

---

Bodo Krause

## **Erkennen diskreter Merkmalsstrukturen. Ein Beitrag zur Untersuchung und Modellierung menschlicher Lernprozesse**

Vortrag in der Klasse Naturwissenschaften der Leibniz-Sozietät am 20. Juni 2002

### **1. Zum Gegenstand**

Die Erkennung diskreter Merkmalsstrukturen ist eine Basisleistung (nicht nur) der menschlichen Intelligenz und ihre Erklärung steht seit jeher im Interesse der Forscher. Es ist eine elementare Leistung, die für die Erhaltung und das Überleben von Arten und Rassen von entscheidender Bedeutung war und ist, und für die es zahlreiche Beispiele gibt:

- das Erkennen von genießbaren und giftigen Pflanzen und Tieren an Farbe, Form, Geruch, ...,
- das Erkennen einer gefährlichen Situation an der Verletzung des Sicherheitsabstands, an Drohgebärden, ...,
- das Erkennen des Unterschieds zwischen Eltern und anderen Erwachsenen durch ein Kleinkind.

Überlebenswichtig ist das schnelle und sichere Erkennen, das dann ein angemessenes Verhaltensprogramm (Vermeidung, Flucht, Zuwendung, ...) auslöst.

Klix (1980, S. 10) kennzeichnet die entwicklungsgeschichtliche Herausbildung und Bedeutung dieser Leistung wie folgt:

„Lernen ... bildet individuelles Gedächtnis aus. Denken beruht auf der Nutzung des Gedächtnisses. ... Suchen aktualisiert Erfahrungen, ermöglicht Vergleiche zwischen ihnen und dabei das Erkennen neuer Zusammenhänge. Denken ist, seinem Wesen nach, Beschaffung von Information durch kognitive Prozesse. Die Funktion dieser Prozesse muss sich bei Entscheidungssicherheit bewähren. ... Dies ist die Motivlage für die Extrapolation von Künftigem durch Denkprozesse.“

## 2. Kennzeichnung diskreter Merkmalsstrukturen und ihrer Bedeutung für den Erkennungsprozess

Diskrete Merkmalsstrukturen sind Verknüpfungen von Merkmalen, die die Zugehörigkeit von Objekten  $O_i$  zu einer Objektmenge  $O = \{O_1, \dots, O_m\}$  kennzeichnen oder entscheidbar machen. Dabei sind die einzelnen Objekte hinsichtlich von Merkmalsdimensionen  $D_1, \dots, D_r$  beschrieben. Jede einzelne Dimension  $D_i$  ist diskret gestuft (z.B. in die wahrnehmbaren Einheiten), im einfachsten Fall dichotom (ein Merkmal ist vorhanden oder nicht). Der potentielle Objektraum  $OR$ , der durch die Dimensionen  $D_1, \dots, D_r$  aufgespannt wird, ist dann durch das direkte Produkt  $D_1 \times \dots \times D_r$  beschreibbar. Eine ausgewählte Objektmenge  $O$  ist demnach stets eine Teilmenge dieses Objektraums  $OR$ , also  $O \subseteq OR$ . So wählt die Kombination der Merkmale sparsam und klein beim Autokauf eine bestimmte Teilmenge von potentiellen Kandidaten wie Smart, Mini, Lupo, ... aus der Gesamtmenge der PKW aus.

In diesem Kontext bezieht sich der Begriff Merkmalsstruktur darauf, dass eine Teilmenge  $OT$  von Objekten aus  $OR$  durch invariante Merkmale und / oder invariante Merkmalsverknüpfungen entsteht und sich von anderen Teilmengen von Objekten unterscheiden lässt. In der Psychologie bezeichnet man den Prozess der Herausbildung solcher Merkmalsstrukturen als Diskriminationslernen (in elementarer Form auch als Konditionieren) oder, wenn er sprachgebunden ist, als Begriffsbildung. Aristoteles kennzeichnet diesen Aspekt des Denkens (in Band 13 „Über die Seele“, S. 59) „Das Denken der ungeteilten Begriffe gehört zu dem Gebiet, wo es keinen Irrtum gibt. Wo aber Irrtum und Wahrheit herrscht, da gibt es schon eine Verknüpfung der Begriffe zu einer Einheit“.

Dies macht deutlich, dass für die Kennzeichnung diskreter Merkmalsstrukturen neben der Merkmalscharakteristik der Objekte auch deren Verknüpfungen sowie ihr Wahrheitswert bedeutsam sind. Kennzeichnet man also Schüler einer bestimmten Klasse durch die Noten in den einzelnen Unterrichtsfächern (Merkmalscharakteristik), dann können sich Teilmengen von Schülern dadurch auszeichnen, dass sie gute Ausprägungen in den naturwissenschaftlichen Fächern haben oder andere sich durch ihre Leistungen in Sport, Musik und Ethik kennzeichnen lassen. Fragt man nun umgekehrt nach der Merkmalsstruktur einer Schülergruppe, die durch häufige Fehlzeiten auffällig werden, dann ist man bei dem angesprochenen Problem der Erkennung einer diskreten Merkmalsstruktur, wobei diese dann auch mehr oder weniger zutreffend sein kann. Allgemein ist dies das Problem einer Diagnosefindung.

### 3. Klassische Beispiele und Erklärungsansätze der Erkennung diskreter Merkmalsstrukturen

Für das Erkennen diskreter Merkmalsstrukturen im Sinne dieses Konditionierens oder Diskriminationslernens gibt es typische Beispiele und zwei grundsätzliche Erklärungsansätze, die bis zur Gegenwart in der psychologischen Literatur z.T. auch kontrovers diskutiert werden:

#### 3.1 Klassische Ansätze auf der Grundlage des Assoziationsprinzips.

Sie beruhen auf der raum-zeitlichen Kontiguität oder Kontingenz von Merkmalsausprägungen oder Reizen. Klassisches Beispiel ist die Ausbildung eines bedingten Reflexes (nach Pavlov, 1927), bei der das gleichzeitige Auftreten eines Tons bei der Futtergabe eines Hundes dazu führt, dass im eingelernten (konditionierten) Zustand auch die alleinige Darbietung des Tones zur Reaktion Speichelfluss führt. Experimentell mehrfach nachgewiesen ist, dass die Stärke dieser Konditionierung der Manifestationshäufigkeit des Tones proportional ist. Genauer, die Effektivität des Konditionierens  $E(K)$  ist eine Funktion der Manifestationshäufigkeit des unkonditionierten Reizes US:

$$E(K) = f \{ P(US | CS) \}.$$

Dies benennt bereits Thorndike (1911, S. 244) als das "Law of Effect";

"Of several responses made to the same situation, those which are accompanied or closely followed by satisfaction to the animal will, other things being equal, be more firmly connected with the situation."

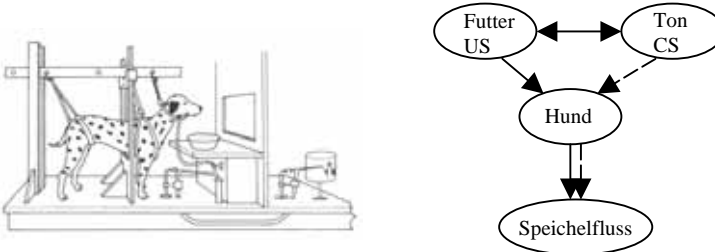


Abbildung 1:

Untersuchungssituation des Hundes (Grafik aus Pearce, 1997) und Strukturschema der Ausbildung eines bedingten Reflexes.

Auf diesem Zugang basieren die klassischen S-R-Theorien des Lernens, wie sie von Hull (1884–1952), Guthrie (1886–1959) und Tolman (1886–1959) (letzterer für das instrumentelle Bedingen) entworfen wurden.

