

## Teil I

# Grundlagen auf der Seite des Menschen



# 1 Grundmodell menschlicher Informationsverarbeitung

Wir Menschen verarbeiten Informationen auf vielfältige Weise. Die Kognitionspsychologie betrachtet den Menschen als informationsverarbeitendes System. Grundlage der Wahrnehmung und damit die Informationsquelle sind Reize, die wir über unsere Sinnesorgane erhalten und die zu einem Gesamteindruck zusammengeführt werden. Kognition (lat. *cognoscere* = erkennen) bezeichnet die Verarbeitung dieser Information. Die Mechanismen der **Wahrnehmung** und der **Kognition** sind daher von zentraler Bedeutung für die Mensch-Maschine-Interaktion. Computer dienen dazu, Menschen bei bestimmten Aufgaben zu unterstützen. Sie sind ein kognitives Werkzeug, ein Kommunikationswerkzeug, eine Gedächtnishilfe und vieles mehr. Damit sie diese Rolle ausfüllen können, ist für den Entwurf interaktiver Systeme Wissen über die Grundlagen menschlicher Informationsverarbeitung, deren Fähigkeiten und Grenzen, notwendig. Der kognitive Ansatz versteht Menschen als informationsverarbeitende Organismen, die wahrnehmen, denken und handeln. Da das Ziel bei der Gestaltung interaktiver Systeme zumeist darin besteht, effizient und einfach zu benutzende Systeme zu entwickeln, ist eine optimale Anpassung an die Aufgaben und Bedürfnisse menschlicher Benutzer sehr wichtig.

Abbildung 1.1 zeigt das Schema eines Mensch-Maschine-Systems. Es besteht aus Nutzungskontext, Benutzer, Aufgabe und Werkzeug. Diese Elemente stehen untereinander in Beziehung. Der Benutzer möchte mithilfe des Werkzeugs, in diesem Fall mit einem Computersystem, eine Aufgabe lösen. Die Aufgabe bestimmt das Ziel des Benutzers. Das Computersystem als Werkzeug stellt bestimmte Mittel bereit, um die Aufgabe zu lösen. Die Passgenauigkeit dieser Mittel auf die Anforderungen der Aufgabe, aber auch auf die Stärken und Schwächen des Benutzers bestimmen die Schwierigkeit der Aufgabe. Eine Änderung der Eigenschaften des Werkzeugs verändert die Aufgabe, da sich

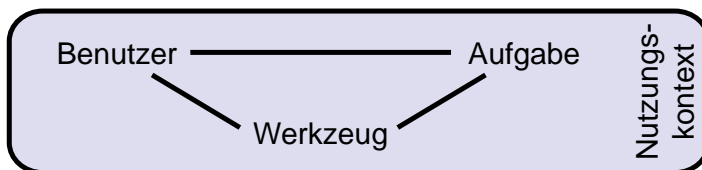


Abbildung 1.1: Schema eines Mensch-Maschine Systems: Nutzungskontext, Benutzer, Aufgabe und Werkzeug (vgl. Wandmacher [119])



mmibuch.de/s/1.1

die Schritte zum Lösen der Aufgabe verändern. Schließlich stellt der Nutzungskontext bzw. die Arbeitsumgebung einen wichtigen Einflussfaktor dar, der Auswirkungen auf die Qualität der Aufgabenbearbeitung hat.

Das primäre Ziel der Mensch-Maschine-Interaktion ist die Verbesserung des Mensch-Maschine-Systems. Die Optimierung des Gesamtsystems bedeutet, dass die Aufgabe möglichst **effizient** (geringer Aufwand) und **effektiv** (gutes Resultat) sowie mit möglichst großer subjektiver Zufriedenheit erledigt werden kann. Da es sich bei Computersystemen um intellektuelle Werkzeuge handelt, die menschliche kognitive Funktionen unterstützen, die Externalisierung von Gedächtnisinhalten (Visualisierung), die Wahrnehmung von computergenerierten Objekten, sowie die Kommunikation mit anderen Menschen erlauben, ist die Anpassung an und Unterstützung von kognitiven Funktionen von zentraler Bedeutung. Heutige Computersysteme sind allerdings nicht ausschließlich als kognitive Werkzeuge zu verstehen. Der Zweck vieler Computersysteme ist die Generierung künstlicher Wahrnehmungsobjekte, die über verschiedene Modalitäten, z.B. visuell, auditiv oder haptisch, vermittelt werden. Hier ist nicht nur die Kognition, sondern auch andere Aspekte der Wahrnehmung relevant. Ein klares Beispiel sind immersive Anwendungen, z.B. Flugsimulatoren oder Virtual Reality CAVES. Neben der Kognition und Wahrnehmung sind auch die **motorischen** Eigenschaften des Menschen von zentraler Bedeutung, da dieser ja wieder mechanisch Eingaben tätigen muss. Damit ist die Relevanz menschlicher Wahrnehmung, Informationsverarbeitung und Motorik unmittelbar klar. Das Ziel eines optimal funktionierenden Mensch-Maschine-Systems ist nur erreichbar, wenn die grundlegenden Fähigkeiten und Einschränkungen des Menschen bekannt sind und die bestmögliche Anpassung der Maschine an diese Eigenschaften vorgenommen wird.

Menschen und Computer haben unterschiedliche Stärken und Schwächen. Als Gesamtsystem sind die Eigenschaften von Menschen und Computern oft komplementär zueinander. So sind Menschen in der Lage, unscharfe Signale in der Gegenwart von Rauschen zu entdecken und komplexe Signale (z.B. Sprache) oder komplexe Konfigurationen (z.B. räumliche Szenen) zuverlässig zu erkennen. Sie können sich an unerwartete und unbekannte Situationen anpassen, und sich große Mengen zusammengehöriger Informationen merken. Ebenso sind sie aber auch in der Lage, sich bei der ungeheuren Menge an Informationen, die sie wahrnehmen, auf das Wesentliche zu konzentrieren. Umgekehrt sind Computer uns Menschen darin überlegen, algorithmisch formulierbare Probleme schnell zu lösen, bekannte Signale schnell und zuverlässig zu detektieren, große Mengen nicht zusammenhängender Daten zu speichern und Operationen beliebig oft zu wiederholen.

## 1.1 Menschliche Informationsverarbeitung und Handlungssteuerung

Abbildung 1.2 zeigt die Komponenten der menschlichen Informationsverarbeitung und Handlungssteuerung [119]: Das **Wahrnehmungssystem** verarbeitet **Reize** durch die **Sinnesorgane** und speichert diese kurzzeitig in **sensorischen Registern**. Die Zeichenerkennung schließlich führt zu einer symbolischen bzw. begrifflichen Repräsentation des Wahrnehmungsgegenstands. Das **Kurzzeitgedächtnis (KZG)** ist der Ort kon-

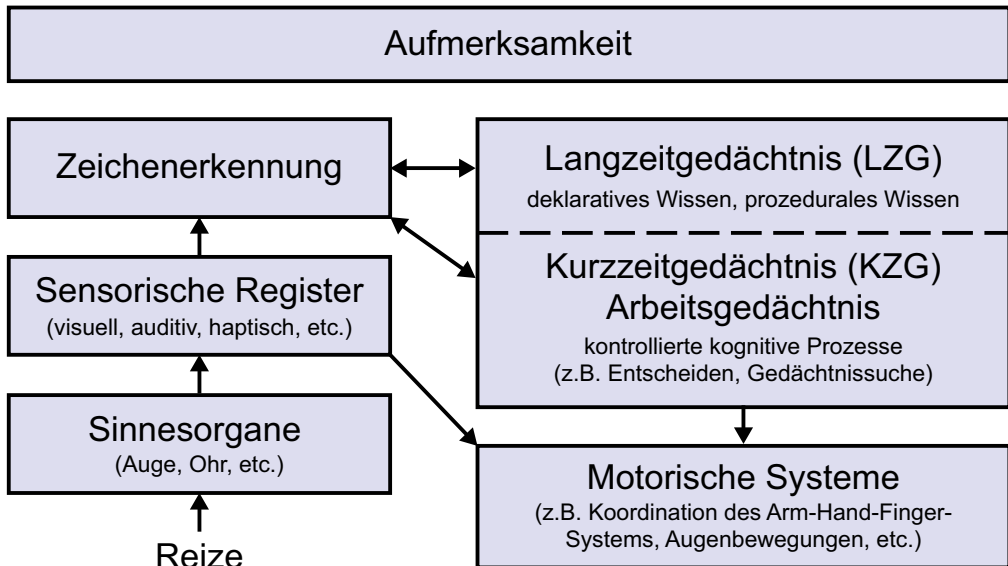


Abbildung 1.2: Komponenten menschlicher Informationsverarbeitung und Handlungssteuerung (vgl. Wandmacher [119])



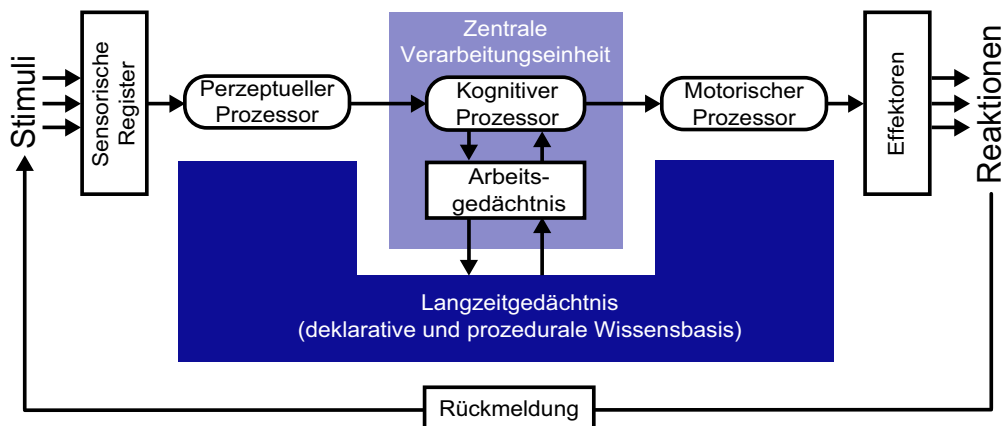
trollerter kognitiver Prozesse, wie z.B. des Entscheidens und der Gedächtnissuche. Im **Langzeitgedächtnis (LZG)** sind deklaratives und prozedurales Wissen repräsentiert. Die **motorischen** Systeme umfassen die Bewegungen des Arm-Hand-Finger-Systems, Augen- und Kopfbewegungen, sowie das Sprechen.

Die Aufmerksamkeit stellt eine übergeordnete Komponente der menschlichen Informationsverarbeitung und Handlungssteuerung dar. Sie dient der Zuweisung kognitiver Ressourcen zu bestimmten Wahrnehmungs- und Handlungsaspekten im Rahmen einer kontrollierten Verarbeitung. Die kontrollierte Verarbeitungskapazität ist begrenzt. Die für die Steuerung einer Handlung erforderliche Kapazität kann durch Üben der Handlung verringert werden. Üben führt zu sensomotorischen Fertigkeiten, die mehr oder weniger automatisiert ablaufen. Selbst bei der körperlich relativ passiven Verwendung von Computersystemen spielen **sensomotorische** Fertigkeiten, wie z.B. die Nutzung der Maus und die Bedienung der Tastatur eine große Rolle. Diese Fertigkeiten sind eine Voraussetzung für die Lösung anspruchsvoller Aufgaben, weil so die kontrollierte Verarbeitungskapazität nicht maßgeblich zur Bedienung der Benutzerschnittstelle eingesetzt, sondern zur Lösung der eigentlichen Aufgabe verwendet wird.

## 1.2 Model Human Processor

Die Vorgänge der menschlichen Kognition und Handlungssteuerung sind schon seit der Antike Gegenstand wissenschaftlicher Untersuchungen. Auch wenn das Wissen über kognitive Vorgänge und die Art und Weise wie wir mit der Umwelt interagieren stetig zugenommen hat, sind wir noch weit davon entfernt das Phänomen natürlicher **Kognition** umfassend zu verstehen. Es stellt sich aber die Frage wie menschliche Kognition und die mit ihr verbundene Informationsverarbeitung trotzdem sinnvoll modelliert und eine systematische Verwendung dieses Modells, z.B. bei der Entwicklung oder der Verbesserung von Benutzerschnittstellen, gewährleistet werden kann.

