manchmal war beides vorhanden, da habe ich gemerkt, dass meine Einteilung völlig falsch ist, aber da war es schon zu spät und mir fiel keine bessere Einteilung mehr ein...		x	x		x			x	x		x	
Vpn12	Meine Kategorien habe ich versucht nach dem äusseren Erscheinungsbild zu wählen, das heisst hauptsächlich nach der Farbenhäufigkeit. Wenn mehr blau vorkam war es die eine, bei mehr rot und braun Tönen die andere; als dritte Kategorie habe ich die weisse Fläche gewählt.	x			x			x			x		
Vpn13	erste K. :symbole mit mehr Blautönen zweite K. :symbole mit mehr rot-grüntönen dritte K. :wenn eine Farbe richtig eindeutig zu sehen war	x			x			x			x		
Vpn14	Ich bin nach dem Kriterium der Farbmehrheit vorgegangen, d.h. ich habe die Bilder nach der Farbe zugeordnet, die meinem ersten Eindruck nach am stärksten vertreten war. 1 grün/ gelb überwiegt 2 rot 3 blau	x			x			x			x		
Vpn15	Kategorien eingeteilt über Farben (blau vs. rot) bei vielen blauen Items links = 1 bei vielen blauen Items Mitte = 2 bei vielen blauen Items rechts bzw. nicht eindeutig bestimmbar = 3 manche Bilder bekannt in ihrer Kategorienugehörigkeit	x			x	x		x			x	x	
Vpn16	form und farbe		x		x				x		x		
Vpn17	kategorie 1: mengenanteil von rot und gelb überwiegt, zum teil symmetrische anordnung der symbole kategorie 2: mengenanteil von rot, gelb, grün und blau ausgewogen, keine symmetrie der icons, sondern durcheinander gemischt kategorie 3: mengenanteil von grün und blau überwiegt	x			x			x			x	x	
Vpn18	keine Angaben												
Vpn19	eigentlich habe ich fast ausschließlich auf die farbliche verteilung geachtet, zeitweise war z.b. viel rot gleich 1, blau und rot gleich 2 und viel blau und wenig rot gleich 3. die einzelnen zeichen oder bestimmte muster habe ich nicht beachtet.	x			x			x			x		
Summe		13	5	3	15	2	0	13	5	3	15	6	0

Vpn11	Überwiegend Kombinationen aus rot und gelb ist rechte Kategorie Alles andere: links im Laufe des Versuchs: bei mehr als drei rot-gelb-Kombinationen: rechts, in jedem anderen Falle links.	x			x	x		x			x	x	x
Vpn12	Habe die Bilder nach ihrem Anteil an der blauen Farbe eingeteilt. Die Bilder, die kühler erschienen, zusammen und umgekehrt.	x			x			x				x	
Vpn13	Anhand der Randdicke der einzelnen Figuren habe ich in Figuren mit dünnen und dicken Ränder unterteilt		x		x					x		x	
Vpn14	Rechts bestimmtes Icon vorhanden - links bestimmtes Icon nicht vorhanden: Icon wechselte wenn überhaupt nicht mehr aufkommend!!!		x	x						x	x		
Vpn15	Aussehen und Vergleich wie sie angeordnet sind		x		x					x		x	
Vpn16	Erst nach Farben - hat nicht funktioniert dann: ein markantes Symbol ausgesucht wo es vorkam= erste Gruppe, wo nicht= 2. Gruppe Merkmal musste einmal gewechselt werden		x	x		x	x			x	x		x
Vpn17	keine Angaben												
Vpn18	Gesamteindruck	x			x					x			
Summe		5	10	7	8	3	1	5	10	6	9	1	2

Erstrater

Zweitater

Farben
Iconform
einzelnes Icon
mehrere Icons
Restkategorie
Wechsel
Farben
Iconform
einzelnes Icon
mehrere Icons
Restkategorie
Wechsel