Pearce (1997) betont in diesem Rahmen, dass „associative learning ... takes place when there is a change in an animal’s behaviour as a result of one event being paired with another.“ Die klassischen Untersuchungsmethoden hierfür sind das Pavlovsche Konditionieren und das instrumentelle Bedingen, bei dem das Verhalten das Ereignis hervorruft.

Pavlov (1927) berichtet auch über ein weiteres Experiment, das 1917 mit Hunden zum Diskriminationslernen durchgeführt wurde. Hungrigen Hunden wurden zwei Reize dargeboten:

- S1 ein bunt ausgemalter Kreis
- S2 ein bunt ausgemaltes Quadrat,

wobei der bunte Kreis die bevorstehende Fütterung signalisierte. Auch hier kam es zu dem Effekt, das mit zunehmender Versuchsdauer der Hund auf den S1 mit Speichelfluss reagierte.

Einen weiterführenden Erklärungsansatz begründet die Theorie des inkrementellen Lernzuwachses von Spence (1936, 1937). Sie geht davon aus, dass jede Belohnung einer Reizmenge zu einem Zuwachs (increment) in der Tendenz führt, künftig diese Verhaltensantwort bei jedem dieser Reize zu geben. Dabei gilt auch umgekehrt, dass das Ausbleiben der Belohnung diese Tendenz verringert.

Dies begründet excitatorische und inhibitorische Auswirkungen auf den Lernprozess, genauer: die Stärke der assoziativen Bindungen werden durch die Belohnung bzw. deren Ausbleiben verstärkt oder gehemmt.

Die Assoziationsstärke oder Effizienz  $E(\text{CS})$  wird damit eine Funktion, die von dem Verhältnis des Eintretens des Ereignisses (US, Belohnung) nach dem Eintreten und dem Nichteintreten des CS abhängt ist:

$$E(\text{CS}) = f \{ P(\text{US} | \text{CS}), P(\text{US} | \sim\text{CS}) \}.$$

### 3.2 Weiterführende Ansätze zum Assoziationslernen

In Erweiterung der Gedanken von Spence wurden eine Reihe weiterer Erklärungsmodelle entwickelt, die im Sinne eines Kontingenztmodells (Jenkins & Ward, 1965; Rescorla, 1968; Salmon, 1965) inkrementelles Lernen durch die Effizienz (Kontingenz- oder Bindungsstärke) eines verursachenden Stimulus  $E(\text{CS})$  für ein Ereignis  $e$  (den US) wie folgt zu fassen:

$$E(\text{CS}) = P(e | \text{CS}) - P(e | \sim \text{CS}).$$

Als weithin akzeptierter Modellansatz ist das Modell von Rescorla – Wagner (1972) sehr verbreitet. Es ist daran orientiert, die excitatorischen und inhibitorischen Wirkungen im Lernen als Veränderungen in der Effizienz (Bindungsstärke) zu modellieren:

- für excitatorische Wirkungen für  $\text{CS}_i$ :  $\Delta E_i = \alpha_i \cdot \beta_1 \cdot (\lambda - \sum E_k)$   
mit  $\alpha_i$  = bestehende Intensität von  $\text{CS}_i$   
 $\beta_1$  = Lernrate von US-trials  
 $\lambda$  = Maximal mögliches Level der Assoziationsstärke konditionierbar zur US-Intensität

$E_k$  = alle anderen auftretenden CS – Stimuli und US

- für inhibitorische Wirkungen des  $\text{CS}_i$ :  $\Delta E_i = \alpha_i \cdot \beta_2 \cdot (0 - \sum E_k)$   
mit  $\beta_2$  = Veränderungsrate bei Nichtbegräftigung.

Für das Diskriminationslernen auf der Grundlage von vier dichotomen Merkmalen könnte sich ein neuronales Netz (vgl. Gluck & Bower, 1988; Estes u.a., 1989) wie folgt aufgebaut sein:

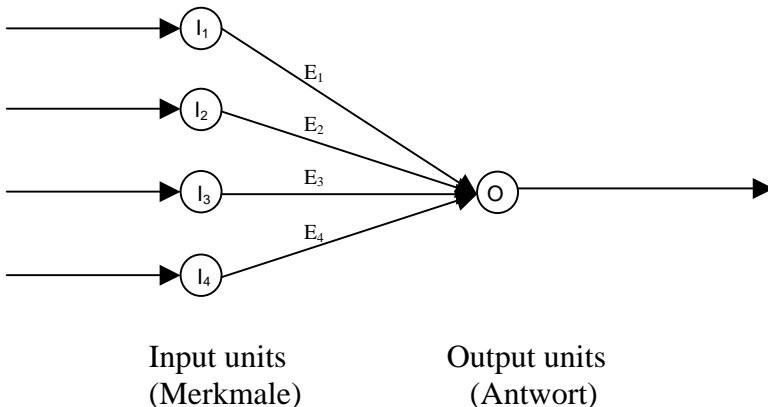


Abbildung 2:

Neuronales Netz zum Rescorla-Wagner-Modell

Wichtige Erkenntnisse zum Rescorla-Wagner-Modell (R-W-M) (nach Pearce, 1997) sind:

- Dies so formalisierte Modell erlaubt es, empirisch prüfbare Hypothesen aufzustellen.
- Diese Vorgehensweise führte in den meisten Fällen zur Modellkonformität.

- Obwohl das R-W-M nicht alle Lernbefunde erklären kann verbleibt die Konklusion von Pearce „... that the Rescorla-Wagner model provides the ideal account of the conditions of learning“.
- Das R-W-M ist auch in Form eines neuronalen Netzwerkes (Perzeptron) darstellbar (Gluck & Bower, 1988; Estes u.a., 1989). Dabei konnten Sutton und Barto (1981) nachweisen, dass die Lösungen nach dem R-W-M und nach dem Prinzip der LMS (least mean squares) eines neuronalen (backpropagation) Netzwerkes übereinstimmen: „the LMS rule is essentially identical to the Rescorla-Wagner (1972) model of associative learning in animals“.

### Fazit 1:

Assoziationslernen, wie es beim Konditionieren und Diskriminationslernen auftritt, ist weitgehend durch die Mechanismen des R-W-M erklärbar und damit durch neuronale Netze modellierbar.

### 3.3 Lernen nach dem Abstraktionsprinzip

Grundlage des Lernens nach dem Abstraktionsprinzip ist eine Vorgehensweise, bei der von konkreten Merkmalsausprägungen abstrahiert wird und dann Merkmalsbeziehungen zur Erkennung nutzt (Relationenlernen, Funktionenlernen, Merkmalsgeneralisierung, Regellernen). Dies sind Formen des induktiven und analogen Schließens. Ein klassisches Beispiel hierfür sind die Untersuchungen Köhlers (1918) am Haushuhn:

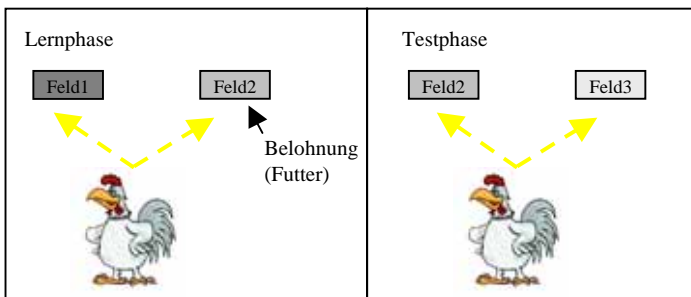


Abbildung 3:  
Versuchsschema zum Diskriminationslernen mit Lern- und Testphase

Dargeboten werden zwei Reizelemente, ein helleres und ein dunkleres, von denen eins z.B. das hellere durch Futter belohnt wird. Nach der Lernphase folgt der Kontrollversuch: Dargeboten wird weiterhin das bisher hellere Reiz-

element, dazu kommt aber jetzt als zweites Reizelement ein noch helleres. Die Verhaltensentscheidung des eingelernten Haushuhns traf nun nicht das bisher hellere Reizelement (das ja auch vielfach dargeboten war), sondern das neue noch hellere Reizelement, das vorher nie dargeboten war. Die Befundinterpretation ist eindeutig: Das Huhn hat nicht eine Merkmalsausprägung „gelernt“ sondern die Merkmalsrelation „auf das hellere Reizelement“.