Ein sehr vereinfachtes Modell der menschlichen Informationsverarbeitung wurde 1983 von Card, Moran und Newell mit dem **Model Human Processor** vorgestellt [21]. Das Hauptaugenmerk dieses Modells liegt auf der Modellierung und Vorhersage der Zeitspannen, die bei der menschlichen Informationsverarbeitung in verschiedenen Stadien entstehen. Bewusst greifen die Autoren die Analogie zu Computersystemen auf und sprechen vom *menschlichen Prozessor*, der Information in Zyklen verarbeitet. Das Modell schätzt die Zeiten der unterschiedlichen Zyklen und durch Summierung entsteht eine Abschätzung der menschlichen Informationsverarbeitungszeit.



mmlibuch.de/a/1.3

Abbildung 1.3: Schematische Sicht der menschlichen Informationsverarbeitung (angelehnt an Eberleh und Streitz [31])

Abbildung 1.3 gibt einen Überblick zu den einzelnen Komponenten, wie von Eberleh und Streitz (in Analogie zu Card [21]) beschrieben [31]. Sie identifizieren drei Prozessoren: den **perzeptuellen Prozessor**, den **kognitiven Prozessor** und den **motorischen Prozessor**. Wie in Abschnitt 1.1 beschrieben, verwenden diese Prozesse sowohl das Kurzzeit- als auch das Langzeitgedächtnis. Sinnesreize (**Stimuli**) erreichen die Menschen zunächst durch die Sinnesorgane und werden direkt im sensorischen Register gespeichert. Die Information im sensorischen Register ist eine sehr direkte Repräsentation,

die u.a. von der Intensität des Sinnesreizes abhängt. Diese Repräsentation wird nun vom perzeptuellen Prozessor unter Verwendung des Langzeitgedächtnisses verarbeitet und das Ergebnis an die zentrale Verarbeitungseinheit und den kognitiven Prozessor weitergeleitet. Unter Berücksichtigung der Inhalte des Arbeits- und Langzeitgedächtnisses werden Handlungen geplant, die durch den motorischen Prozessor ausgeführt werden und dann zu Aktionen des motorischen Apparats (**Effektoren**) führen.

Zu den sensorischen und motorischen Systemen des Menschen gehören neben dem Sehsinn, dem Hörsinn und dem Tastsinn auch der Geruchs- und Geschmackssinn. Dazu kommen der vestibuläre Sinn (Balance und Beschleunigung), sowie die eigentlich zum Tastsinn gehörende Propriozeption (Körperbewusstsein, Kinästhetik), Thermorezeption (Temperatur) und Schmerzempfindung. Zum motorischen System zählen Arme, Hände und Finger; Kopf, Gesicht und Augen; Kiefer, Zunge; sowie Beine, Füße, Zehen. In die Mensch-Maschine-Interaktion werden immer mehr dieser sensorischen und motorischen Systeme einbezogen, beispielsweise bei Spielekonsolen, die Körperbewegungen und Gesten des Spielers erkennen.

Die Hauptmotivation für das vorliegende Modell war die Vorhersage der Bearbeitungszeiten und des Schwierigkeitsgrads von Steuerungsaufgaben in der Mensch-Maschine-Interaktion. Dazu gibt das Modell die Zeitdauern elementarer Operationen des perzeptuellen, kognitiven und motorischen Prozessors an. Mit dem Modell lassen sich die Ausführungszeiten einfacher Reiz-Reaktions-Experimente abschätzen.

Nehmen wir an, die modellierte Aufgabe bestehe darin, eine Taste zu drücken, sobald ein beliebiges Symbol auf dem Bildschirm erscheint. Das Symbol erscheint zum Zeitpunkt  $t = 0\text{s}$ . Das Symbol wird zunächst durch den Wahrnehmungsprozessor verarbeitet. Hierfür kann eine Verarbeitungszeit von  $\tau_P = 100\text{ms}$  veranschlagt werden. Anschließend übernimmt der kognitive Prozessor das Ergebnis und benötigt für die Weiterverarbeitung und die Bestimmung der Reaktion z.B. die Zeit  $\tau_C = 70\text{ms}$ . Der motorische Prozessor schließlich sorgt für das Drücken der Taste und benötigt dafür z.B. die Zeit  $\tau_M = 70\text{ms}$ .

Die Gesamtreaktionszeit  $\tau$  kann nun durch die Summe der einzelnen Zeiten  $\tau = \tau_P + \tau_C + \tau_M$  bestimmt werden. Im vorliegenden Beispiel ist damit die Reaktionszeit  $\tau = 240\text{ms}$ . Die jeweiligen Zeiten für  $\tau_P$ ,  $\tau_C$  und  $\tau_M$  können aus der Grundlagenforschung der Wahrnehmungspsychologie übernommen werden. Es handelt sich dabei aber nur um Durchschnittswerte. Konkrete Werte können von Mensch zu Mensch erhebliche Unterschiede aufweisen.



### Übungsaufgaben:

1. Abbildung 1.2 zeigt die Komponenten menschlicher Kognition. Die Leistung des Sinnesorgans Auge kann beispielsweise durch eine Brille erhöht werden. Überlegen Sie sich, wie bei jeder einzelnen Komponente Leistung bzw. Effizienz erhöht werden kann. Kennen Sie solche Techniken aus Ihrem Alltag? Erläutern Sie, wie und warum jeweils die Leistung erhöht wird.
2. Wie kann das in diesem Kapitel vorgestellte menschliche Informationsverarbeitungsmodell (Abbildung 1.3) sinnvoll bei der Gestaltung von Computerspielen eingesetzt werden? Diskutieren Sie anhand eines Computerspiels ihrer Wahl hierzu die Stimuli, die Reaktionen und die Rolle des perzeptuellen und des motorischen Prozessors, sowie der zentralen Verarbeitungseinheit für grundlegende Spielzüge. Erläutern Sie, wieso eine genaue Abschätzung der Prozessorgeschwindigkeiten wichtig für ein gutes Spielerlebnis ist.



# 2 Wahrnehmung

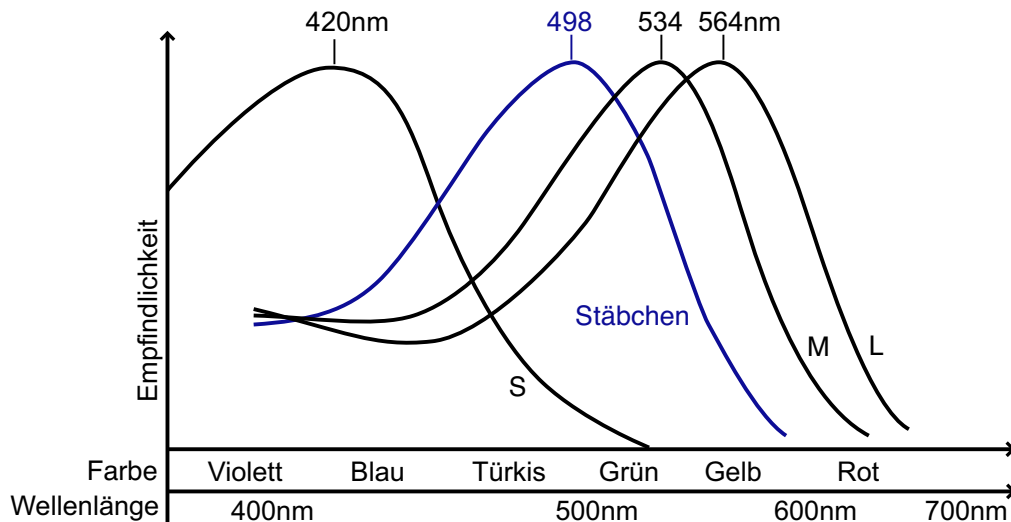
## 2.1 Sehsinn und visuelle Wahrnehmung

Der **Sehsinn** ist derzeit der am meisten genutzte Sinn bei der Interaktion zwischen Mensch und Computer. Mit den **Augen** nehmen wir Ausgaben wie Text und grafische Darstellungen auf dem Bildschirm wahr und verfolgen unsere Eingaben mit den verschiedenen Eingabegeräten, wie Tastatur oder Maus. Um grafische Ausgaben gezielt zu gestalten, benötigen wir Einiges an Grundwissen über die Physiologie des Sehannes, die **visuelle Wahrnehmung** von Farben und Strukturen, sowie einige darauf aufbauende Phänomene.

### 2.1.1 Physiologie der visuellen Wahrnehmung

Sichtbares **Licht** ist eine elektromagnetische Schwingung mit einer Wellenlänge von etwa 380nm bis 780nm. Verschiedene Frequenzen dieser Schwingung entsprechen verschiedenen **Farben** des Lichts (siehe Abbildung 2.1). Weißes Licht entsteht durch die Überlagerung verschiedener Frequenzen, also die Addition verschiedener Farben. Das eintreffende Licht wird durch eine optische Linse auf die **Netzhaut** projiziert und trifft dort auf lichtempfindliche Zellen, die sogenannten **Stäbchen** und **Zäpfchen**, die für bestimmte Spektralbereiche des sichtbaren Lichtes empfindlich sind (siehe Abbildung 2.1). Die Zäpfchen gibt es in drei Varianten, die jeweils ihre maximale Empfindlichkeit bei verschiedenen Wellenlängen (S = short, M = medium, L = long) haben (siehe Abbildung 2.1). Bei weißem Licht sprechen alle drei Zäpfchentypen gleich stark an. Wahrgenommene Farben entstehen durch verschiedenen starke Reize an den verschiedenen Zäpfchentypen. Werden beispielsweise nur L-Zäpfchen angesprochen, dann sehen wir Rot. Werden die L- und M-Zäpfchen etwa gleich stark angesprochen, S-Zäpfchen jedoch gar nicht, dann sehen wir die **Mischfarbe** Gelb. Dies kann bedeuten, dass tatsächlich gelbes Licht mit einer Frequenz zwischen Rot und Grün einfällt, oder aber eine Mischung aus rotem und grünem Licht (entsprechend einem Zweiklang in der akustischen Wahrnehmung). In beiden Fällen entsteht die gleiche Reizstärke an den verschiedenen Zäpfchen, so dass in der Wahrnehmung kein Unterschied besteht. Nur dieser Tatsache ist es zu verdanken, dass wir aus den drei Farben Rot, Grün und Blau, mit denen die drei Zäpfchenarten quasi getrennt angesprochen werden, alle anderen Farben für das menschliche Auge zusammenmischen können, ähnlich wie eine Linearkombination aus drei Basisvektoren eines Vektorraumes. Die drei Farben Rot, Grün und Blau heißen deshalb (additive) **Grundfarben** und spannen den dreidimensionalen **RGB-Farbraum** auf (für Details siehe auch Malaka et al. [79]).

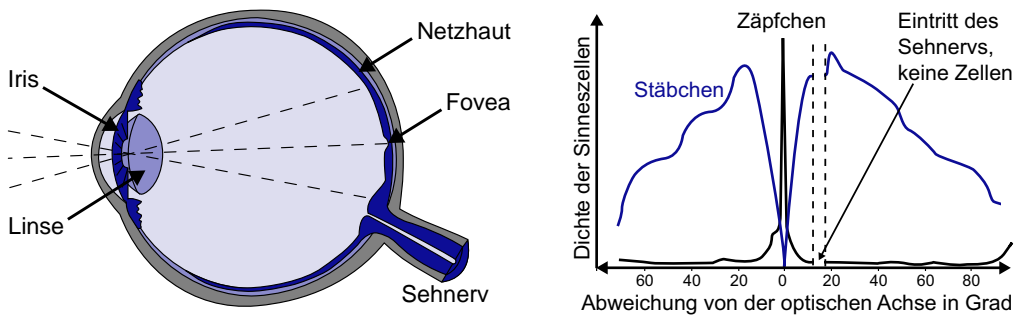
Die Stäbchen sind hauptsächlich für die Helligkeitswahrnehmung verantwortlich. Sie sind absolut gesehen wesentlich empfindlicher und kommen daher vor allem bei dunk-



**Abbildung 2.1:** Frequenzspektrum des sichtbaren Lichts und spektrale Empfindlichkeit der Sinneszellen im Auge (schematisch für S-, M-, L-Zäpfchen und Stäbchen)

leren Lichtverhältnissen zum Einsatz. Da sie alle die gleiche spektrale Empfindlichkeit besitzen, ermöglichen sie kein Farbsehen wie die Zäpfchen. So ist es zu erklären, dass wir bei schlechten Lichtverhältnissen weniger Farben unterscheiden können (*Bei Nacht sind alle Katzen grau.*).