Nicht-erfolgreiche Löser

Vpn1	Kategorienbildung nach "innerer" Färbung des Motives, also 1. Kategorie: blau + rot 2. Kategorie: gelb + grün unabhängig von Form und Farbe des Rahmens.	x			x			x				x	
Vpn2	Wechsel vor Wechsel 1 Motiv rund/zackig 2. Teil an Symbol orientiert.		x	x				x		x	x		
Vpn3	keine Angaben												
Vpn4	Gesamteindruck	x			x				x			x	
Vpn5	keine Angaben												
Vpn6	Farbe Form einzelnes Icon	x		x						x	x		
Vpn7	Farben	x			x				x			x	
Vpn8	anfangs gar keine Kategorien, dann verschiedens ausprobiert, am Ende nach farben geordnet (mehr blau oder mehr rot).	x			x			x	x			x	
Vpn9	Farbe und Lage	x			x				x			x	

Vpn10	hell - dunkel	x			x			x			x									
Vpn11	keine Angaben																			
Vpn12	Kategorie links: mehr blaue Icons Kategorie rechts: mehr rote Icons	x			x			x				x								
Vpn13	Farbe und geometri	x			x			x				x								
Vpn14	eckig vs rund		x		x				x			x								
Vpn15	Farbeindruck	x			x			x				x								
Vpn16	Zuerst nach helleren und dunkleren Bildern, dann hab ich mir gewisse Zeichen gemerkt und danach sortiert, also eher große helle Zeichen in die 1. Kategorie und viele kleine kantige dunkle Zeichen in die 2. Kategorie, großes "&" Zeichen in die 1. Kategorie usw. Aber ein richtiges System hatte ich glaub ich nicht wirklich, dazu warn die Bilderabfolge zu schnell.		x		x			x	x			x								
Vpn17	ueberblick	x			x			x				x								
Vpn18	Form		x		x					x			x							
Summe			11		4		3	12		0		3	11		4	3	12		0	0

Anhang 9:

Computersimulation.

Konnektionistisches Delta-Modell I

Das Modell bestand aus 72 Eingabeknoten und 3 Ausgabeknoten. Jeder Eingabeknoten repräsentierte genau 1 Icon. Durch die orthogonale Kodierung wurden eventuell bestehende Ähnlichkeiten zwischen den Icons in der Eingabe nicht berücksichtigt. Alle Eingabeknoten waren mit allen Ausgabeknoten über veränderbare Gewichte verbunden. Die Gewichte wurden am Anfang der Simulation mit Zufallswerten aus einer Gleichverteilung über dem Intervall 0 bis $\frac{2}{72}$ initialisiert. Durch die Gleichverteilung konnte erwartet werden, dass die realisierten Gewichte sich symmetrisch um $\frac{1}{72}$ verteilen und somit der Erwartungswert für die Gewichte $\frac{1}{72}$ beträgt. Damit sollten sich die Gewichte über alle 72 Eingabeknoten hinweg zu ungefähr 1 aufsummieren.

Nach jeder „Darbietung“ eines Stimulus wurden die Gewichte zu dem gewinnenden Ausgabeneuron nach folgender Regel aufgefrischt:

$$\Delta w_{ij} = \frac{\beta(1 - o_j)a_i}{n}; i \in \{\text{aktive Eingabeknoten}\}, n \text{ Anzahl aktiver Eingabeknoten}$$

und

$$\Delta w_{ij} = \beta(1 - o_j)w_{ij}; i \in \{\text{inaktive Eingabeknoten}\}$$

Der Lernparameter Beta wurde über die Simulationen hinweg von 0.001 bis 0.5 in Schritten der Größe 0.01 erhöht.

Für jeden Parameterwert β wurden 24 Simulationen durchgeführt. Die Eingabeneurone wurden entsprechend den 24 unterschiedlichen Präsentationsfiles, die den Versuchspersonen präsentiert worden waren, initialisiert. In der Bildschirmpräsentation dargebotene Icons wurden durch Eingabeneurone repräsentiert, welche die Aktivierungsstärke 1 erhielten. Alle anderen Eingabeneurone erhielten die Aktivierungsstärke 0.

Die Kategorienentscheidung entsprach dem am stärksten aktivierten Ausgabeneuron.