Ungeklärt bleibt, warum das Haushuhn nicht die Reizausprägung sondern die Reizrelation konditioniert hat. Köhlers Erklärung war, dass das Diskriminationslernen in diesem Fall dadurch entsteht, dass die Hühner beide Reize inspizieren und sich dann für das hellere Feld entscheiden.

Aber auch Spence (1937), zu dessen ursprünglichem Erklärungsansatz dieser Befund im Widerspruch stand, fand eine Erklärung, die die Diskriminationsleistung auf einer Dimension als Reizgeneralisierung (und damit, wie wir unten zeigen werden, eigentlich als Abstraktionsleistung) erklärt.

### **Fazit 2:**

Schon bei einfachen Diskriminationsleistungen ist das rein assoziationsbasierte Erklärungsprinzip nicht ausreichend. Höhere Erkennungsprinzipien, die auf dem Abstraktionsprinzip beruhen, greifen.

## **4. Ausgewählte neuere Untersuchungs- und Modellierungsansätze zum Erkennen diskreter Merkmalsstrukturen**

Wir wollen ausgewählte Untersuchungsansätze mit wesentlichen Ergebnissen und Problemen darstellen, um damit den eigenen Zugang zur Untersuchung und Modellierung dieser Lernprozesse zu begründen.

### **4.1 Psychophysiologisch gestützte Befunde zum Diskriminationslernen (Lachnit und Mitarbeiter, 1992–1996)**

Lachnit und Mitarbeiter verwendeten die Konditionierungsparadigmen des „positive patterning“ und des „negative patterning“ auf der Grundlage der Darbietung von Buchstabenpaaren. Schematisch sah das so aus:



*Abbildung 4:  
Buchstabenpaare für das Konditionierungsparadigma von Lachnit u. Mitarb.*

Der Unterschied der beiden Paradigmen bestand in dem auslösenden Ereignis, das einen spürbaren Elektroschock bewirkte. Dies konnte ein Einzelreiz (negative patterning) oder das Reizpaar (positive patterning) sein, wie dies in der linken Seite der nachfolgenden Abbildung 5 gekennzeichnet ist:

**positive patterning**

	B	~B
A	1	0
~A	0	?

1 = event (Schock)  
0 = no event

**AND**

	B	~B
A	1	0
~A	0	0

1 = wahr  
0 = falsch

**negative patterning**

	B	~B
A	0	1
~A	1	?

**XOR**

	B	~B
A	0	1
~A	1	0

Abbildung 5:

Positive / negative patterning untereinander auf linker Hälfte, die „vergleichbaren“ logischen Verknüpfungen AND und XOR rechte Hälfte.

Durch die Variation der Anzahl der verwendeten Buchstabenpaare bei gleicher Länge der Lernliste konnte die Manifestationshäufigkeit der auslösenden Reize variiert werden. Danach erfolgte der kritische Versuch, bei dem ein neues, bisher nicht verwendetes Buchstabenpaar dargeboten und das ausgelöste Verhalten über physiologische Parameter registriert wurde.

Das Ergebnis war ein interessanter, aber auch widersprüchlicher Befund: Bei beiden Anforderungen zeigte sich, dass die Variation der Manifestationshäufigkeit keinen Einfluss auf den Konditionierungseffekt hatte. Eine Diskrimination der neuen Reizpaare wurde nur beim positive patterning übertragen. Der erste Befund belegt, dass in dieser Anforderung die *Manifestationshäufigkeit keine entscheidende Einflussgröße auf die Effektivität des Konditionierens* ist, der Befund also nicht durch reines Assoziieren erklärbar wird. Als Erklärung bieten die Autoren einen Regelerwerb an. Aus unserem Ansatz (s.u.) ergibt sich auch eine Erklärung im Sinne einer Merkmalsabstraktion, der aber über die Erklärungen von Köhler und Spence hinausgeht. Der zweite



Befund kennzeichnet, dass *im Fall des negative patterning offensichtlich kein generalisiertes Lernergebnis* eintritt. Dies bedeutet eigentlich, obwohl von den Autoren nicht so interpretiert, dass damit die logische Verbindung XOR nicht ausgebildet wird !

Dennoch ziehen die Autoren eine Parallele der beiden Konditionierungsparadigmen zu den Wahrheitstabellen der logischen Funktionen AND und XOR. Diese Parallele zeigt sich in der Gegenüberstellung der obigen Abbildung 4 auf der linken Seite, wobei wir hier darauf verweisen, dass sich die Definitionsbereiche beider Strukturen dadurch unterscheiden, dass in den Konditionierungsparadigmen ein Fall, nämlich die (0, 0)-Darbietung fehlt. Hier setzt ein weiterführender Gedanke an, nämlich durch Realisierung dieses Falls (0, 0) eine Entscheidungsmöglichkeit zu schaffen, ob beim Konditionieren wirklich die angegebene logische Struktur erworben wird und damit die Konditionierungen dem Erlernen logischer Verknüpfungen entsprechen.

#### 4.2 Die Arbeiten von Thornton zum Parity-Problem (1993, 1996 a, b)

Der Zugang von Thornton ist, in unserem Zusammenhang, besonders interessant, weil er den Aspekt der Generalisierung besonders hervorhebt: „I will begin by noting that the problem of learning is essentially the problem of generalization (Learning problems which do not involve generalization are effectively storage problems). Thus, solving a learning problem must involve exploiting whatever generalization information is contained within feedback available to the learner.“ (Thornton, 1994, S. 2).

Dies ist genau der Zugang, den bereits Klix (1971, S. 347) mit folgender Lerndefinition kennzeichnet: „Mit Lernen bezeichnen wir danach jede umgebungsbezogene Verhaltensänderung, die als Folge einer individuellen (systemeigenen) Informationsverarbeitung eintritt.“. Mit diesem Zugang haben wir versucht, einen Ansatz einer kognitiven Lerntheorie zu begründen (Krause, 1997), der genau diesen Generalisierungsaspekt im Sinne eines induktiven Lernprozesses kennzeichnet.

Clark & Thornton (1993, 1997) stellen allgemein Lernprobleme als Abbildung zwischen Inputs und Outputs dar. Lernziel ist es dann, zu Inputs mit hoher Wahrscheinlichkeit den korrekten Output zuzuordnen. Entscheidend dafür sind die bedingten Wahrscheinlichkeiten  $P\{O | I=i\}$  eines Outputs O unter der Bedingung des Anliegens eines konkreten Input i.