Eine Art Blende im Auge, die **Iris**, regelt die Öffnung der optischen Linse und damit die Menge des durchfallenden Lichtes, genau wie die Blende eines Kameraobjektivs. Ohne diese sogenannte **Akkommodation** durch die Iris kann das Auge einen **Dynamikumfang** von etwa 10 fotografischen Blendenstufen, also etwa  $1 : 2^{10}$  oder  $1 : 1.000$  wahrnehmen, mit Akkommodation sogar 20 Stufen oder  $1 : 1.000.000$ . Dabei beträgt die **Helligkeitsauflösung** des Auges 60 Helligkeits- oder Graustufen. So ist es zu erklären, dass bei einem Monitor mit einem darstellbaren **Kontrastumfang** von  $1 : 1.000$  die üblichen 8 Bit (= 256 Stufen) pro Farbkanal ausreichen, um stufenlose Farbverläufe darzustellen. Die Fähigkeit, konkrete Farben wiederzuerkennen, ist jedoch viel geringer und beschränkt sich in der Regel auf etwa 4 verschiedene Helligkeitsstufen je Farbkanal. Hierbei ist zu beachten, dass das menschliche **Farbsehen** im Spektralbereich zwischen Rot und Grün wesentlich besser ausgeprägt ist, als im Blaubereich, da das Auge etwa 5x mehr Sinneszellen für Rot und Grün enthält, als für Blau. Eine plausible evolutionäre Erklärung hierfür ist, dass in der Natur diese Farben sehr häufig vorkommen und insbesondere bei der Nahrungssuche (Früchte vor Blattwerk) wichtig waren. In der Kunst hat es dazu geführt, dass Gelb als weitere **Primärfarbe** neben Rot, Grün und Blau angesehen wird. Für grafische Darstellungen bedeutet es, dass Farben im Bereich



**Abbildung 2.2:** Links: Vereinfachtes Schnittbild des menschlichen Auges, Rechts: Verteilung der Sinneszellen in der Netzhaut in Abhängigkeit vom Winkel zur optischen Achse



von Rot über Gelb bis Grün mit weniger Anstrengung und detaillierter wahrgenommen werden können als im Bereich von Blau bis Violett. Dunkelblaue Schrift auf hellblauem Grund ist also beispielsweise die schlechteste Kombination bzgl. Lesbarkeit (niedriger Kontrast, wenige Sinneszellen) und wird daher sogar oft für Beschriftungen verwendet, die gar nicht gelesen werden sollen, wie z.B. Inhaltsangaben auf Lebensmitteln.

Die Lichtsinneszellen sind auf der Netzhaut nicht gleichmäßig verteilt. Stattdessen befinden sich in der Mitte, der sogenannten **Fovea** sehr viele Zäpfchen und im Außenbereich zunächst viele, dann immer weniger Stäbchen (siehe Abbildung 2.2 rechts). Dies bedeutet, dass wir nur im Innenbereich wirklich scharf und farbig sehen, im Außenbereich zunehmend unschärfer und nur hell/dunkel. Trotzdem ist der Außenbereich und somit unsere **periphere Wahrnehmung** empfindlich für andere Dinge, wie beispielsweise Bewegungen.

Um trotzdem ein detailliertes Bild unserer gesamten Umgebung zu liefern, ist das menschliche Auge in ständiger Bewegung. In sogenannten **Sakkaden** von etwa 200ms Dauer springt das Auge ruckartig zu verschiedenen Punkten im Sichtfeld. Beim Lesen beispielsweise tasten wir so die Textzeilen ab und rekonstruieren ein detailliertes Bild der Buchstaben und Wörter. Umgekehrt bedeutet dies aber auch, dass unsere **visuelle Aufmerksamkeit** nicht in allen Bereichen des Sichtfeldes gleichzeitig sein kann. Wird sie z.B. durch Blinken oder andere Störungen zu einer bestimmten Stelle gelenkt, so nehmen wir gleichzeitige Veränderungen an einer anderen Stelle nicht wahr. Dieses als **Veränderungsblindheit** oder **change blindness** bezeichnete Phänomen gilt es bei der Gestaltung grafischer Anzeigen zu beachten. Es kann sogar ausgenutzt werden, um gezielt Information zu verstecken.

## 2.1.2 Farbwahrnehmung

Die menschliche **Farbwahrnehmung** wurde bereits sehr früh in den bildenden Künsten studiert und gezielt angesprochen. Eine sehr systematische Beschreibung aus dieser Perspektive liefert beispielsweise **Johannes Itten** in seinem Buch *Kunst der Farbe* [56].

Seine Theorie zu Farben, Kontrasten und Farbklingen wird in diesem Abschnitt für die Verwendung bei Bildschirmausgaben adaptiert. Farben unterscheiden wir als Menschen vor allem anhand ihres **Farbtons**, ihrer **Sättigung** und ihrer **Helligkeit**. (Für den Computer wäre im Gegensatz dazu eine Beschreibung durch den Rot- Grün- und Blauanteil naheliegender.) Eine mathematische Beschreibung entlang dieser drei Dimensionen ist der **HSV-Farbraum** (HSV = *Hue, Saturation, Value*, siehe Abbildung 2.3). Die voll gesättigten Farben (S maximal) bilden dabei entlang der Farbtton-Achse (H) die Farben des Regenbogens bzw. des Farbkreises. Rot und Violett haben dabei als elektromagnetische Schwingung ein Frequenzverhältnis von etwa 1:2, entsprechend einer **Oktave** in der akustischen Wahrnehmung. Dies ist eine plausible Erklärung, warum sich das Farbspektrum an dieser Stelle nahtlos zu einem Kreis schließen lässt, genau wie sich die Töne der Tonleiter nach einer Oktave wiederholen. Nimmt man die Sättigung (S) zurück, so vermischt sich die entsprechende Farbe zunehmend mit Grau. Bei maximalem Helligkeitswert (V) geht sie irgendwann in Weiß über, bei minimalem in Schwarz. Im Gegensatz zu den **bunten Farben** des Regenbogens bezeichnet man Schwarz, Weiß und die Grauwerte dazwischen als **unbunte Farben**.

Ein **Farbkontrast** ist ein deutlich sichtbarer Unterschied zwischen zwei (oft räumlich benachbarten) Farben. Folgende Farbkontraste sind im Zusammenhang mit der Gestaltung grafischer Präsentationen von Bedeutung (siehe Abbildung 2.3): Der **Farbe-an-sich-Kontrast** bezeichnet die Wirkung einer einzelnen Farbe im Zusammenspiel mit anderen. Je näher diese Farbe an einer der Primärfarben liegt, umso stärker ist ihr Farbe-an-sich-Kontrast. Solche starken Farben eignen sich besonders gut, wichtige Inhalte einer grafischen Präsentation zu unterstreichen. Sie haben eine Signalwirkung, können jedoch auch mit Symbolwirkung belegt sein (z.B. Ampelfarben Rot und Grün). Ein **Hell-Dunkel-Kontrast** besteht zwischen zwei Farben mit verschiedenen Helligkeiten. Verschiedene bunte Farben haben verschiedene Grundhelligkeiten, wobei Gelb die hellste und Blauviolett die dunkelste Farbe ist. Schwarz-Weiss ist der größtmögliche Hell-Dunkel-Kontrast. Bei der Verwendung verschiedener Farben in einer Darstellung bietet es sich oft an, reine Farbkontraste zusätzlich mit Hell-Dunkel-Kontrasten zu kombinieren, also z.B. Farben verschiedener Grundhelligkeiten auszuwählen. So ist sichergestellt, dass sie auch bei wenig Licht, in Schwarz-Weiss-Ausdrucken und von Farbenblinden (siehe Exkurs auf Seite 17) noch unterscheidbar sind. Ein **Kalt-Warm-Kontrast** besteht zwischen zwei Farben, deren Temperaturwirkung unterschiedlich ist. Farben im Bereich von Rot über Orange bis Gelb werden als warme Farben bezeichnet, während Grün, Blau und Violett als kalte Farben gelten. Diese Terminologie geht auf das menschliche Empfinden dieser Farben zurück (siehe auch Itten [56]) und ist strikt zu trennen vom technischen Messwert der **Farbtemperatur** (ironischerweise wird Temperatur dort genau umgekehrt verwendet: Licht mit einer hohen Farbtemperatur hat einen hohen Anteil der kalten Farbe Blau). Farben verschiedener Temperaturwirkung



mmibuch.de/a/2.3

**Abbildung 2.3:** Die Farben im HSV Farbraum, sowie die im Text genannten Farbkontraste. Dieses Bild ist im Schwarzweiß-Druck leider nicht sinnvoll darstellbar, dafür aber im verlinkten Bild auf der Webseite zum Buch (<http://mmibuch.de/a/2.3/>).

können beispielsweise auf subtile Art unser Verständnis einer grafischen Präsentation unterstützen, indem warme Farben für aktive oder aktuelle Objekte verwendet werden. Ein **Komplementärkontrast** besteht zwischen zwei Farben, die in der additiven Mischung zusammen weiß ergeben. Solche Farbpaare sind beispielsweise Rot – Grün und Blau – Orange. Der Komplementärkontrast ist der stärkste Kontrast zwischen zwei Farben und sollte mit Vorsicht verwendet werden. Ein **Simultankontrast** entsteht durch die direkte räumliche Nachbarschaft unterschiedlicher Farben. Diese stoßen sich gegenseitig im Farbkreis ab. Das bedeutet, dass ein neutrales Grau in der Nachbarschaft eines starken Grüntones rötlich wirkt, während dasselbe Grau in der Nachbarschaft eines Rottones grünlich wirkt. Dies kann in der Praxis dazu führen, dass identische Farben an verschiedenen Stellen einer grafischen Präsentation deutlich unterschiedlich wahrgenommen werden. Ein **Qualitätskontrast** besteht zwischen verschiedenen stark gesättigten Farben, beispielsweise zwischen Rot und Rosa. Verschiedene Farbqualitäten können beispielsweise die Elemente einer grafischen Präsentation in verschiedene Kategorien unterteilen: gesättigt = aktiviert oder im Vordergrund, weniger gesättigt = inaktiv oder im Hintergrund.

#### Exkurs: Farbenblindheit und der Umgang damit

Die am weitesten verbreitete **Farbschwäche** ist die Rot-Grün-Schwäche (Dyschromatopsie), bei der die Farben Rot und Grün schlechter oder gar nicht unterschieden werden können. Von ihr sind über 8% der Männer und fast 1% der Frauen betroffen. Dies bedeutet, dass eine Bildschirmdarstellung, die eine Bedeutung alleine als Unterschied zwischen Rot und Grün ausdrückt, schlimmstenfalls für jeden zwölften Nutzer unbenutzbar ist. Leider wird dieser einfache Sachverhalt auch im Alltag immer noch oft missachtet, wie z.B. bei Kontrolllampen, die lediglich ihre Farbe von Rot auf Grün wechseln. Bei Ampeln ist dieser Bedeutungsunterschied redundant in der Position enthalten: Rot ist dort immer oben. Bei Fußgängerampeln kommt außerdem das dargestellte Symbol hinzu.

Zur Beurteilung dessen, was Farbenblinde sehen, erlauben Bildbearbeitungsprogramme wie z.B. **Adobe Photoshop** – teilweise auch mittels Plugins – die Umrechnung von Bildern in eine entsprechende Darstellung. Sind alle Informationen auch danach noch klar erkennbar, so kann davon ausgegangen werden, dass die getestete grafische Darstellung auch für Rot-Grün-Schwache funktioniert.

Laut Itten [56] lassen sich mittels geometrischer Konstruktionen im Farbkreis bzw. in einer Farbkugel systematisch Farbakkorde konstruieren, deren Farben eine ausgewogene Verteilung besitzen und die damit beispielsweise als **Farbschema** in grafischen Präsentationen dienen könnten. Einen anderen populären Zugang zur Farbauswahl bietet **Cynthia Brewer** mit Ihrer Webseite *Colorbrewer*<sup>1</sup>, auf der fertige Farbskalen verschiedener Kategorien für verschiedene Anwendungszwecke zu finden sind.