Konnektionistisches Delta-Modell II

Die einzigen Unterschiede zu Modell I bestanden darin, dass am Anfang der Phase 2, die Gewichte zu Ausgabeknoten 3 erneut zufällig initialisiert wurden (Wiedererkennungshypothese) und die Veränderungsregel bei aktiven Gewichten um den Salienzparameter (Habituationshypothese) ergänzt wurde:

$$\Delta w_{ij} = \frac{\beta(1-o_j)a_i s_{icon}}{n}; i \in \{\text{aktive Eingabeknoten}\}$$

Hierbei gibt „ s_{icon} “ die Salienzstärke der jeweiligen aktiven Icons an.

Die Salienzstärke wurde bei der ersten Präsentation eines Icons auf 1 gesetzt. Die Salienzstärke nahm nach jeder Iconpräsentation um einen durch den Salienzparameter festgelegten Prozentwert ab:

$$s_{icon}(t) = s_{icon}(t-1) * (1 - s_{parameter})$$

„ t “ indiziert dabei den Zeitpunkt bzw. die Anzahl stattgefundener Iconpräsentationen. Der Salienzparameter variierte, wie die Lernrate, im Bereich von 0 bis 0.5 und mit Schrittgröße 0.01.

Für jede Parameterkombination erfolgten 24 Simulationen.

Konnektionistisches Rumelhart & Zipser-Modell I

Das Netzwerk entsprach dem Delta-Modell I. Lediglich änderte sich der Update-Algorithmus zu:

$$\Delta w_{ij} = \begin{cases} 0 & \text{Falls Einheit j bei Vorliegen von Stimulus i verliert} \\ \beta_1 \frac{a_i}{n} - \beta_2 w_{ij} & \text{Falls Einheit j bei Vorliegen von Stimulus i gewinnt} \end{cases}$$

Gegenüber dem Rumelhart und Zipser Modell (1986) wurde als weiterer Parameter β_2 eingeführt, um dem Modell mehr Möglichkeiten einzuräumen und damit einer Widerlegung zu entgehen. Während β_1 die Lernrate repräsentiert, regelt β_2 den Zerfall von konkurrierenden Gewichtsverbindungen und wird deshalb mit dem Ausdruck „Decay“-Parameter referenziert.

Konnektionistisches Rumelhart & Zipser-Modell II

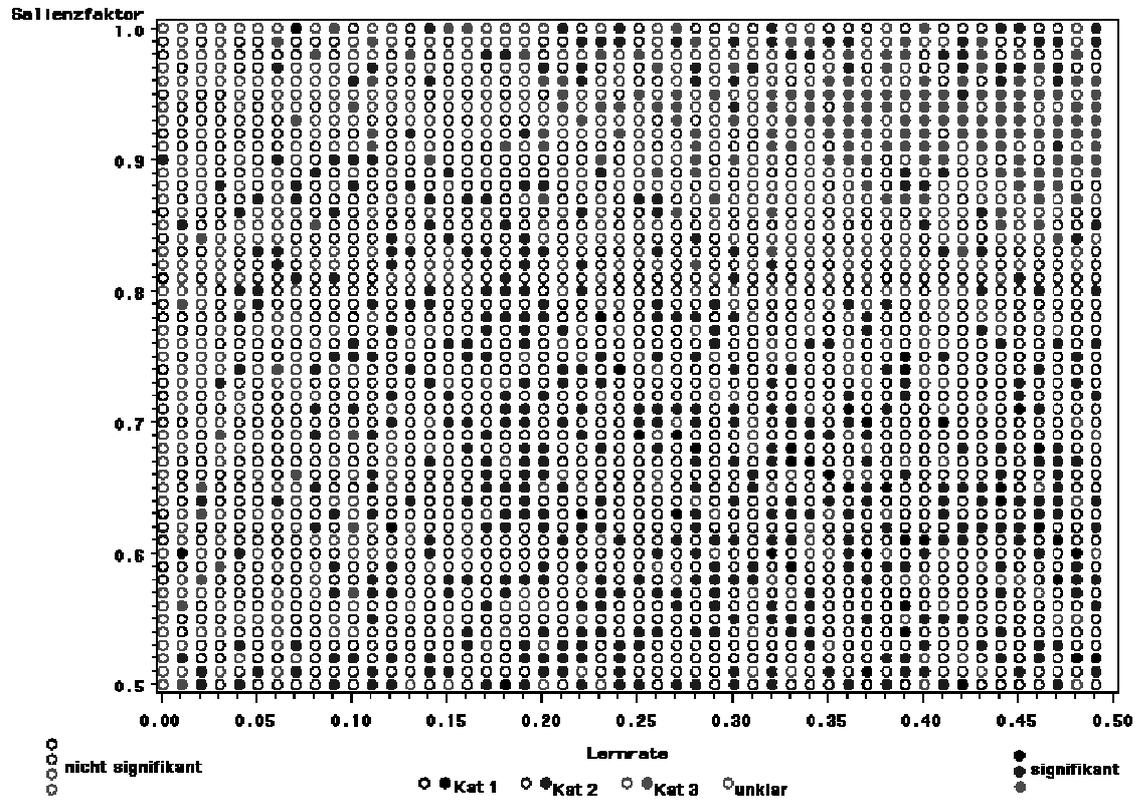
Das Netzwerk entsprach dem Delta-Modell II. Der Update-Algorithmus veränderte sich zu:

$$\Delta w_{ij} = \begin{cases} 0 & \text{Falls Einheit j bei Vorliegen von Stimulus i verliert} \\ \beta_1 \frac{a_i}{n} s_{icon} - \beta_2 w_{ij} & \text{Falls Einheit j bei Vorliegen von Stimulus i gewinnt} \end{cases}$$

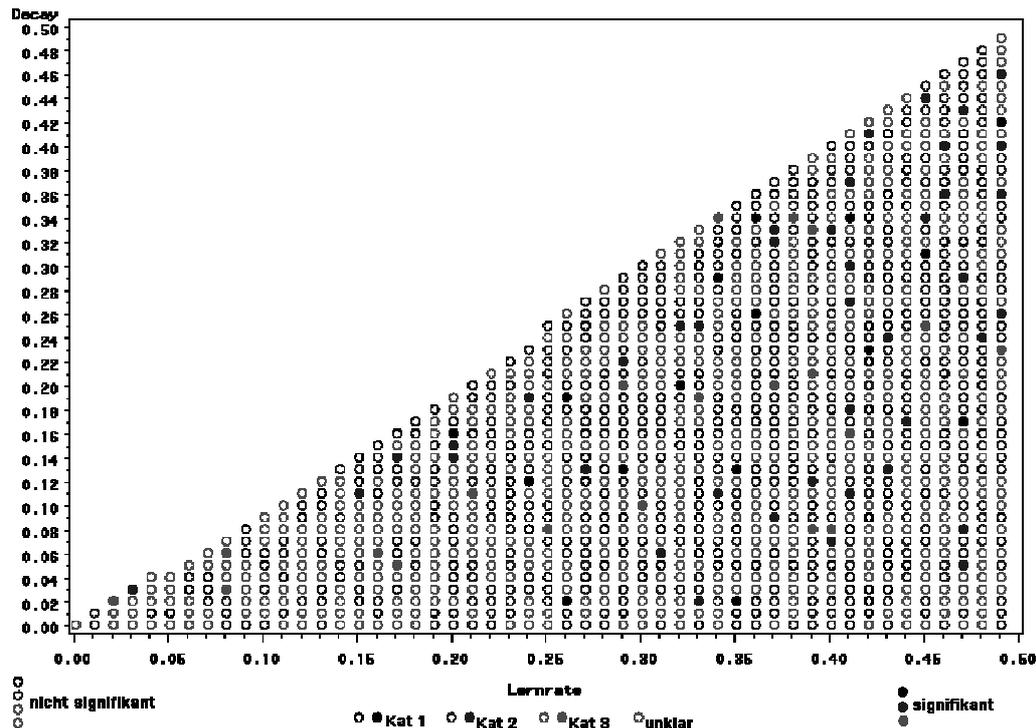
Anhang 10:

Grafische Darstellung der Parameterauswirkung auf die Kategorienentscheidung.