Unter diesem Gesichtspunkt sollten dann die bedingten Wahrscheinlichkeiten  $P(O | I)$  eine besondere Rolle für das Lernen spielen. Sind sie groß /

klein, dann werden sie den Lernprozess stützen, sind sie alle gleich, das sind dann die sogenannten Parity-Probleme, dann liefern diese bedingten Wahrscheinlichkeiten keine lernrelevante Information. Solche Probleme werden als statistisch neutral bezeichnet. Das XOR ist allgemein eine solche Verknüpfung, wie man am zweistelligen XOR einfach sieht:

Wertetabelle der Funktion  $I_1 \text{ XOR } I_2$  :

	<b><math>I_2</math></b>	<b>1</b>	<b>0</b>
<b><math>I_1</math></b>			
<b>1</b>		0	1
<b>0</b>		1	0

Betrachtet man nun die bedingten Wahrscheinlichkeiten in Abhängigkeit vom Input, dann ergibt sich die nachfolgende Tabelle, die diese Paritäten ausweist:

<b>Bedingung B</b>	<b><math>P\{ O=0 \mid B \}</math></b>	<b><math>P\{ O=1 \mid B \}</math></b>
<b>Keine</b>	$\frac{1}{2}$	$\frac{1}{2}$
<b><math>I_1 = 1</math> (Reiz 1 liegt an)</b>	$\frac{1}{2}$	$\frac{1}{2}$
<b><math>I_1 = 0</math> (Reiz 1 liegt nicht an)</b>	$\frac{1}{2}$	$\frac{1}{2}$
<b><math>I_2 = 1</math> (Reiz 2 liegt an)</b>	$\frac{1}{2}$	$\frac{1}{2}$
<b><math>I_2 = 0</math> (Reiz 2 liegt nicht an)</b>	$\frac{1}{2}$	$\frac{1}{2}$

Thornton begründet nun, dass der *Grad der statistischen Neutralität im Sinne eines Schwierigkeitsindex den Lernprozess beeinflusst* und damit Parity-Probleme durch neuronale Netze und damit assoziatives Lernen nicht lösbar sind. Dies ist eine Generalisierung des Beweises von Minsky und Papert (1962), die für Perzeptrone die Unmöglichkeit der Lösung des XOR-Problems nachgewiesen haben. Thorntons (1996 a) Schlussfolgerung ist „Backpropagation in fact reliably false to solve parity problems . . . ., i.e., as problems involving generalisation.“

Für uns wird dies Anlass zur Frage, ob diese Eigenschaft der statistischen Neutralität ein Indikator für die Unterscheidung von menschlichem und dem auf der Assoziationstheorie basierenden modellierten Lernverhalten neuronaler Netze wird.

## 5. Begründung unseres Untersuchungsansatzes zur Analyse und neuronalen Modellierung elementarer Lernprozesse (Experiment 1)

**Grundhypothese:** Wenn das Assoziationslernen den Lernprinzipien des Error-Back-Propagation entspricht, dann sollten sich unterschiedliche Lernverläufe für Menschen und neuronale Netze unter folgenden Bedingungen ergeben:

- wenn die Anforderungen nicht durch „reine“ Assoziationen erlernbar sind (XOR, allgemein statistisch neutrale Probleme).
- wenn Generalisierungen des Erlernten im Sinne von Transferleistungen auf bis dahin unbekannte Objekte vergleichbaren Typs gefordert sind.
- wenn hierarchische Strukturbildungen gefordert sind.
- wenn lineare Separierbarkeit nicht gegeben ist.
- wenn statistische Neutralität vorliegt.

Für die Entwicklung eines Untersuchungsparadigmas sollte gesichert werden, dass die Verhaltensanforderung sowohl für Menschen als auch für neuronale Netze kodierbar ist und zusätzlich die offenen Stellen aus dem Lachnit-Zugang einbezogen werden können. Wir entschieden uns in Anleh-

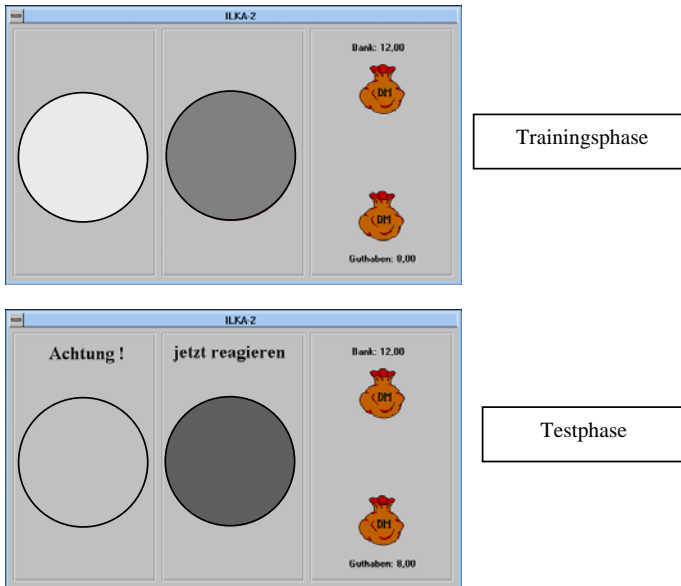


Abbildung 6:  
Experimentelles Paradigma der weiterführenden Untersuchungen

nung an die Lachnit-Untersuchungen für ein Paradigma, bei dem farbige Kreise auf einem neutralen Hintergrund dargeboten wurden. Dies zeigt der obere Teil der Abbildung 6.

Es ist deutlich, dass bei diesem Zugang mit zwei potentiellen Farbflächen insgesamt vier Ereignisse auftreten können ( $S,0$ ),  $(0,S)$ ,  $(S,S)$  und insbesondere das bei Lachnit fehlende  $(0,0)$ . Außerdem lässt sich durch die Variation der Anzahl von verwendeten Farbpaaren die Auftrittshäufigkeit der Einzelreize wie bei Lachnit verändern. Dem Auftreten eines Reizpaares folgt dann immer die entsprechende Reaktion.

Im unteren Teil der Abbildung ist dann der Typ des Prüfens des Gelernten dargestellt. Hier bleibt die automatische Reaktion aus und die Person hat mit einer Taste zu reagieren, die die zugehörige Reaktion des Systems zuzulassen oder zu vermeiden gestattet:

Der gesamte Versuchsablauf, der weiterführende Fragestellungen einschließt, ist in folgender Abbildung dargestellt:

Phase	Kennzeichnung des Versuchsablaufs	Bedeutungszuordnung
1.	<b>Trainingsphase (32 Stimuli)</b> , davon - 16 Darbietungen von Einzelreizen, d.h. ein Farbkreis links oder rechts: $(0, S_1)$ oder $(S_1, 0)$ - 16 Darbietungen von Doppelreizen, d.h. der beiden unterschiedlichen Farbkreise eines Reizpaares: $(S_1, S_2)$ bzw. $(S_2, S_1)$ .	Dies entspricht exakt der Gestaltung der Lernphase bei Lachnit (vgl. Text)
2.	<b>Vorbereitung der Testphase</b> durch die Darbietung der Reize eines neuen Reizpaares a) als neues Paar von Farbkreisen $(S_1, S_2)$ im Falle des negative patterning b) als neue Einzelreize $(S_1, 0)$ und $(0, S_2)$ im Falle des positive patterning	Dies entspricht exakt der Vorbereitung der Testphase bei Lachnit
3.	<b>Testphase</b> , Darbietung von a) einem Einzelreiz aus dem neuen Reizpaar (bei negative patterning) b) dem neuen Reizpaar (bei positive patterning)	Dies entspricht exakt der Gestaltung der Testphase bei Lachnit
4.	<b>Erweiterte Testphase</b> (für das negative patterning) <b>Item 1:</b> $(S_1, 0)$ <b>Item 2:</b> $(0, 0)$ <b>Item 3:</b> $(S_1, 0)$ <b>Item 4:</b> $(0, 0)$ <b>Item 5:</b> $(S_2, S_2)$ <b>Item 6:</b> $(S_1, S_1)$	dies entspricht dem Testitem bei Lachnit  Testitem zur Prüfung der Generalisierung auf XOR Replikation Item 1 zur Konsistenzprüfung  Replikation Item 2 der Generalisierung auf XOR  neuer Itemtyp für Transferleistungen  neuer Itemtyp für Transferleistungen

Abbildung 7:  
Versuchsablauf

Wir wollen hier die Befunde für unsere Probanden beim negative patterning darstellen:

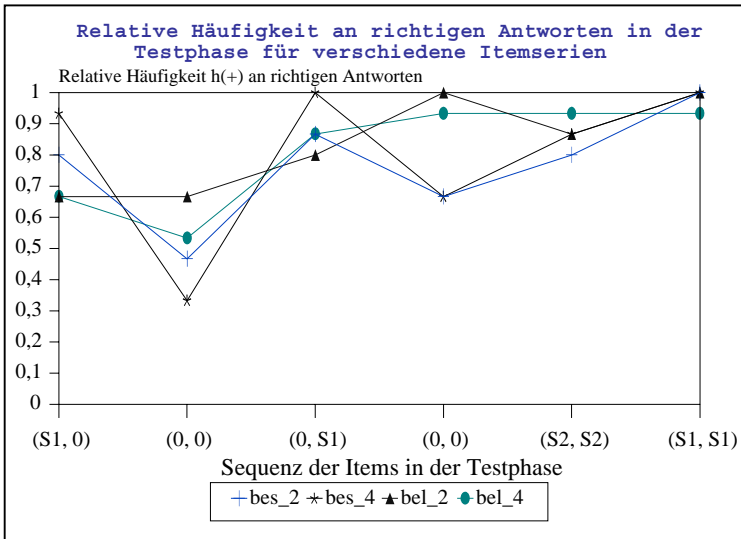


Abbildung 8:

Darstellung der relativen Häufigkeit richtiger Antworten gemäß der XOR-Funktion in der Testphase (Die Versuchsbedingungen der vier Kurven unterscheiden sich im event bestrafen/belohnen und der Manifestationshäufigkeit durch Verwendung von 2/4 Farbpaaren bei gleicher Länge der Lernserie).

Schrittweise lassen sich diese Befunde wie folgt interpretieren:

1. Es ist überraschend, dass sich das Verhalten unserer Versuchspersonen bei der ersten Darbietung des (0,0)-Items nicht von einem Zufallsverhalten unterscheidet ( $p > 0,05$ ). Dies bedeutet, dass wir hier keinerlei Effekt eines Regellernens finden, die Verhaltensvariation ist maximal. Deshalb folgern wir:
2. Lernen nach den eingeschränkten Paradigmen des Konditionierens im Sinne des positive und negative patterning sind weder durch das Lernen logischer Funktionen noch durch Regelerwerb zu erklären. Deshalb ist Lachnits Annahme einer Korrespondenz zwischen beiden widerlegt.
3. Aus der Verhaltensvariation folgt außerdem, dass unsere Probanden zumindest keine einheitliche Entscheidungsstruktur erworben haben. Aber was dann?

Eine erste Antwort finden wir, wenn wir die erste Replikation des (0,0)-Items betrachten, das an vierter Position in der Testphase dargeboten wird. Jetzt wird das Antwortverhalten sehr viel einheitlicher und unterscheidet sich deutlich vom Zufallsraten ( $p < 0,05$ ). Dies bedeutet, dass unsere Versuchspersonen in der Lage sind, ihr Entscheidungsverhalten sehr effektiv zu verändern und nach der ersten Rückmeldung zu guten Prädiktionen kommen. Und dies spricht für einen Wissenstransfer und ist somit auch ein Hinweis dafür, dass in der Lernphase sehr wohl Generalisierungen erworben wurden. In Krause & Gauger (1997) haben wir diesen Befund dadurch erklärt, dass weder eine logische Verknüpfung noch eine Regel erworben und transferiert wurde, sondern hier eine Merkmalsabstraktion vorliegt, die z.B. in der Ausbildung des komplexen Merkmals „farbig“ bestehen könnte. Dann erklärt sich der obige Befund dahingehend, dass nur beim ersten Auftreten des (0,0)-Items (zufällig) zu prüfen ist, ob die beide grauen Kreise nun als „farbig“ anzusehen sind oder nicht.

**Fazit 3:**

Das Lernen ist nicht durch die Manifestationshäufigkeiten bestimmt, es wird offensichtlich nicht die adäquate Regel erworben. Die Befunde sprechen jedoch für eine Merkmalsabstraktion und bestätigen damals die alte Köhlersche Erkenntnis der Dimensionsgeneralisierung.

**6. Leistungsvergleich von Mensch und neuronalem Netz (Experiment 2)**

Für eine zweite Betrachtung, bei der wir menschliche Lernleistungen mit dem Lernverhalten neuronaler Netze vergleichen wollten, haben wir uns im experimentellen Paradigma auf den Erwerb logischer Funktionen beschränkt. Dabei ging es uns in Anlehnung an die offenen Stellen in den obigen Konditionierungsparadigmen darum, Paare logischer Funktionen zum Lerngegenstand zu machen, die sich nur an einer Stelle im Wert unterscheiden und ansonsten identisch sind. Außerdem sollten die Funktionen unterschiedliche Schwierigkeitsanforderungen realisieren. Dafür wählten wir folgende Funktionenpaare aus:

Wahrheitswerte ausgewählter Funktionspaare			
1. AND und ÄQUI			
AND		ÄQUI	
$A \backslash B$	0	1	
0	0	0	
1	0	1	
$A \backslash B$	0	1	
0	1	0	
1	0	1	
2. NAND und XOR			
NAND		XOR	
$A \backslash B$	0	1	
0	1	1	
1	1	0	
$A \backslash B$	0	1	
0	0	1	
1	1	0	

Abbildung 9:

Tabellen der Wahrheitswerte der beiden ausgewählten Funktionenpaare

Nach den klassischen, strukturorientierten Ansätzen der Schwierigkeitskennzeichnung (Neisser & Weene, 1962; Bourne, 1972) ist die jeweils links stehende Funktion die leichtere, ÄQUI und XOR also die beiden schwierigeren. Dieser Schwierigkeitsunterschied sollte sich auf den Lernaufwand auswirken.

Wichtig nun zum Versuchsablauf: In der Lernphase wurden nur Elemente des übereinstimmenden Bereiches gewählt und danach (Erreichen eines Lernkriteriums von 6 aufeinander folgenden richtigen Prädiktionen) wurde dasjenige Objekt dargeboten, in dem sich die beiden potentiell erlernbaren Funktionen unterschieden. Dies galt auch für die neuronalen Netze, bei denen wir (Krause, Gauger, 1997, Gauger 2001) zweischichtige Netze mit folgender Struktur verwendeten:

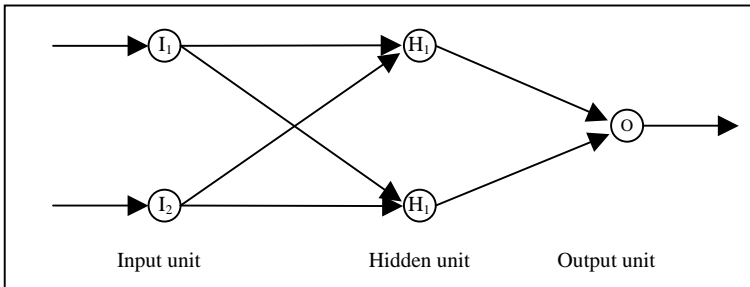


Abbildung 10:  
Netzwerkstruktur der verwendeten neuronalen Netze

Interessant nun der Befundvergleich für die Lernprozesse bei Mensch und neuronalem Netz:

	Lerninhalt		Lernaufwand (mittlere Anzahl der benötigten Items)	
	Versuchspersonen	Netzwerke	Versuchspersonen	Netzwerke
AND	14	28	11,5	26,4
ÄQUI	14	0	10,2	–
NAND	4	28	9,7	56,5
XOR	24	0	11,4	–

Abbildung 11:  
Vergleich der Lerneleistungen von Mensch und neuronalem Netzwerk

Die Ergebnisse zeigen die deutlichen Unterschiede im Lernverlauf zwischen unseren und den zufällig initialisierten neuronalen Netzen:

- Bei der *Betrachtung der Lerninhalte* kann der erwartete deutliche Unterschied der Schwierigkeit bei unseren nicht ausgewiesen werden. Im Gegenteil zeigt der Befund, dass im ersten Paar keinerlei Wirkung der Schwierigkeit auftritt und sich auch die Lernaufwände nicht unterscheiden. Überraschend wird im zweiten Paar sogar die schwierigere Verknüpfung dominant erlernt und benötigt den höheren Lernaufwand. (Es sei vermerkt, dass hier in der Literatur widersprüchliche Befunde vorliegen, die bisher nicht stimmig erklärt sind.)
- Im *Vergleich der gleichschweren Verknüpfungen* in den beiden Paaren



wird deutlich, dass *der Lernaufwand* für unsere Probanden nicht gleich ist. Damit kann hier nicht bestätigt werden, dass die Negationen eines Merkmals keinen Einfluss auf die Schwierigkeit haben. (Wir vermerken, dass bei diesem Vergleich zwischen den Paaren kein Effekt von linearer Separierbarkeit und statistischer Neutralität auftreten kann, da in diesen Vergleichen diese Eigenschaften jeweils identisch sind.)