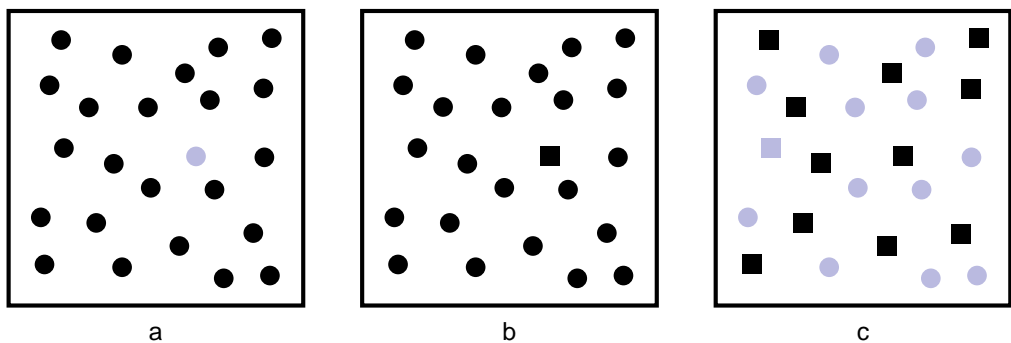
Generell ist bei der Wahl eines Farbschemas für eine grafische Darstellung auf verschiedenste Aspekte zu achten: Zunächst sollte das Farbschema so viele Farben wie nötig, aber nicht unnötig viele Farben enthalten. Unterschiede im Farbton eignen sich eher für

<sup>1</sup><http://colorbrewer2.org/>

die Unterscheidung von Kategorien. Unterschiede in Helligkeit oder Sättigung können eine begrenzte Anzahl von Werten auf einer kontinuierlichen Achse annehmen (z.B. Wichtigkeit oder Aktualität). Gruppen von gleich hellen oder gleich gesättigten Farben wiederum vermitteln logische Gleichartigkeit der so gefärbten visuellen Elemente (z.B. weniger Sättigung = ausgegraut = inaktiv). Kontraste zwischen unbunten und bunten Farben vermitteln eine Signalwirkung. Betrachten Sie unter diesem Aspekt bitte auch einmal das Titelbild dieses Buches.

### 2.1.3 Attentive und präattentive Wahrnehmung

Die Verarbeitung bestimmter visueller Informationen (z.B. Erkennung einfacher Formen, Farben oder Bewegungen) erfolgt bereits im Nervensystem des Auges, nicht erst im Gehirn. Diese Art der Wahrnehmung verläuft daher hochgradig parallel und sehr schnell. Da sie passiert, bevor das Wahrgenommene überhaupt unser Gehirn und unsere Aufmerksamkeit erreicht, heißt sie **präattentive Wahrnehmung** (lat. *attentio* = Aufmerksamkeit). Folgerichtig heißt die Art von Wahrnehmung, der wir unsere Aufmerksamkeit widmen müssen, **attentive Wahrnehmung**. Präattentive Wahrnehmungsprozesse laufen in konstanter Zeit von ca. 200-250ms ab, unabhängig von der Anzahl der präsentierten Reize. Abbildung 2.4a und b zeigen Beispiele hierfür.



**Abbildung 2.4:** Beispiele für präattentive Wahrnehmung (a, b) und attentive Wahrnehmung (c): Im Bild c ist es deutlich schwerer, die eine Form zu finden, die nicht zu den anderen passt.

Der blaue Kreis in Bild 2.4a und das Quadrat in Bild 2.4b springen uns sofort ins Auge. Ohne das Bild sequenziell absuchen zu müssen, sehen wir sofort, wo das unterschiedliche Objekt ist. Farbe und Form sind Merkmale, die präattentiv wahrgenommen werden können. Sobald wir jedoch nach einer Kombination von zwei oder mehreren solcher Merkmale suchen müssen, bricht die präattentive Wahrnehmung zusammen. In Abbildung 2.4c finden wir das helle Rechteck nicht sofort, sondern müssen nacheinander alle Formen nach der Kombination der beiden Merkmale *Helligkeit und Quadrat* absuchen. Die benötigte Zeit wächst linear mit der Anzahl der dargestellten Objekte, und obwohl in der Abbildung nur 24 Objekte in einem Bild enthalten sind, dauert es

schon eine Weile, das helle Quadrat in Abbildung 2.4c zu entdecken. Weitere Merkmale, die präattentiv wahrnehmbar sind, sind Größe, Orientierung, Krümmung, Bewegungsrichtung und räumliche Tiefe. Für die Gestaltung grafischer Präsentationen bedeutet dies, dass Objekte, die sich in genau einem präattentiv wahrnehmbaren Merkmal unterscheiden, sehr schnell in einer großen Menge von Objekten wahrgenommen werden können, während die Kombination von Merkmalen nur relativ langsam zu finden ist. So kann Information in einer Darstellung gezielt hervorgehoben (**Pop-out Effekt**) oder aber auch gut versteckt werden.

### 2.1.4 Gestaltgesetze

Die **Gestaltgesetze** sind eine Sammlung von Regeln, die auf die Gestaltpsychologie Anfang des letzten Jahrhunderts zurückgehen. Sie beschreiben eine Reihe visueller Effekte bei der Anordnungen von Einzelteilen zu einem größeren Ganzen, der Gestalt. Abbildungen 2.5 und 2.6 zeigen Beispiele hierfür. Die Gestaltgesetze sind nirgends als eigentlicher Gesetzestext niedergeschrieben, sondern es findet sich eine Vielzahl von Quellen, die im Kern gemeinsame Gedanken enthalten, diese jedoch oft auf eine bestimmte Anwendung (Musik, Malerei, UI design, ...) beziehen. In diesem Sinne werden auch hier nur die für unseren Anwendungsfall Mensch-Maschine-Interaktion interessanten Gestaltgesetze in einen direkten Bezug dazu gesetzt.

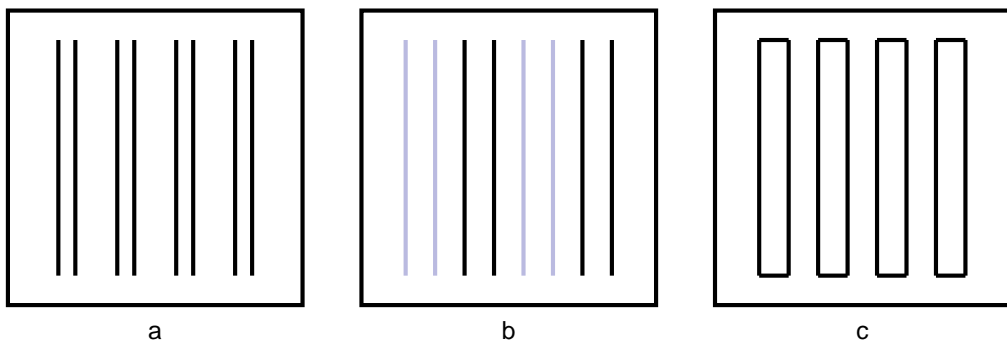


Abbildung 2.5: Beispiele zum Gesetz der Nähe (a), Gesetz der Ähnlichkeit (b) und Gesetz der Geschlossenheit (c)



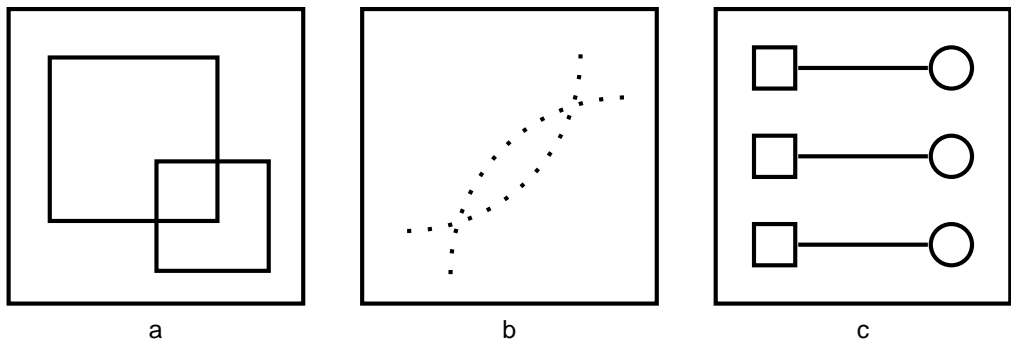
Das **Gesetz der Nähe** besagt, dass Objekte als zueinander gehörig empfunden werden, die nahe beieinander platziert sind, und dass umgekehrt weit voneinander entfernte Objekte als getrennt empfunden werden. Einfachstes Beispiel für eine praktische Anwendung dieses Gesetzes ist die Beschriftung von interaktiven Elementen wie Schalter oder Textfelder in einer grafischen Benutzerschnittstelle: Die Beschriftung wird automatisch dem am nächsten dazu gelegenen Element zugeordnet. Sind die Interface-Elemente zu dicht gepackt, so kann diese Zuordnung mehrdeutig werden, da nicht mehr genügend Abstand zu den anderen Elementen vorhanden ist. Im Inhaltsverzeichnis dieses Buches



werden auf diese Weise auch die Unterabschnitte eines Kapitels visuell als zusammengehörig gruppiert und die verschiedenen Kapitel voneinander getrennt.

Das **Gesetz der Ähnlichkeit** besagt, dass gleichartige Objekte als zueinander gehörig empfunden werden und verschiedenartige als getrennt. Dies wird beispielsweise bei der Darstellung von Tabellen ausgenutzt, in der Zeilen immer abwechselnd weiß und grau eingefärbt sind. Die durchgehenden Farbstreifen führen den Blick und bewirken, dass die Inhalte einer Zeile jeweils als zusammengehörig empfunden werden. Sind umgekehrt die Spalten abwechselnd weiß und grau eingefärbt, so nehmen wir die Einträge einer Spalte als zusammengehörig wahr. Beides funktioniert ohne etwas an den regelmäßigen Zeilen- oder Spaltenabständen (siehe Gesetz der Nähe) zu verändern.

Das **Gesetz der Geschlossenheit** besagt, dass geschlossene Formen als eigenständige Objekte wahrgenommen werden und darin enthaltene Objekte als zueinander gehörig gruppieren. Dabei werden nicht vollständig geschlossene Formen oft zu solchen ergänzt. Dieses Gesetz wirkt z.B. immer dann, wenn wir UI-Elemente in einer rechteckigen Fläche organisieren, die visuell umrandet ist. Viele UI Toolkits unterstützen eine solche inhaltliche Gruppierung durch Layout-Algorithmen und grafische Mittel wie Linien oder räumliche Effekte (herausheben, absenken).



**Abbildung 2.6:** Beispiele zum Gesetz der Einfachheit (a), Gesetz der guten Fortsetzung (b) und Gesetz der Verbundenheit (c)

Das **Gesetz der Einfachheit** (oder auch Gesetz der guten Form) besagt, dass wir Formen immer so interpretieren, wie sie geometrisch am einfachsten sind. Am konkreten Beispiel der Abbildung 2.6a sehen wir automatisch zwei sich überschneidende Quadrate und nur mit einiger Mühe können wir uns davon überzeugen, dass das auch zwei aneinander stoßende winkelförmige Figuren sein könnten, zwischen denen ein Leerraum eingeschlossen ist. Dieses Gesetz müssen wir vor allem beachten, um ungewünschte aber visuell einfachere grafische Anordnungen zu erkennen und zu vermeiden.

Das **Gesetz der guten Fortsetzung** besagt, dass Objekte, die auf einer kontinuierlichen Linie oder Kurve liegen, als zusammengehörig wahrgenommen werden. Im Beispiel in Abbildung 2.6b sehen wir zwei Gruppen von Punkten, die jeweils auf einem Kreisbo-



gen liegen. Dieses Gesetz kommt beispielsweise zum Tragen, wenn sich Verbindungslinien in Diagrammen überschneiden oder wenn viele Messpunkte in einem Diagramm eine ideale Kurve annähern. Es hat außerdem Anwendungen in der Typografie. Im Index dieses Buches liegen beispielsweise Begriffe und Spezialisierungen davon auf verschiedenen Einrückungsebenen. Die Anfangsbuchstaben der Hauptbegriffe ergeben dabei eine senkrechte Linie, ebenso die Anfangsbuchstaben der Unterbegriffe.