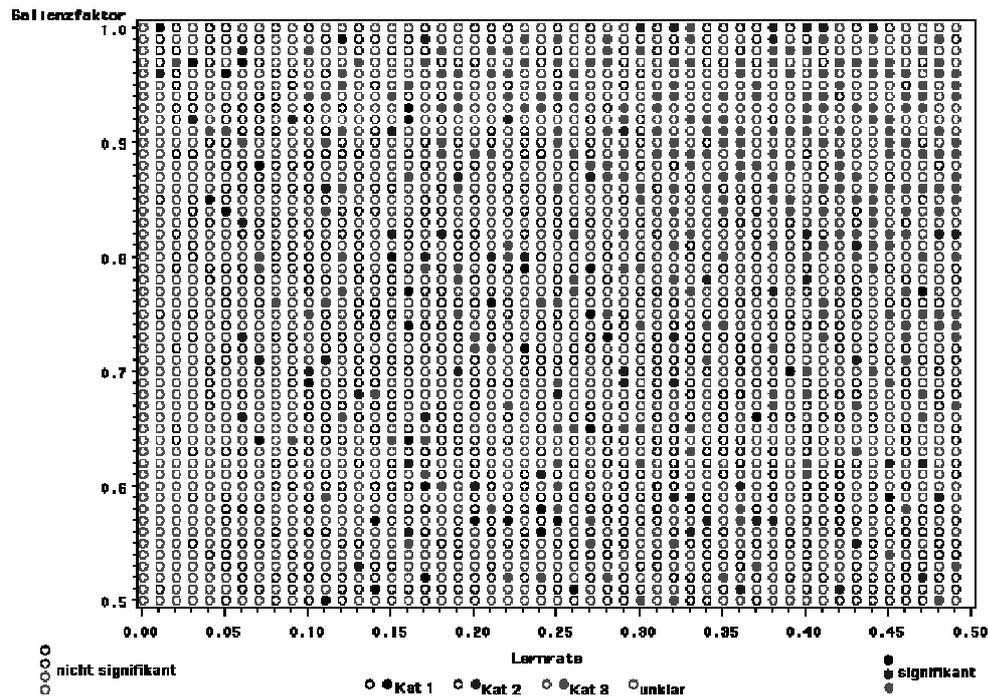
a) divisionsbegrenzter Algorithmus



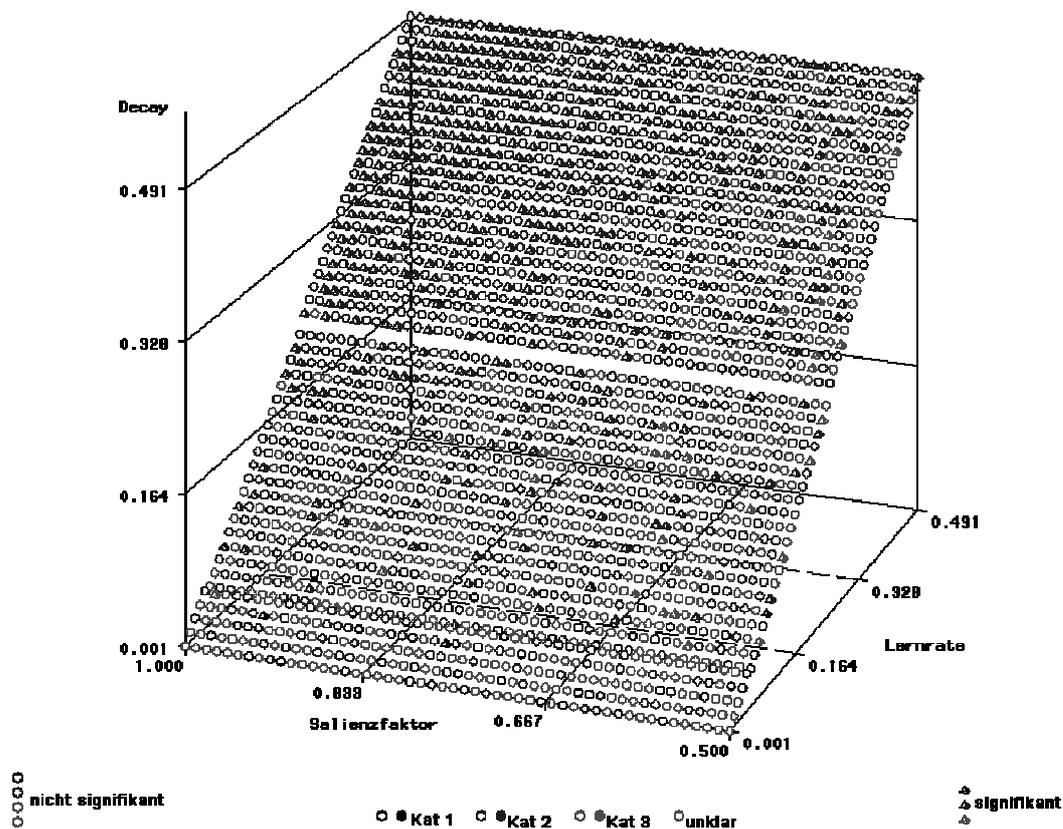
b) Rummelhart & Zipser Modell, mit ratio rule. $\Theta=20$.