- c. Bei den neuronalen Netzen zeigt sich im Gegensatz zudem der erwartete deutliche *Effekt der Schwierigkeit*: Alle Netzwerke erlernen durchgängig die jeweils leichteren Verknüpfungen eines Paares. Diese Diskrepanz zwischen Versuchspersonen und Netzwerken ist ein erster Hinweis darauf, dass mit den assoziativen Lernprinzipien der Netzwerke der menschliche Lernprozess nicht adäquat nachgebildet wird. Dies ist damit ein Befund, der die früher ausgewiesenen Literaturbefunde erhärtet, die Diskriminationslernen nicht ausschließlich auf das Assoziationslernen zurückführen.
- d. Beim Vergleich der jeweils durch die Netzwerke erlernten leichteren Verknüpfung wird außerdem deutlich, dass sich die *Lernaufwände der als gleich schwierig* postulierten Verknüpfungen AND und NAND deutlich voneinander unterscheiden. Dies belegt, dass auch bei den neuronalen Lernprozessen die Negation nicht schwierigkeitsneutral ist, sondern einen erhöhten Lernaufwand hervorruft.

#### **Fazit 4:**

Ausgehend von der Übereinstimmung des rein assoziativen Lernens mit dem Lernverhalten unserer neuronalen Netze wird deutlich, dass sich bei diesen Lernaufgaben das menschliche Lernverhalten vom rein assoziativen Lernen unterscheidet und damit Generalisierungen (Abstraktionsleistungen) im menschlichen Lernen auftreten, die für diese Unterschiede verantwortlich zu machen sind. Was aber macht dann diesen Unterschied aus und unter welchen Bedingungen treten sie auf?

### **7. Komponenten unterschiedlichen Lernverhaltens (Experiment 3)**

Für einen ersten Zugang zur Beantwortung dieser Frage folgern wir im Sinne eines kognitiven Ökonomieprinzips, dass wachsende Schwierigkeiten das Verlassen des reinen Assoziationsprinzips verursachen und damit den Übergang zum Abstraktionsprinzip bahnen. Damit entsteht die Frage nach möglichen Komponenten von Schwierigkeit. Die bisherigen zeigen, dass die in der Psychologie ausgewiesenen Schwierigkeitsmaße nur bedingt geeignet sind, Anhaltspunkte für die Einbeziehung von Abstraktionsleistungen zu kenn-

zeichnen. Folgen wir daher der zweiten Linie, die sich aus den Modellierungsansätzen mit neuronalen Netzen ergibt. Dort sind es vor allem zwei Eigenschaften, die Schwierigkeiten ausmachen und zu Grenzpunkten rein assoziativen Lernens führen könnten:

- das *Prinzip der linearen Separierbarkeit* der Objekte im Merkmalsraum. Hieraus resultierte die Nichtlösbarkeit des XOR-Problems mit einfachen Netzen des R-W-Typs (Minsky&Papert, 1962).
- das *Prinzip der statistischen Neutralität*, für das Thornton (1996) nachwies, dass es durch Backpropagation nicht lösbar sei.

Um den Einfluss von statistischer Neutralität und linearer Separierbarkeit auf den Lernprozess zu differenzieren, wählten wir einen Untersuchungsansatz, der von Hahn in einem Forschungsprojekt (2000) ausgeführt wurde. Gewählt wurde ein dreidimensionaler Merkmalsraum, aus dem in unterschiedlichen Versuchsserien Lernmengen ausgewählt wurden, die sich in der Variation dieser beiden Eigenschaften unterschieden. Dabei konnte Gauger (2001) beweisen, dass vollständige Variation der Bedingungen in diesem Merkmalsraum objektiv nicht möglich war. Für unseren experimentellen Zugang wählten wir folgenden Versuchsplan, der die drei im Dreidimensionalen möglichen Aufgabentypen enthält:

<b>Statistische Neutralität</b>	<b>vorhanden</b>	<b>nicht vorhanden</b>
<b>Lineare Separierbarkeit</b>		
<b>vorhanden</b>	gibt es nicht	Typ (LS und ~SN)
<b>nicht vorhanden</b>	Typ (~LS und SN)	Typ (~LS und ~SN)

Diese Aufgabentypen wurden in ein Diagnoseparadigma eingebettet, wie dies von Gluck und Bower (1988) und Estes u. Mitarbeitern (1989) zur Untersuchung des Lernverhaltens eingeführt wurde. Unsere drei dichotomen Dimensionen waren das Auftreten von drei Stoffen im Blut, die für die Behandlungsbedürftigkeit bedeutsam waren. Die Probanden wurden in der Lernserie über die Nachweisreaktionen dieser drei Stoffe wie folgt informiert:

<b>Test auf</b>	<b>Ergebnis</b>
Thetacepine	Positiv
F-Kreafonin	Positiv
Mylositon	Negativ

und hatten danach die Behandlungsbedürftigkeit (ja/nein) einzuschätzen. Die Diskrimination galt es erkannt, wenn sechs aufeinander folgende Probanden richtig klassifiziert wurden. Die Anzahl benötigter Lernschritte konnte somit als Kennwert des Lernaufwands verwendet werden. Den experimentellen Befund zeigt folgende Tabelle:

Problemtyp	Anzahl von Probanden mit Zielerreichung	Mittlere Schrittzahl bis zur Zielerreichung	Standardabweichung der Schrittzahl
LS und ~SN	20 (100%)	21,5	10,09
~LS und ~SN	19 (95%)	31,37	16,94
~LS und SN	16 (80%)	25,5	14,89

Zur Beurteilung der Befunde dieses unvollständigen Versuchsplans ist es hier zweckmäßig, die faktoriellen Einflüsse getrennt zu beurteilen. Dann ergeben sich durch Zusammenfassung bzgl. der mittleren Schrittzahl folgende (gepoolte) faktorielle Effekte:

<p>Für den <b>Untersuchungsfaktor LS:</b>  bei LS: 21,5 Schritte;  bei ~LS: 28,69 Schritte,</p> <p>Für den <b>Untersuchungsfaktor SN:</b>  bei SN: 25,5 Schritte;  bei ~SN: 26,31 Schritte.</p>
---

Für die zufallskritische Beurteilung der Mittelwerte ergeben sich keine signifikanten Unterschiede.

Der Befund belegt, dass

- unsere Probanden in hohem Grade in der Lage waren, auch diejenigen Probleme zu lösen, die backpropagation Netzwerke nicht lösen können, und
- die Eigenschaft der linearen Separierbarkeit einen stärkeren Einfluss auf den Lernaufwand unserer Probanden hat als die statistische Neutralität.