Das **Gesetz der Verbundenheit** besagt, dass miteinander verbundene Formen als zusammengehörig wahrgenommen werden. Seine Wirkung ist stärker als die der Nähe oder Ähnlichkeit. So kann beispielsweise in einem Diagramm aus Objekten und Verbindungslinien – nahezu unabhängig von der räumlichen Anordnung und der konkreten grafischen Darstellung der einzelnen Objekte – Zusammengehörigkeit vermittelt werden. Im Beispiel in Abbildung 2.6c ist klar, dass je ein Quadrat und ein Kreis zusammengehören, obwohl sich ohne die Verbindungen nach den Gesetzen der Nähe und Ähnlichkeit eine völlig andere Gruppierung ergäbe.

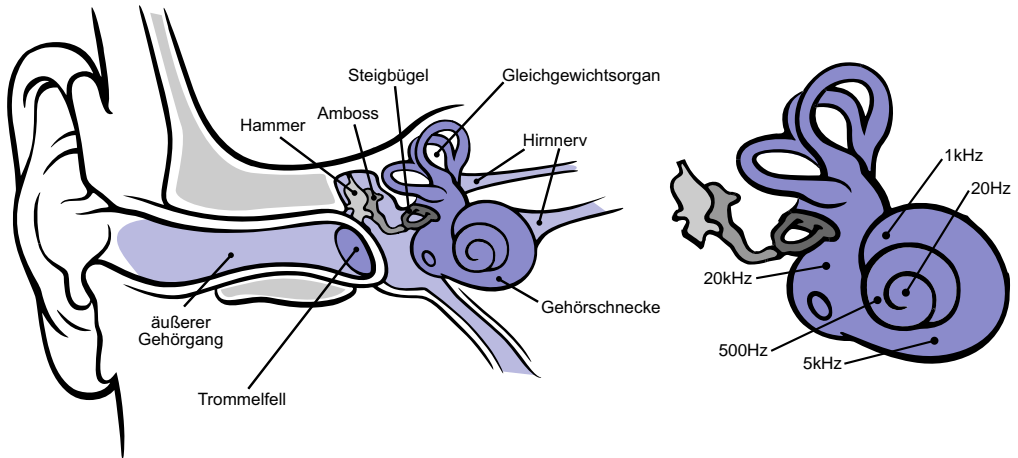
## 2.2 Hörsinn und auditive Wahrnehmung

Der **Hörsinn** ist – wenn auch mit großem Abstand – der am zweithäufigsten genutzte Sinn bei der Mensch-Maschine-Interaktion. Praktisch alle heutigen Desktop-Computer, aber auch Tablets und Mobiltelefone haben die Möglichkeit, Audio auszugeben. Außer zur Musikwiedergabe wird dieser Kanal jedoch meist nur für einfache Signal- oder Warntöne genutzt. Eine Ausnahme bilden Sprachdialogsysteme (siehe Abschnitt 8.1).

### 2.2.1 Physiologie der auditiven Wahrnehmung

**Schall** ist – physikalisch gesehen – eine zeitliche Veränderung des Luftdrucks. **Töne** sind periodische Veränderungen des Luftdrucks mit einer klar hörbaren Grundfrequenz. **Geräusche** haben – im Gegensatz hierzu – einen unregelmäßigen Signalverlauf und enthalten oft viele und zeitlich wechselnde Frequenzanteile. Das menschliche **Ohr** nimmt akustische Frequenzen zwischen etwa 20Hz und 20.000Hz wahr, und zwar nach folgendem Prinzip (siehe hierzu auch Abbildung 2.7):

Die eintreffenden Schallwellen erreichen durch **Ohrmuschel** und äußeren **Gehörgang** das **Trommelfell** und versetzen es in Schwingung. Die **Gehörknöchelchen** (Hammer, Amboss und Steigbügel) leiten diese Schwingungen an die **Gehörschnecke** (lat. **Cochlea**) weiter. Diese ist eine aufgewickelte Röhre mit einer sie längs teilenden Trennwand, der **Basilarmembran**. Auf dieser wiederum sitzen die sogenannten **Flimmerhärchen**, die an verschiedenen Positionen jeweils durch verschiedene Frequenzen angeregt werden und dann einen Sinnesreiz zur weiteren Verarbeitung im Nervensystem abgeben. Verschiedene eintreffende Frequenzen regen somit verschiedene Bereiche der Basilarmembran an und können gleichzeitig wahrgenommen werden: zwei verschiedene Töne werden damit als Zweiklang wahrgenommen, nicht etwa als in der Mitte dazwischen liegenden Mischton, was etwa den Mischfarben im Auge entsprechen würde. Frequenzen, die zu dicht beieinander liegen, maskieren sich gegenseitig und nur noch das stärkere Signal wird wahrgenommen. Weitere Details hierzu finden sich z.B. in Malaka et al. [79].



**Abbildung 2.7:** Links: Vereinfachtes Schnittbild des menschlichen Ohrs, Rechts: Frequenzempfindlichkeit entlang der Gehörschnecke

Die **Frequenzauflösung** des Ohrs bei einer Grundfrequenz von 1kHz beträgt etwa 3Hz. Die Empfindlichkeit des Ohrs für verschiedene Frequenzen ist unterschiedlich und für Frequenzen zwischen 2 und 4kHz am höchsten. Dies entspricht dem Frequenzbereich gesprochener Sprache. Die **Hörschwelle** für die Frequenz von 2kHz wird dabei als Lautstärke von 0dB (0 Dezibel) definiert. Tiefere und höhere Frequenzen werden erst ab höheren Lautstärken wahrgenommen. Der hörbare Lautstärkebereich erstreckt sich bis zur **Schmerzgrenze** oberhalb etwa 120dB, wobei der meist genutzte Bereich (ohne kurzfristige Gehörschäden) eher zwischen 0dB (Hörschwelle) und etwa 100dB (Discolautstärke) liegt, wobei jeweils 6dB einer Verdopplung des **Schalldrucks** entsprechen. Somit beträgt der **Dynamikumfang** des Ohrs etwa  $1 : 2^{17}$  oder etwa  $1 : 125.000$ .

Der Hörsinn ist unter bestimmten Einschränkungen auch in der Lage, den Ort einer Schallquelle zu bestimmen (**räumliches Hören**). Dies wird durch drei verschiedene Effekte ermöglicht (siehe hierzu Abbildung 2.8). Zur Nutzung der ersten beiden Effekte benötigen wir beide Ohren. Die Tatsache, dass die Ohren an verschiedenen Seiten des Kopfes angebracht sind und in entgegengesetzte Richtung zeigen, führt dazu, dass ein Schallsignal von einer Seite beim gegenüberliegenden Ohr wesentlich leiser ankommt, da dieses quasi im Schatten des Kopfes liegt (Abbildung 2.8 rechts). Dieser Lautstärkeunterschied wird englisch als *interaural intensity difference* (**IID**) bezeichnet. Außerdem legt der Schall von einer Schallquelle, die nicht in der Mitte vor, über, unter oder hinter uns liegt, unterschiedlich lange Wege zu den verschiedenen Ohren zurück. Da die Ausbreitungsgeschwindigkeit des Schalls in Luft mit etwa  $300\text{m/s}$  recht niedrig ist, ist das Signal am entfernteren Ohr gegenüber dem nahen Ohr deutlich zeitlich verzögert. Dieser Zeitunterschied heißt auf englisch *interaural time difference* (**ITD**). Wegen der

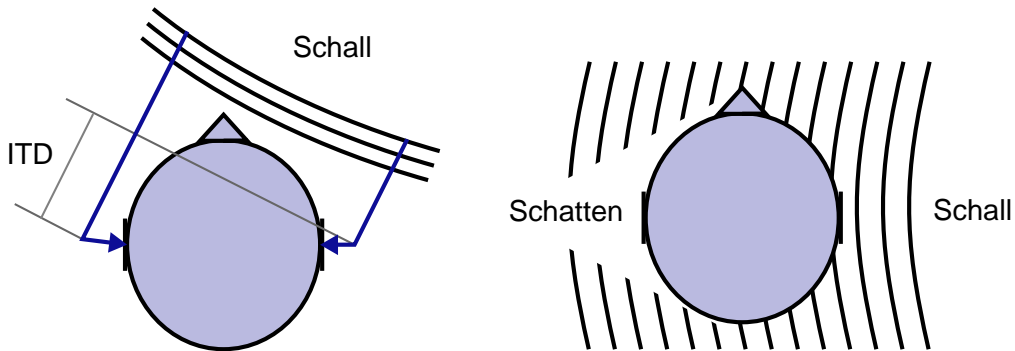


Abbildung 2.8: Links: räumliches Hören mittels Zeitunterschied (ITD), Rechts: räumliches Hören mittels Lautstärkeunterschied (IID)



Beugungs-Eigenschaften des Schalls bei verschiedenen Frequenzen und wegen der großen Wellenlänge tiefer Frequenzen funktioniert das räumliche Hören mittels IID und ITD nicht für tiefe Frequenzen. Diese lassen sich nicht klar orten, was auch der Grund dafür ist, dass Basslautsprecher (Subwoofer) ohne hörbaren Unterschied quasi an beliebiger Stelle im Raum aufgestellt werden können. Hinzu kommt, dass IID und ITD für alle Orte mit gleichem Abstand zu den beiden Ohren (also auf einem Kreis um die Achse durch beide Ohren) auch gleich sind. Dies bedeutet, dass wir mittels IID und ITD nur zwischen rechts und links, nicht aber zwischen vorne, oben, hinten und unten unterscheiden können.

Der dritte Effekt, den wir zum räumlichen Hören ausnutzen, ist die richtungs- und frequenzabhängige Veränderung des akustischen Signals durch die geometrischen und materiellen Eigenschaften des Ohrs und des Kopfes. Signale von oben oder hinten werden beispielsweise durch Haare gedämpft, und die Form der Ohrmuschel bildet Resonanzmuster oder Auslöschungen für bestimmte Frequenzen. Eine Funktion, die diese Veränderungen des Schalls in Abhängigkeit von seiner Richtung und Frequenz beschreibt, heißt *head-related transfer function* (**HRTF**). Jeder Mensch hat eine spezifische HRTF, die sich experimentell ermitteln lässt. Man kann jedoch auch eine gemittelte HRTF verwenden, die für viele Menschen annähernd korrekt ist, und damit räumliche Klänge erzeugen, in denen wir dann auch vorne, hinten, oben und unten unterscheiden können. Solche HRTF-basierten Raumklänge lassen sich nur mittels Kopfhörer abspielen, da sonst ja noch die reale HRTF des Hörers hinzukäme.

Eine weitere wichtige Komponente räumlichen Hörens ist schließlich die Gewohnheit: das Geräusch eines Hubschraubers oder eines Vogels werden wir unwillkürlich von oben hören, während Mensch, Autos, und andere auf der Erdoberfläche befindliche Schallquellen auch in der uns umgebenden Ebene wahrgenommen werden, solange es keine starken anderen Evidenzen gibt. Räumliches Hören kann auch gezielt in der Benutzerschnittstelle eingesetzt werden, indem beispielsweise einem mobilen Benutzer akustische

Signale aus der jeweils passenden Richtung gegeben werden [45]. Tiefe Frequenzen werden nicht nur mit den Ohren sondern auch als Vibrationen mit anderen Körperteilen wahrgenommen. Ein Beispiel hierfür sind die Bassfrequenzen in der Disco. Extrem tiefe Frequenzen können sogar unsere Stimmung beeinflussen, obwohl wir mit den Ohren bewusst keinen Ton wahrnehmen. So wurden in manchen Kirchen Orgelpfeifen unterhalb des hörbaren Frequenzbereichs eingebaut, da diese scheinbar in der Lage sind, uns das Gefühl zu vermitteln, es gehe etwas Übersinnliches vor sich [111].