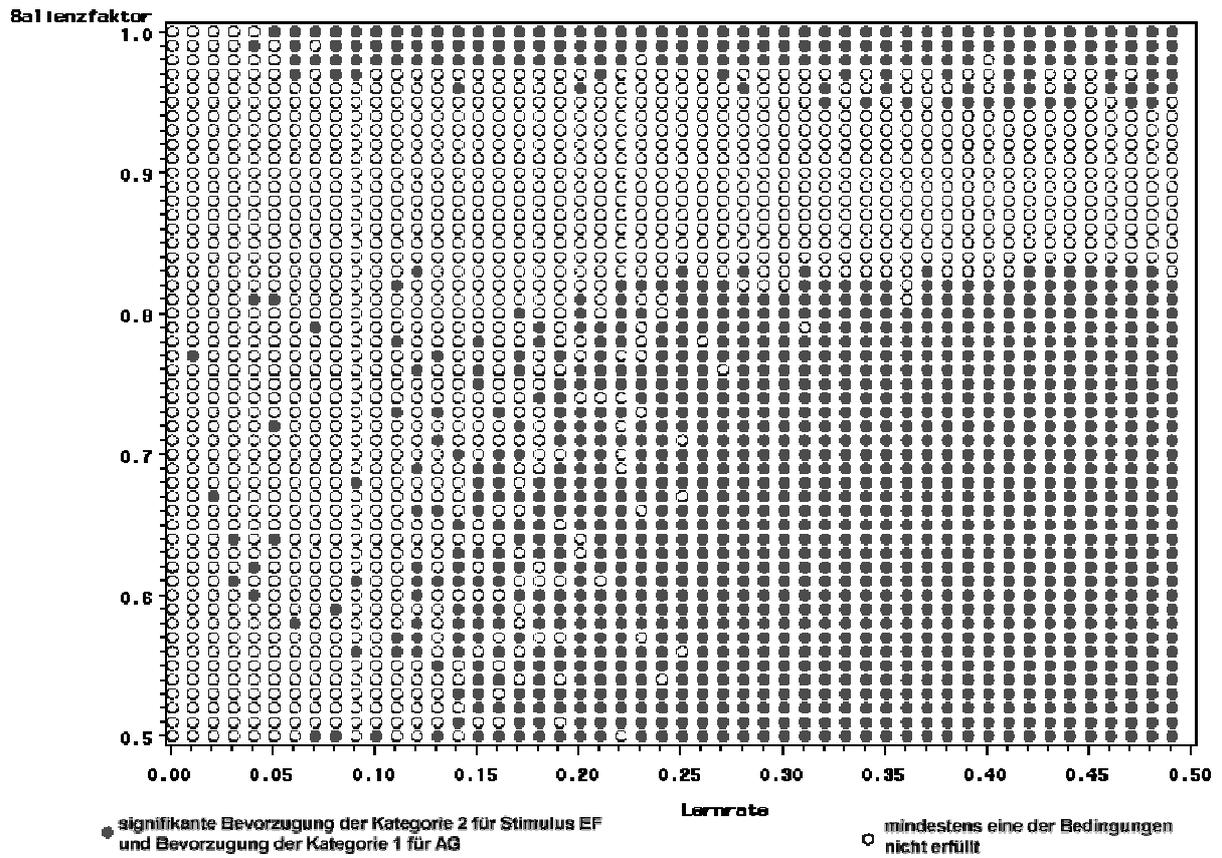
c) Delta Modell, $\Theta=20$.