Um diesen Befund im Vergleich zum Lernverhalten der Netzwerke weiter zu differenzieren, führte Gauger (2001) eine Reihe von Simulationen durch, bei denen den Netzwerken in der Testphase ein Objekt dargeboten wurde, das in der Lernphase nicht auftrat. Dies begründet die Möglichkeit, die Generali-

sierungsfähigkeit des neuronalen Lernens in diesen Aufgabenklassen zu beurteilen. Dabei wurden zweischichtige Netzwerke verwendet, von denen nachgewiesen ist, dass sie auch XOR-Probleme lösen können.

Die Hauptbefunde nach Gauger (2001, S. 95) für die Simulation mit neuronalen Netzen sind:

- Probleme des Typs (LS, ~SN) werden nach dem Einlernen vollständig richtig generalisiert, alle 20 zufällig initialisierten Netze (also 100%) lösen die acht Transferaufgaben dieses Problemtyps.
- Probleme des Typs (~LS, ~SN) werden nach dem Einlernen fast völlig falsch generalisiert, nur in 10 von 160 Fällen (das sind 6% der Fälle) gelingt eine korrekte Generalisierung.
- Probleme des Typs (~LS, SN) werden nach dem Einlernen völlig falsch generalisiert (d.h. zu 100 %).

Die Befunde stimmen mit den Beweisen der Nichterlernbarkeit von linear nicht separierbaren und statistisch neutralen Problemen überein.

## 8. Zusammenfassende Interpretation und Ausblick

Aus der dargestellten Befundlage und weiteren Literaturbefunden lassen sich folgende Schlussfolgerungen begründen:

1. Die Erklärung assoziativer Lernprozesse mit dem Modell von Rescorla-Wagner und dessen Äquivalenz zu neuronalen Netzen mit dem Backpropagation Lernmechanismus gestattet es, in diesem Kontext Grenzen des assoziativen Lernens aufzuzeigen. Dies sind z.B. die Eigenschaften der linearen Separierbarkeit und der statistischen Neutralität, die demnach nicht über rein assoziatives Lernen ausgebildet werden können.
2. Im Gegensatz zu den neuronalen Netzen sind unsere Probanden in der Lage, auch Probleme mit diesen Eigenschaften zu lösen und für weitere Verhaltensanforderungen zu generalisieren. Dies ist ein deutlicher Beleg dafür, dass selbst in den einfachen Anforderungen des Erkennens von Merkmalsstrukturen neben den Assoziationsprozessen auch Strukturerkennungen im Sinne von Abstraktionsleistungen auftreten, die die beschriebenen neuronale Netze nicht leisten können.
3. Offen bleibt, wie bei Köhlers Haushuhn, die Frage, wann und warum im Lernen der reine Assoziationsprozess durch einen Abstraktionsprozess ergänzt wird. Eine erste Antwort bestünde darin, dass dies dann geschieht, wenn die Komplexität/Schwierigkeit der Anforderung zu groß wird, d.h. eine merkmalsbezogene Entscheidungsstruktur nicht erkennbar oder in der verfügbaren Gedächtnisspanne nicht erfassbar ist.

4. Bzgl. der merkmalsbezogenen Entscheidungsstruktur bietet Gauger (2001) einen Erklärungsansatz an, den er als Rekonstruktion der Objektmenge bezeichnet und der darauf basiert, an Hand der bedingten Wahrscheinlichkeiten der Input-/Output-Zuordnungen die Objektzuordnungen vollständig zu rekonstruieren. Ist dies der Fall, dann lernen auch die neuronalen Netze und die Anzahl der nicht rekonstruierbaren Objekte erhöht die Schwierigkeit. Da Wahrscheinlichkeiten mit Manifestationshäufigkeiten und damit Assoziationen verbunden sind, könnte der Grad der Abweichung einen Indikator für die Grenzen der Lernfähigkeit der neuronalen Netze und des Wechsels auf das Abstraktionsprinzip unserer Probanden begründen.
5. Bzgl. der verfügbaren Gedächtnisspanne kommt ein zweiter Gedanke zum Tragen, den Thornton mit seiner Typisierung von „justified problems“ anspricht. Er kennzeichnet sie in den Input-/Outputbeziehungen wie folgt:
  - a)  $P(O) = p$ .
  - b)  $P(O|X) = p$
  - c)  $P(O|g(X)) = p$

Dabei sind die „direkten“ Probleme a) und b) nach dem Assoziationsprinzip beherrschbar, während der Typ c im Wesentlichen dadurch bestimmt wird, dass die Funktion  $g$  erkannt werden muss. So ist z.B. auch von uns gefundenes unterschiedliches Lernverhalten der XOR-Funktion dadurch erklärbar, dass die schwer erkennbare Zuordnung ganz einfach wird, wenn man die Anzahl auftretender Merkmale als diskriminationsrelevant erkennt. Die Funktion  $\text{anz}(X) = \sum x_i$  hat beim XOR den Wert 1 und ist bei den Parity-Problemen ungerade. Damit entsteht eine Entscheidungsstruktur, die nicht nur Merkmale miteinander in Beziehung setzt, sondern auf die Anzahl generalisiert. Das Erkennen von Merkmalsstrukturen wird damit auf die Ebene des hierarchisch-genetischen Prinzips von Duncker (1932) gehoben, nämlich der Umformung des Gegebenen bis die Lösung einsichtig wird.

### **Fazit 5:**

Ausgehend von der Existenz unterschiedlicher Lernprinzipien Assoziation vs. Abstraktion wird menschliches Lernen entscheidend von der erfahrungsabhängigen, internen Repräsentation der Aufgabenstellung bestimmt. Für direkte Probleme, die über die bedingten Wahrscheinlichkeiten rekonstruierbar sind, ist das (im Rahmen der Gedächtnisspanne) die Merkmalsrepräsentation

mit dem Assoziationsprinzip, bei den mittelbaren Problemen wird die Entscheidungsstruktur hierarchische Merkmalsverknüpfungen erfordern, die dann Abstraktionsprozesse und Repräsentationswechsel enthalten. In diesem Fall gehen die interindividuellen Unterschiede hinsichtlich stationärer und prozeduraler Gedächtnisbestände entscheidend in den Lernprozess ein.

**Ausblick:**

Dieses Fazit begründet die Frage, wie man an dieser Stelle zu weiteren Erkenntnissen kommen kann. Wir wollen dabei weiterhin den Zugang des Vergleichs von menschlichem Lernverhalten mit künstlichen Modellen verfolgen und glauben, dass die Entwicklungen der Multi-Agenten-Systeme (MAS) ein geeigneter Zugang sein kann, um den Effekt unterschiedlicher Annahmen über den Lernprozess zu prüfen. Ein solches MAS formuliert z.B. ein Modell, in dem unterschiedliche Agenten in jeweils einer Merkmalsdimension agieren und versuchen, im Rahmen ihres vorgegeben Verhaltensrepertoires eine Stellung zu beziehen, die im Kontext mit den anderen Agenten eine Problemlösung darstellt. Für „constraint satisfaction“ Probleme (CSP) kennzeichnen Liu, Jing & Tang (2002) dies u.a. am Beispiel des n-Damen Problems (n Damen sind auf einem  $n \times n$  Schachbrett korrekt zu positionieren wie folgt:

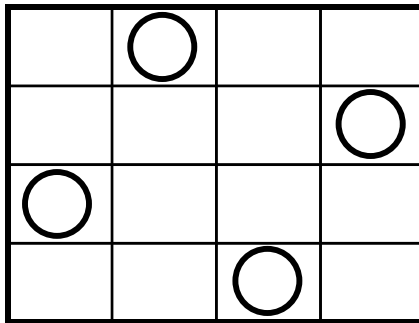


Abbildung 12:  
Schematische Darstellung einer Lösung des 4-Damen-Problems

Ein Multi-Agenten System kann dann wie folgt gekennzeichnet werden:

*Ein Agent ist eine virtuelle Einheit* mit folgenden Eigenschaften:

- der Fähigkeit, in der Umgebung zu leben und zu handeln,
- der Fähigkeit, die lokale Umgebung wahrzunehmen,
- be driven by certain objectives
- einem gewissen Reaktionspotential.