## 2.2.2 Besonderheiten der auditiven Wahrnehmung

Die akustische Wahrnehmung bildet mit dem zugehörigen **sensorischen Register** ein spezielles Konstrukt, die sogenannte **phonologische Schleife**. In ihr werden akustische Ereignisse wie z.B. Sprache bis zu einer Dauer von etwa 2 Sekunden abgespeichert und durch innerliches Wiederholen immer wieder aufgefrischt. Dabei bestimmen die verfügbare Zeit und die Dauer der einzelnen Einträge die Anzahl der in der Schleife gehaltenen Einträge: wir können uns im Register mehr kurze Worte merken als lange. Zudem lassen sich einige wahrnehmungspsychologische Effekte daraus ableiten und auch experimentell nachweisen, beispielsweise der phonologische Ähnlichkeitseffekt (ähnlich klingende Wörter lassen sich schwerer merken als verschieden klingende) und die artikulatorische Unterdrückung (ständige Wiederholung irrelevanter Äußerungen wie „Ähs“ oder Füllwörter vermindert die Gedächtnisleistung [42]).

Im Alltag nutzen wir die phonologische Schleife beispielsweise dazu, uns Telefonnummern oder andere kurze Informationen zu merken, bis wir sie beispielsweise durch Aufschreiben oder Eingeben dauerhaft speichern können. Wird die Schleife durch neu eintreffende akustische Reize gestört, z.B. durch jemanden, der uns anspricht, so verblasst der darin gespeicherte Inhalt und wird durch den neu wahrgenommenen ersetzt.

Die einfachste Form akustischer Ausgaben am Computer sind kurze Signaltöne oder -geräusche. Besitzen diese Geräusche einen Bezug zur echten Welt, beispielsweise das Zerknüllen oder Zerreißen von Papier beim Entleeren des Papierkorbs in einer Desktop-PC Umgebung, dann spricht man von **Auditory Icons**. Diese symbolisieren – genau wie grafische Icons – bestimmte Objekte oder Vorgänge und werden automatisch mit diesen assoziiert. Abstraktere Klänge, wie z.B. Tonfolgen, deren Bedeutung erlernt werden muss, heißen **Earcons**. Während der Vorteil der direkten Assoziation einer Bedeutung hier wegfällt, bieten diese mehr Spielraum zur Veränderung von Parametern wie Tonhöhe, Tonfolge oder Geschwindigkeit und sind damit ein potenziell mächtigeres Ausdrucksmittel.

## 2.3 Tastsinn und Propriozeption

Die **haptische Wahrnehmung** liefert uns eine Fülle von Informationen über die physikalische Welt. Zur Interaktion mit dem Computer wird sie dagegen nur sehr selten genutzt. Ausnahmen sind Controller für Computerspiele, die Vibration oder Druckkräfte vermitteln, der allseits bekannte Vibrationsalarm am Mobiltelefon, oder die Ausgabe von Braille-Schrift am Computer, die für Blinde den Sehsinn durch den Tastsinn ersetzt.

Als **Tastsinn** oder **taktile Wahrnehmung** bezeichnet man die Wahrnehmung von Berührungen und mechanischen Kräften mit den Tastsinneszellen der Haut, während der allgemeinere Begriff haptische Wahrnehmung auch andere Wahrnehmungsformen wie Temperatur oder Schmerz, sowie Sinneszellen in anderen Bereichen des Körpers umfasst. Die Sinnesorgane der Haut umfassen Sensoren zur Messung von Berührung, mechanischem Druck, und Vibration (Mechanorezeptoren), Temperatur (Thermorezeptoren), sowie Schmerz (Schmerzrezeptoren). Außerdem gibt es mechanische Sensoren in Muskeln, Sehnen und Gelenken sowie inneren Organen. Die Dichte der Sensoren und damit die Genauigkeit und Auflösung des Tastsinnes ist ungleichmäßig am Körper verteilt. Hände und insbesondere die Fingerspitzen, sowie die Zunge weisen eine hohe Sensordichte auf, sind daher sehr empfindlich und können feinere Strukturen erfassen. Andere Körperbereiche wie Bauch oder Rücken enthalten weniger Sensoren und haben ein wesentlich niedrigeres Auflösungsvermögen.

Bisher werden Ausgaben an den Tastsinn nur in ganz wenigen Kontexten verwendet. So verwenden Spiele auf Mobiltelefonen den ohnehin eingebauten Vibrationsmotor, um Spielereignisse wie Kollisionen zu signalisieren. Antiblockiersysteme in Autos versetzen das Bremspedal in Vibration, um das Eingreifen der ABS-Automatik anzuzeigen, und Spurwarnsysteme vermitteln Vibrationen im Fahrersitz auf der Seite, auf der die Spur gerade verlassen wird. Außerdem haben haptische Reize im Bereich der Computerspiele eine gewisse Verbreitung gefunden und werden beispielsweise mittels sogenannter **Force Feedback** Joysticks oder Vibrationsmotoren in Game Controllern vermittelt. Mit wachsender Verbreitung interaktiver Oberflächen (siehe Kapitel 17) wächst auch der Wunsch, die darauf befindlichen grafischen Ausgaben ertastbar zu machen, doch bisher scheitert dies meist an der technischen Umsetzbarkeit.

Mit **Propriozeption** bezeichnet man die Wahrnehmung des eigenen Körpers bezüglich seiner Lage und Stellung im Raum. Diese Selbstwahrnehmung wird durch das **Gleichgewichtsorgan** im Ohr (siehe Abbildung 2.7 links) sowie durch Mechanorezeptoren im Muskel- und Gelenkapparat ermöglicht und liefert uns eine Wahrnehmung der Stellung der Gliedmaßen zueinander und im Raum. Propriozeption ist beispielsweise im Zusammenhang mit Gesten wichtig, da sie das blinde Ausführen kontrollierter Bewegungen ermöglicht. Sie hilft uns außerdem beim Umgang mit dem körpernahen Raum, beispielsweise bei der Raumaufteilung auf Tischen und in Fahrzeugen (Abschnitt 17.2.3).

In manchen Kontexten wirkt die Propriozeption jedoch auch störend, beispielsweise in einfachen Fahrsimulatoren, die die umgebende Landschaft zwar grafisch korrekt darstellen und damit den Eindruck von Bewegung und Beschleunigung vermitteln, während das Gleichgewichtsorgan eine Ruheposition ohne Beschleunigungen signalisiert. Diese Diskrepanz zwischen Sehsinn und Propriozeption kann zu Übelkeit und Schwindel führen, der sogenannten Simulatorübelkeit oder **Cyber Sickness** und führt dazu, dass etwa 10% der Probanden in Simulatorversuchen diese Versuche vorzeitig abbrechen.

## 2.4 Geruchs- und Geschmackswahrnehmung

Die **olfaktorische** oder **Geruchswahrnehmung** ist in der Nase angesiedelt und wird durch eine Vielzahl (über 400) spezialisierter Rezeptoren für bestimmte Substanzen

realisiert. Ein bestimmter Geruch spricht dabei eine oder mehrere Rezeptorarten an und erzeugt den entsprechenden Sinneseindruck. Im Gegensatz zur Farbwahrnehmung ist es nicht möglich, aus wenigen Grundgerüchen alle anderen zu mischen, sondern für jeden spezifischen Geruch muss die entsprechende chemische Kombination abgesondert werden. Der Geruchssinn ist außerdem recht träge, was das Nachlassen von Gerüchen angeht, und besitzt eine starke Akkomodationsfähigkeit: Gerüche, die in gleicher Intensität vorhanden bleiben, werden immer schwächer wahrgenommen.

Die **gustatorische** oder **Geschmackswahrnehmung** ist im Mundraum angesiedelt, besitzt im Wesentlichen die gleichen oben beschriebenen Eigenschaften wie die olfaktorische Wahrnehmung und interagiert auch mit ihr. Die grundlegenden Eigenschaften wie Trägheit, Nichtvorhandensein von Grundgerüchen zur Mischung aller anderen, sowie persönliche und hygienische Vorbehalte gegenüber der Nutzung dieser Sinne bewirken, dass sie zur Interaktion mit Computern bisher nicht nennenswert genutzt werden.



### Übungsaufgaben:

1. Das menschliche Auge hat im Zentrum eine Auflösung von etwa einer Winkelminute. Dies bedeutet, dass zwei Lichtstrahlen, die aus Richtungen kommen, die mindestens eine Winkelminute ( $= 1/60$  Grad) auseinanderliegen, gerade noch als verschiedene Strahlen wahrgenommen werden. Bilder werden normalerweise aus einem Abstand betrachtet, der ungefähr der Bilddiagonalen entspricht. Hierdurch entsteht ein Bildwinkel entlang der Diagonale von etwa 50 Grad (entspricht dem sogenannten Normalobjektiv). Leiten Sie daraus her, ab welcher ungefähren Auflösung (bzw. ab welcher Anzahl von Megapixeln bei einem Seitenverhältnis von 3:2) die Auflösung des Auges erreicht ist, eine weitere Verfeinerung also keinen sichtbaren Qualitätsgewinn mehr bringt.
2. In verschiedenen Kulturkreisen verwenden Fußgängerampeln verschiedene Methoden zur Darstellung der Information, ob die Fußgänger stehen bleiben oder gehen sollen. Während die deutschen Ampelmännchen dreifach redundant encodieren (Farbe, Symbol, Position), verwenden ältere amerikanische Modelle z.B. Schrift (Walk – Don't Walk) oder ein Hand-Symbol. Recherchieren Sie mindestens vier grundlegend verschiedene Ausführungen, beispielsweise mithilfe einer Bilder-Suchmaschine im Web, und diskutieren Sie deren Verständlichkeit für alle Bevölkerungsgruppen.

# 3 Kognition

## 3.1 Gedächtnistypen

Wie in Kapitel 1 beschrieben, besteht das menschliche Gedächtnis aus verschiedenen Teilsystemen mit unterschiedlichen Aufgaben. Die sensorischen Register enthalten eine physische Repräsentation von Wahrnehmungseindrücken, also beispielsweise Lichtreizen auf der Netzhaut oder Tasteindrücken vom Greifen eines Gegenstands. Das Kurzzeit- oder Arbeitsgedächtnis ist das Gedächtnis der Gegenwart und der Ort kognitiver Prozesse. Das Langzeitgedächtnis ist das Gedächtnis der Vergangenheit.

Die **sensorischen Register** sind Kurzzeitspeicher für durch Sinnesorgane wahrgenommene Reize. Sie enthalten Rohdaten, die nicht in interpretierter oder abstrakter Form vorliegen, sondern analog zum physikalischen Stimulus sind. Für die verschiedenen sensorischen Kanäle existieren jeweils eigene sensorische Register, nämlich das ikonische oder visuelle Register, das echoische Register und das haptische Register. Um Inhalte der sensorischen Register in das Kurzzeitgedächtnis zu überführen, ist Aufmerksamkeit erforderlich. Eine vertiefte Darstellung hierzu, auf der auch der folgende Abschnitt beruht, findet sich beispielsweise in Wandmacher [119].

### 3.1.1 Kurzzeitgedächtnis und kognitive Prozesse

Das **Kurzzeitgedächtnis** dient der kurzzeitigen Speicherung einer kleinen Menge symbolischer Informationen als Resultat von Wahrnehmungen oder als Resultat von Aktivierungen im Langzeitgedächtnis. Die Inhalte des Kurzzeitgedächtnisses sind bewusst verfügbar. Das Kurzzeitgedächtnis ist darüber hinaus der Ort, an dem kognitive Prozesse, wie Figur- und Grundunterscheidung und Zeichenerkennung, stattfinden. Kognitive Prozesse bestehen aus elementaren kognitiven Operationen, die als *recognize-act*-Zyklus [21] verstanden werden können. Dabei bedeutet *recognize* die Aktivierung einer Einheit im Langzeitgedächtnis und *act* das Verfügbar werden für kognitive Prozesse im Kurzzeitgedächtnis. Ein Beispiel ist das Erkennen von Buchstaben. Die im visuellen Register vorliegenden sensorischen Daten führen zur Aktivierung einer Repräsentation des Buchstabens im Langzeitgedächtnis (*recognize*) und damit zur symbolischen Repräsentation im Kurzzeitgedächtnis (*act*). Dies dauert zwischen 30ms und 100ms.