d) Rummelhart & Zipser mit Salienzparameter ohne Θ -Rauschen. Aus Darstellungsgründen sind nur Punkte für maximale Decay-Rate dargestellt.



e) Delta-Modell II mit Performanz



Anhang 11

Herleitung der Wahrscheinlichkeit, dass ein vertauschtes Icon die Kategorienzuordnung bestimmt.

Im Folgenden wird davon ausgegangen, dass für jede der 3 Kategorien genau 1 indikatives Icon existiert, welches bei alleinigem Auftreten zu der entsprechenden Kategorienentscheidung führt. Wenn das in Kategorie Y transferierte indikative Icon x gemeinsam mit einem indikativen Icon aus Kategorie Y auftritt, beträgt die Wahrscheinlichkeit, dass Icon x die Kategorienzuordnung bestimmt, da zwei gleichwertige Alternativen bestehen, 0.5. Mit Wahrscheinlichkeit 0.10 ist aber zusätzlich das indikative Icon aus Y ausgetauscht worden. In diesem Fall fällt die Entscheidung garantiert entsprechend Icon x aus:

$P(\text{Entscheidung basiert auf Icon } x \mid x \text{ ist indikatives Icon und tritt im Stimulus einer anderen Kategorie auf}) = 0.9 \cdot 0.5 + 0.10 \cdot 1 = 0.55.$

Herleitung der Wahrscheinlichkeit, dass über alle Versuchspersonen in Experiment 2 hinweg bei Vorliegen der Ein-Icon-Strategie kein einziges mal alle 3 Icons identifizieren werden.

Die Wahrscheinlichkeit, dass ein bestimmtes Icon vertauscht wird beträgt 0.10. Die Wahrscheinlichkeit für ein bestimmtes Icon vertauscht und erkannt zu werden beträgt damit $0.10 \cdot 0.55 = 0.055$.

Dass das indikative Icon bei 8 Stimuluspräsentationen wenigstens einmal vertauscht und erkannt wird, ergibt sich aus der Gegenwahrscheinlichkeit:

$$1 - (1 - 0.055)^8 = 0.36.$$

Dass in einer Phase aus allen 3 Kategorien die entscheidenden Icons erkannt werden, ergibt sich zu:

$$0.36^3 = 0.05.$$

Dass sowohl in Phase 1 als auch Phase 2 nicht alle 3 Icons erkannt werden, beträgt:

$$(1 - 0.05)^2 = 0.91.$$

Dass dieses Ereignis für alle 17 Versuchspersonen eintritt, beträgt:

$$0.91^{17} = 0.19.$$

Herleitung der Wahrscheinlichkeit, dass über alle Versuchspersonen in Experiment 2 hinweg bei Vorliegen der Ein-Icon-Strategie wenigstens 2 Icons identifiziert werden.

Dass in einer Phase wenigstens 2 entscheidende Icons erkannt werden, ergibt sich zu:

$$0.36^2 \cdot 0.64 \cdot 3 + 0.36^3 = 0.25$$

Dass in mindestens einer Phase wenigstens 2 Icons identifiziert werden können, beträgt:

$$(1 - 0.25)^2 = 0.56.$$

Dass dieses Ereignis für alle 17 Versuchspersonen höchstens 1 mal auftritt, beträgt:

$$0.56 \cdot (1 - 0.56)^{16} \cdot 17 + (1 - 0.56)^{17} = 2.09 \cdot 10^{-5}.$$