*Ein Multi-Agenten-System* enthält folgende Elemente:

- a. *Eine Umgebung E*, in dem die Agenten leben,
- b. *Eine Menge R von Reaktionsregeln*, die die Interaktionen zwischen den Agenten und der Umwelt betreffen,
- c. *Eine Menge A = {a<sub>1</sub>, ..., a<sub>n</sub>}* von Agenten.

Für das n-Damen-Problem induziert dies ein Multi-Agenten-System mit dem Grundgedanken:

- Jeder Agent repräsentiert eine Variable und die Position auf der Domäne.
- Jeder Agent kann sich lokal innerhalb seiner Domäne bewegen und hat dafür ein eigenes Bewegungspotential.
- Jeder Agent strebt eine Position an, bei der seine Verstöße gegen die Nebenbedingungen 0 wird.
- die Anzahl der Verstöße wird zu jedem Zeitpunkt festgestellt.

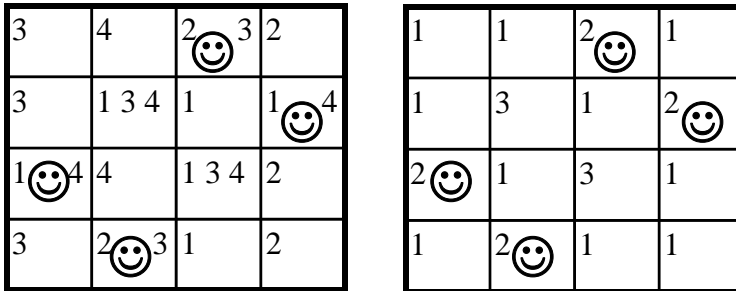


Abbildung 13:  
 Darstellung zur Funktion eines Multi-Agenten-Systems:  
 Links: Kennzeichnung der bedrohten Felder durch die Damen in der i-ten Zeile (i) bzgl. der anderen Zeilen  
 Rechts: Summe der Verstöße (violations) aus der Stellung der 4 Agenten. Ein Zielzustand ist dadurch gekennzeichnet, dass die Felder der Agenten 0 Verstöße aufweisen.

Für unsere weiterführenden Überlegungen ist wichtig, dass diese MAS es gestatten, den Agenten präzise Verhaltensmöglichkeiten aufzuerlegen und auch das Informationsangebot für die Verhaltenssteuerung vorzugeben. Damit sollte es möglich werden, den Einfluss oben beschriebener Eigenschaften und Effekte auf das Lernverhalten zu simulieren und mit menschlichem Lernverhalten zu vergleichen.

Damit sollten Agenten-Systeme auch den Übergang zu hierarchisch organisierten Merkmalsstrukturen ermöglichen, und damit einen neuen Zugang zur Modellierung kognitiver Prozesse, insbesondere individueller Lernstrategien, eröffnen.

## Literatur

- Aristoteles (1986). Aristoteles Werke in deutscher Übersetzung (Original etwa 330 v. Chr.). Berlin: Akademie-Verlag.
- Bourne, L.E. (1974). An inference model for conceptual rule learning. In: Solso, R.L. (Ed.), *Theories in cognitive psychology: The Loyola Symposium*, 231–255. Potomac, Maryland: L. Erlbaum Ass.
- Cheng, P.W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, 104, 367–405.
- Clark, A. & Thornton, C. (1993). Trading spaces: Computation, representation, and the limits of learning. *Cognitive and Computation Sciences, CSRP* 291.
- Clark, A. & Thornton, C. (1997). Trading spaces: Computation, representation, and the limits of uniformed learning. *Behavioral and Brain Sciences*, 20, 57–90.
- Estes, W.K. u.a. (1989). Base rate effects in category learning: A comparison of parallel network and memory storage-retrieval models. *Journal of Exp. Psychology; Learning, Memory and Cognition*, 15, 556–571.
- Gauger, U. (2001). Möglichkeiten und Grenzen der Modellierung von Lernprozessen mit neuronalen Netzen. Dissertationsschrift. Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II.
- Gluck, M. & Bower, G.H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, 117, 227–247.
- Jenkins, H. & Ward, W. (1965). Judgement of contingency between responses and outcomes. *Psychological Monographs*, 7, 1–17.
- Klix, F. (1971). *Information und Verhalten*. Berlin: VEB Deutscher Verlag der Wissenschaften.
- Klix, F. (1980). *Erwachendes Denken*. Berlin: VEB Deutscher Verlag der Wissenschaften.
- Klix, F. (Hrsg.) (1984). *Gedächtnis, Wissen, Wissensnutzung*. Berlin: VEB Deutscher Verlag der Wissenschaften.
- Köhler, W. (1918). Nachweis einfacher Strukturfunktionen beim Schimpansen und beim Haushuhn. Über eine neue Methode zur Untersuchung des bunten Farbensystems. *Abhandlungen der Königlich Preussischen Akademie der Wissenschaften, Physikalisch-Mathematische Klasse*.
- Krause, B. (1997). Towards a theory of cognitive learning. In: Brzesinski, J., Krause, B. & Maruszewski, T. (Eds). *Idealization in psychology*. Amsterdam: Rodopi.
- Krause, B. (2000). Entwicklungstendenzen mathematischer Methoden in der psychologischen Forschung. *Zeitschrift für Psychologie*, 208, 357–384.
- Krause, B. (2001). Methoden der Beobachtung und Modellierung elementarer Lern- und Gedächtnisprozesse. In: Krause, B. (Hrsg). *Ebbinghaus-Symposium. ZeE-Publikationen. Reihe: Wissenschaftliche Veranstaltungen*. Berlin: ZeE



- Krause, B. & Gauger, U. (1997). Learning and use of invariances: Experiments and network simulations. In: Brzesinski, J., Krause, B. & Maruszewski, T. (Eds). *Idealization in psychology*. Amsterdam: Rodopi.
- Kruschke, J.K. (1992). ALCOVE: An exemplar-based connectionist model for category learning. *Psychological Review*, 99, 22–44.
- Kruschke, J.K. (1993). Human category learning: Implications for backpropagation models. *Connection Science*, 5, 3–36.
- Liu, J. & Jing, H. & Tang, Y.Y. (2002). Multi-agent oriented constraint satisfaction. *Artificial Intelligence*, 136, 101–144.
- Minsky, M. & Papert, S. (1969). *Perceptrons*. Cambridge, MA: MIT Press.
- Neisser, U. & Weene, P. (1962). Hierarchies in concept attainment. *Journal Exp. Psychology*, 64, 640–645.
- Pavlov, I.P. (1927). *Conditioned reflexes*. New York: Dover.
- Pearce, J.M. (1997). *Animal learning and cognition*. Hove, East Sussex: Psychology Press.
- Rescorla, R. A. (1968). Pavlovian conditioning: It's not what you think it is. *American Psychologist*, 43, 151–160.
- 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*, pp. 216–271. New York: Appleton-Century-Crofts.
- Salmon, W. C. (1965). The status of prior probabilities in statistical explanation. *Philosophy of Science*, 32, 137–146.
- Spence, K.W. (1936). The nature of discrimination learning in animals. *Psychological Review*, 43, 427–449.
- Spence, K.W. (1937). The differential response to stimuli varying within a single dimension. *Psychological Review*, 44, 430–444.
- Thornton, C. (1996a). Backpropagation can't do parity generalization. *Cognitive and Computation Sciences, CSRP* 450.
- Thornton, C. (1996b). Parity: the problem won't go away. In: McCalla, G. (Ed.). *Proceedings of AI-96*, 362–374. Toronto: Canada.
- Tolman, E.L. (1911). *Animal intelligence: Experimental studies*. New York: Teacher's College.