Das Kurzzeitgedächtnis als Ort kognitiver Prozesse ist in seiner Kapazität beschränkt. Es ist nicht möglich, mehrere bewusst kontrollierte kognitive Prozesse wirklich parallel auszuführen. Beispielsweise können wir nicht einen Text mit Verständnis lesen und gleichzeitig eine Multiplikationsaufgabe im Kopf lösen. Diese Beschränkung bewusst kontrollierter kognitiver Prozesse auf serielle Ausführung wird auch **Enge des Bewusstseins** genannt.



Automatische kognitive Prozesse hingegen können parallel ausgeführt werden, da sie keine oder wenig kontrollierte Verarbeitungskapazität des Kurzzeitgedächtnisses beanspruchen. Sie können auch parallel mit bewusst kontrollierten kognitiven Prozessen ausgeführt werden. Ein Beispiel für einen automatischen kognitiven Prozess ist das Erkennen bekannter Wörter. Ein weiteres Beispiel ist die automatische visuelle Suche (siehe auch Abschnitt 2.1.3 zur präattentiven Wahrnehmung). Bewusst kontrollierte Prozesse können durch Übung zu automatischen kognitiven Prozessen werden. Dabei können unterschiedliche Grade der Automatisierung erreicht werden. Reale kognitive Prozesse liegen also auf einem Kontinuum zwischen vollständig kontrolliert und vollständig automatisch.

Die Kapazität des Kurzzeitgedächtnisses beträgt durchschnittlich drei (zwischen zwei und vier) Einheiten (siehe Card et al. [21] sowie Abbildung 1.3). Diese Einheiten werden auch **Chunks** genannt. Die Kapazität des Kurzzeitgedächtnisses ist primär durch die Anzahl der vorliegenden Chunks beschränkt und weniger durch deren Informationsgehalt. So können Chunks Ziffern, Wörter, Begriffe oder visuelle Vorstellungseinheiten sein. Angenommen man möchte den **Morse-Code** lernen. Kurze und lange Töne sind die ersten Chunks, die wir uns merken müssen. Mit ein wenig Übung bilden wir aus dem Signal kurz-kurz-kurz den Chunk *S*. Anschließend werden größere Chunks gebildet, und so wird aus kurz-kurz-kurz, lang-lang-lang, kurz-kurz-kurz, irgendwann der Chunk *SOS*. Jeder, der einmal eine Sprache gelernt hat, wird diesen Prozess nachvollziehen können. Für sehr kurze Intervalle (ca. 2s) und mit der vollen kontrollierten Verarbeitungskapazität, also der vollen Aufmerksamkeit, können sieben (zwischen fünf und neun) Buchstaben oder Ziffern im Kurzzeitgedächtnis repräsentiert werden. Dies ist jedoch keine realistische Abschätzung der Kapazität des Kurzzeitgedächtnisses. Die Dauer der Speicherung von Chunks im Kurzzeitgedächtnis hängt von der Auslastung ab. Nach Card et al. [21] beträgt die Speicherdauer bei einem einzelnen Chunk ca. 70s (zwischen 70s und 226s) und bei drei chunks ca. 7s (zwischen 5s und 34s). Diese Zeiten setzen voraus, dass die kontrollierte Verarbeitungskapazität auf andere kognitive Prozesse gerichtet ist.

Die Eigenschaften des Kurzzeitgedächtnisses sind relevant für die Gestaltung von Benutzerschnittstellen. Die *Enge des Bewusstseins* bedeutet, dass schon eine moderate Auslastung mit bewusst kontrollierten kognitiven Prozessen die Fehlerrate deutlich erhöht. Eine detailliertere Darstellung von Informationen auf dem Bildschirm schafft hier keine Abhilfe, da auch die Verarbeitung dieser Information kognitive Kapazität in Anspruch nimmt. Mögliche Lösungsstrategien sind Automatisierung durch Übung und **Superzeichenbildung** (engl. **chunking**). Mit *Superzeichenbildung* ist die Bildung von Chunks mit höherem Informationsgehalt gemeint. Diese Strategie ist erfolgreich, da die Kapazität des Kurzzeitgedächtnisses primär durch die Anzahl und nicht die Komplexität der referenzierten Einheiten beschränkt ist. Ein Beispiel ist die Darstellung von Zahlen. Die dreistellige Hexadezimaldarstellung 1C8 kann leichter behalten werden als die neunstellige Binärdarstellung 111001000.

### 3.1.2 Langzeitgedächtnis

Das Langzeitgedächtnis enthält die Gesamtheit unseres Wissens, unseres Könnens und unserer Erfahrungen. Die Inhalte des Langzeitgedächtnisses können durch Erinnern



wieder im Kurzzeitgedächtnis verfügbar gemacht werden. Nicht aktivierte Inhalte des Langzeitgedächtnisses sind aus dem Bewusstsein verschwunden. Man unterscheidet zwischen **deklarativem** und **prozeduralem Wissen**.

**Deklaratives Wissen** ist Faktenwissen, also Informationen über Tatsachen, die sich als assoziatives Netzwerk modellieren lassen. Kognitive Prozesse nutzen deklaratives Wissen, um aus im Kurzzeitgedächtnis vorliegenden Informationen weitere Informationen abzuleiten. Dies versetzt uns in die Lage, Schlussfolgerungen zu ziehen und kognitive Problemstellungen zu lösen. Dies geschieht durch das Finden von Analogien, die Anwendung von Regeln und das Überprüfen von Hypothesen. Deklaratives Wissen lässt sich weiter in episodisches Wissen, also das Wissen über Erfahrungen, die wir gemacht und Ereignisse, die wir erlebt haben, sowie semantisches Wissen, also Wissen über Sachverhalte unterscheiden.

**Prozedurales Wissen** ist operatives Handlungswissen. Es umfasst kognitive Fertigkeiten wie Kopfrechnen und motorische Fertigkeiten, wie Fahrradfahren, Klavierspielen oder **Jonglieren**. Prozedurales Wissen ist typischerweise nicht bewusst. Es kann nur schwer verbalisiert werden und häufig nur durch motorische Wiederholungen gelernt werden. Es ist wichtig zu beachten, dass das Langzeitgedächtnis nur unsere Interpretation und unser Verständnis von Sachverhalten und Erfahrungen speichert und sich auch nachträglich durch neue Erfahrungen verändern kann. Die Kapazität des Langzeitgedächtnisses ist praktisch nicht limitiert. Das Problem liegt in einer geeigneten Organisation und dem möglichen Lernaufwand, die es ermöglichen, die Inhalte des Langzeitgedächtnisses später im Kurzzeitgedächtnis wieder verfügbar zu machen.

Es gibt zwei unterschiedliche Arten Wissen aus dem Langzeitgedächtnis wieder in das Kurzzeitgedächtnis zu rufen und es damit situativ verfügbar zu machen: das **Erinnern** (engl. **Recall**) und das **Wiedererkennen** (engl. **Recognition**). Beim Erinnern müssen wir aktiv Elemente aus unserem Langzeitgedächtnis reproduzieren, z.B. wenn wir nach der Marke unseres Fahrrads gefragt werden. Das Wiedererkennen beruht hingegen auf der Präsentation eines Elements, welches wir schon kennen, z.B. wenn wir gefragt werden, ob unser Fahrrad von einer bestimmten Marke ist. Fragen, die auf Wiedererkennung beruhen, können in der Regel mit Ja oder Nein beantwortet werden, während Fragen die Erinnern erfordern, in der Regel das Gedächtniselement selbst produzieren müssen. Wiedererkennen fällt meistens leichter als Erinnern. Moderne Benutzerschnittstellen setzen daher vornehmlich auf Elemente, die ein einfaches Wiedererkennen von Funktionen ermöglichen - dies ist z.B. auch ein Grund für die Verwendung grafischer Bedienelemente und Repräsentationen (**Icons**), da visuelle Elemente von uns in der Regel besser wiedererkannt werden als textuelle. Für die Entwicklung von Benutzerschnittstellen ist die Unterscheidung der beiden Konzepte daher von großer Bedeutung (siehe auch Abschnitt 13.2.2) und die Berücksichtigung der unterschiedlichen Arten uns zu erinnern kann zu Benutzerschnittstellen führen, die weniger frustrierend für die Anwender sind [80].

## 3.2 Lernen

Viele Funktionen von komplexen Benutzerschnittstellen, wie z.B. die eines Textverarbeitungsprogramms, erfordern von Benutzern eine erhebliche Einarbeitungszeit. Benutzer müssen die Funktionen der Benutzerschnittstelle zunächst **lernen**, um diese später schnell und effektiv einsetzen zu können. Aus Sicht des Entwicklers einer solchen Benutzerschnittstelle ist es daher wichtig zu verstehen wie menschliches Lernen funktioniert und auf welchen kognitiven Prozessen es beruht. Basierend auf diesem Verständnis werden dann Benutzerschnittstellen entwickelt, die Lernprozesse effektiv unterstützen. Grundsätzlich kann Information auf unterschiedlichste Art und Weise unterrichtet werden. Dazu gehört der klassische Frontalunterricht (basierend auf einer Verbalisierung der Information) genauso wie das Lernen aus Büchern, anhand von Modellen oder aber direkt während der Ausführung einer Tätigkeit.

Entscheidend für das Design neuer Benutzerschnittstellen ist hierbei die **Trainings-effizienz** [128]. Die Trainingseffizienz eines Lernverfahrens ist am höchsten, wenn 1.) der beste Lerneffekt in kürzester Zeit erzielt wird, 2.) die längste Erinnerungszeit erreicht wird und 3.) dieses einfach und günstig in einer Benutzerschnittstelle umgesetzt werden kann. Dabei spielt es eine große Rolle, wie gut Wissen, das durch ein Lernverfahren erworben wird, dann in der tatsächlichen Benutzung abgerufen werden kann. Man spricht in diesem Zusammenhang auch von der **Transferleistung** [49]. Wie gut die Transferleistung bei einem neuen Lernverfahren ist, wird üblicherweise experimentell ermittelt [128]. Dabei werden die Lernenden in zwei Gruppen unterteilt. In der **Kontrollgruppe** wird ohne Lerntechnik gelernt, z.B. durch *Versuch und Irrtum* bis ein Erfolgskriterium bei der Ausführung der Zieltätigkeit (z.B. fehlerfreie Formatierung eines Textes in einem Textverarbeitungsprogramms) erreicht wird. Die dazu benötigte Zeit wird gemessen. Die zweite Gruppe, die **Transfergruppe**, lernt zunächst eine gewisse Zeit mithilfe der neuen Technik (z.B. einem interaktiven Übungskurs zur Textverarbeitung). Danach wird die Zeit gemessen, die die Kontrollgruppe benötigt, um die Zieltätigkeit erfolgreich auszuführen. Ist die Summe beider Zeiten der Transfergruppe geringer als die Zeit der Kontrollgruppe, dann hat offensichtlich die neue Lerntechnik einen nützlichen Lerntransfer erreicht. Da insgesamt weniger Zeit zum Erreichen eines Lernerfolges benötigt wurde, erzielt die neue Technik eine höhere Trainingseffizienz als die alte Vorgehensweise. Auf diese Art lassen sich auch verschiedene Lerntechniken miteinander vergleichen. Es kann sogar passieren, dass Lerntechniken eine negative Trainingseffizienz besitzen, d.h. sie behindern das Lernen. In diesem Fall wäre die kumulierte Zeit der Transfergruppe zur Erreichung eines Erfolgskriteriums größer als die Zeit der Kontrollgruppe, die keine Lerntechnik verwendet hat.

Häufig werden diese drei Maße unterschieden: die **Transferleistung**, die **Transfereffektivität** und die **Lernkosten** (vgl. hierzu auch Wickens und Hollands [128]). Die Transferleistung wird als prozentuelle Darstellung des Zeitgewinns gegenüber der Kontrollgruppe formalisiert, allerdings ohne die zusätzliche investierte Lernzeit der Transfergruppe zu berücksichtigen. Dabei ist:

$$\text{Transferleistung} = \frac{\text{Zeit}_{\text{kontrollgruppe}} - \text{Zeit}_{\text{transfergruppe}}}{\text{Zeit}_{\text{kontrollgruppe}}} * 100$$

Die Transfereffektivität berücksichtigt hingegen die investierte Lernzeit der Transfergruppe und lässt sich aus dem Verhältnis der gewonnenen Zeit (d.h. der Differenz  $Zeit_{kontrollgruppe} - Zeit_{transfergruppe}$ ) und der Zeit, die von der Transfergruppe mit der neuen Lerntechnik investiert wurde, bestimmen:

$$Transfereffektivität = \frac{Zeit_{kontrollgruppe} - Zeit_{transfergruppe}}{Zeit_{neueLerntechnik}}$$

Das heißt, wenn die Gesamtzeit der Transfergruppe gleich der Zeit der Kontrollgruppe ist, haben wir eine Transfereffektivität von 1. Die neue Lerntechnik richtet zwar keinen Schaden an, hilft aber auch nichts. Erst wenn die Transfereffektivität  $> 1$  wird, erzielt die neue Lerntechnik einen positiven Effekt. Ist die Transfereffektivität hingegen  $< 1$ , erzielt die Lerntechnik einen negativen Effekt. Trotzdem können Transfereffektivitäten von  $< 1$  sinnvoll sein, insbesondere wenn andere Kriterien berücksichtigt werden müssen. Wird z.B. in einem Flugsimulator trainiert, kann dies zwar zu einer Transfereffektivität von  $< 1$  führen (d.h. die gleiche Tätigkeit wird in einem echten Flugzeug in der Luft schneller gelernt). Das Lernen im Flugsimulator ist aber wesentlich sicherer und auch kostengünstiger. Bei der Einführung einer neuen Lerntechnik wird daher häufig das Verhältnis der Lernkosten zwischen neuer Technik und herkömmlicher Technik gebildet:

$$Trainingskostenverhältnis = \frac{\text{Trainingskosten der neuen Lerntechnik (pro Zeiteinheit)}}{\text{Trainingskosten der bisherigen Methode (pro Zeiteinheit)}}$$

Je größer dieser Wert, desto günstiger ist die neue Lerntechnik gegenüber der herkömmlichen Methode. Eine Entscheidungshilfe, ob es sich lohnt eine neue Lerntechnik einzuführen kann durch Multiplikation von Transfereffektivität und Trainingskosten erreicht werden. Ist dieser Wert  $> 1$  lohnt sich die Einführung der neuen Technik, ist der Wert  $< 1$  lohnt sich die Einführung nicht. Nach Wickens und Hollands [128] werden acht unterschiedliche Lernmethoden unterschieden, von denen die wichtigsten hier besprochen werden sollen.

Die **Praktische Ausübung**, häufig auch als *Learning by doing* bezeichnet, ist ein Konzept mit dem jeder vertraut ist. Sie besagt schlicht, dass durch Ausübung einer Tätigkeit diese gelernt wird. Die interessante Frage ist dabei, wie lange eine Tätigkeit ausgeübt werden sollte, um sie angemessen zu beherrschen. In der Regel steigt der Lernerfolg monoton mit der investierten Zeit, d.h. auch über lange Zeit verbessert man seine Fähigkeiten, lernt also kontinuierlich. Allerdings muss man hier je nach Lernziel unterscheiden. Eine Tätigkeit ohne Fehler zu absolvieren ist in der Regel schneller zu erreichen, als die Tätigkeit mit erhöhter Geschwindigkeit auszuüben. Eine Tastatur fehlerfrei zu bedienen, kann beispielsweise recht schnell erreicht werden. Dieselbe Bedienung in erhöhter Geschwindigkeit durchzuführen dauert wesentlich länger. Es kann also sinnvoll sein, eine Tätigkeit auch dann noch zu trainieren, wenn bereits Fehlerfreiheit erreicht wurde.

Das **Trainieren von Teilaufgaben** zerlegt eine komplexe Tätigkeit in einfachere Bestandteile, die dann getrennt gelernt werden. Entscheidend ist, dass die Tätigkeit in entsprechende Teile zerlegt werden kann. Das Einüben eines Klavierstückes, welches aus verschiedenen musikalischen Sequenzen besteht, kann zerlegt werden. In der Regel werden hier schwierige und einfache Passagen getrennt geübt und dann zu dem

Gesamtmusikstück zusammengesetzt. Auch komplexe Vorgänge, die parallel ausgeführt werden, können zerlegt werden. Dabei muss allerdings sorgfältig vorgegangen werden, da parallel ausgeübte Tätigkeiten häufig von einander abhängig sind. Ein Beispiel ist das Fahrradfahren, welches die Stabilisierung des Gleichgewichts, das Treten der Pedale und das Bedienen des Lenkers parallel erfordert. Die Stabilisierung des Gleichgewichts hängt eng mit dem Lenken zusammen. Insofern macht es wenig Sinn, diese Tätigkeiten von einander zu trennen. Das Treten hingegen ist einigermaßen unabhängig von den beiden anderen Tätigkeiten. So erklärt sich der bessere Trainingserfolg bei der Verwendung eines Dreirades und eines Laufrades im Vergleich zur Benutzung von Stützrädern: Beim Dreirad wird das unabhängige Treten geübt, beim Laufrad das Lenken und die Balance. Bei Stützrädern hingegen werden Lenken und Treten gleichzeitig geübt ohne die eng damit verknüpfte Balance.

Das **Lernen durch Beispiele** ist eine weitere wichtige Technik, die vielfach angewandt wird. Es hat sich gezeigt, dass gut gewählte Beispiele hohe Lernerfolge erzielen können. Dies erklärt auch den großen Erfolg von *Youtube Videos* beim Erlernen neuer Tätigkeiten und Konzepte. Allerdings verweist Duffy [28] darauf, dass es nicht nur die Beispiele selbst sind, die den Lernerfolg bestimmen, sondern die Art ihrer Verwendung. Eine tiefere Interaktion mit dem Beispiel ist notwendig, indem beispielsweise geeignete Fragen gestellt werden oder die Tätigkeit direkt nach dem Beispiel geübt wird, das Beispiel also in Kombination mit einer praktischen Ausübung verwendet wird. Aus diesem Grund gehen viele Übungsaufgaben dieses Buches auf praktische Beispiele ein.

### 3.3 Vergessen

Der psychologische Prozess des Vergessens bezeichnet den Verlust der Erinnerung, also der Fähigkeit, sich an Elemente aus dem Langzeitgedächtnis zu erinnern. Was wir vergessen hängt von vielen Faktoren ab, insbesondere von den zuvor gelernten Elementen selbst und wie diese mit bereits gelernten Elementen in unserem Gedächtnis in Verbindung gebracht werden können. So ist es beispielsweise viel leichter, sich Wörter von Objekten zu merken (Fahrrad, Haus, Auto, Boot) als Wörter von abstrakteren Konzepten (z.B. Gesundheit, Vergnügen, Zukunft, Rückenwind). Wann immer wir zu lernende Elemente in Verbindung mit existierenden Gedächtniselementen bringen können, fällt uns das Lernen leichter und wir vergessen das Erlernte nicht so schnell. Das Vergessen als Prozess ist, wie viele andere kognitive Prozesse, noch nicht völlig verstanden. Zwei Theorien werden in diesem Zusammenhang besonders häufig in der Literatur genannt: die Spurenverfallstheorie und die Interferenztheorie.

Die **Spurenverfallstheorie** geht auf den Psychologen Hermann Ebbinghaus zurück, der als einer der ersten in Gedächtnisversuchen gezeigt hat, dass mit verstreichender Zeit die Erinnerung an bedeutungslose Silben immer schwerer fällt [30]. Genau darauf basiert die Idee der Spurenverfallstheorie, die annimmt, dass die *Spuren* im Gedächtnis verblasen. Inzwischen geht man allerdings davon aus, dass der Faktor Zeit dabei zwar eine Rolle spielt, viel wichtiger aber die Anzahl und die Zeitpunkte der Aktivierung des Gedächtniselements sind. Elemente, die wir seltener verwenden, werden demnach schneller vergessen [101]. Dies deckt sich auch mit Ergebnissen aus der Hirnforschung, wo man ähnliche Prozesse auf neuronaler Ebene nachweisen konnte.

Die **Interferenztheorie** geht hingegen davon aus, dass neu erlernte Gedächtniselemente mit alten Elementen interferieren und diese bei bestimmten Bedingungen verloren, bzw. vergessen werden. Ein typisches Beispiel ist, dass man nach einem Umzug in der neuen Stadt neue Straßennamen lernen muss. Diese interferieren mit den gelernten Straßennamen der alten Stadt, die in Konsequenz anschließend schlechter erinnert werden. Manchmal tritt allerdings der gegenteilige Effekt zu Tage und man bezeichnet eine der neuen Straßen mit einem ähnlichen Straßennamen aus der alten Stadt. Dieser Effekt wird **proaktive Inhibition** genannt und ist eine mögliche Fehlerquelle bei der Interaktion (vgl. Abschnitt 5.3).

Man weiß auch, dass das Vergessen von emotionalen Faktoren abhängig ist. Erinnerungen, die mit starken positiven oder negativen Emotionen verknüpft werden, halten sich nachweislich länger, als solche die wir mit neutralen Gefühlen in Verbindung bringen. Nach Jahren noch erinnert man sich an schöne Erlebnisse, z.B. ein besonders wichtiges Geburtstagsfest, in vielen Details, während die Erlebnisse des Alltags schnell in Vergessenheit geraten. Interessant ist auch, dass Gedächtniselemente an Orten besser erinnert werden, an denen sie erlernt wurden. So konnten Godden und Baddeley zeigen, dass Taucher, die im Wasser lernen, sich im Wasser besser an das Gelernte erinnern können [39]. Man spricht in diesem Fall vom **kontextabhängigen Gedächtnis** und es gibt Hinweise darauf, dass man seine Gedächtnisleistung auch dann verbessern kann, wenn man sich den Kontext nur vorstellt, z.B. wenn man sich während einer Prüfung die ursprüngliche Lernumgebung vor Augen führt.

## 3.4 Aufmerksamkeit

Der Begriff der Aufmerksamkeit ist vielschichtig und von großer Bedeutung für viele Aspekte der Mensch-Maschine-Interaktion. Daher bezieht sich nicht nur dieser Abschnitt auf die Phänomene der menschlichen Fähigkeit zur Informationsextraktion und -verarbeitung, sondern auch weitere in diesem Buch, z.B. der Abschnitt 2.1.3 über präattentive Wahrnehmung. An dieser Stelle beschäftigen wir uns mit der Aufmerksamkeit, wie sie uns aus dem Alltag geläufig ist: der Fähigkeit, uns auf einen Informationskanal zu konzentrieren, bzw. der Anfälligkeit, uns ablenken zu lassen. Aufmerksamkeit und Ablenkung spielen auch bei Benutzerschnittstellen eine große Rolle, da z.B. die Fähigkeit, relevante Details zu extrahieren und sich auf mehrere Aufgaben gleichzeitig zu konzentrieren bei funktionierenden Benutzerschnittstellen unabdingbar ist. Gute Benutzerschnittstellen lenken bewusst oder unbewusst die Aufmerksamkeit der Benutzer. Insofern muss ihr Entwickler verstehen, welche Aspekte der Aufmerksamkeit für den Entwurf von Benutzerschnittstellen von Bedeutung sind.

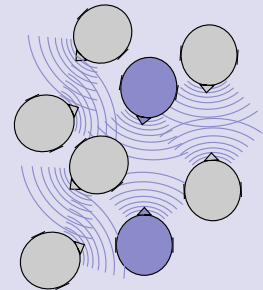
Nach Wickens und Hollands [128] lassen sich in diesem Zusammenhang drei Kategorien von Aufmerksamkeit unterscheiden: die **selektive Aufmerksamkeit**, die **fokussierte Aufmerksamkeit**, und die **geteilte Aufmerksamkeit**. Die selektive Aufmerksamkeit führt dazu, dass wir nur bestimmte Aspekte unserer Umgebung wahrnehmen. Wenn wir beispielsweise beim Gehen unser Mobiltelefon bedienen, werden weitere Umgebungsfaktoren ausgeblendet, was zu gefährlichen Situationen führen kann. Einer der Autoren ist in einer ähnlichen Situation beim Aussteigen aus einem Fernzug zwischen Wagen und Bahnsteigkante gefallen und musste so auf schmerzliche Weise lernen, dass die selekti-

ve Aufmerksamkeit nicht nur Segen, sondern auch Fluch bedeuten kann. Hinter vielen Unfällen im Flug- und Autoverkehr mit *menschlichem Versagen* versteckt sich in Wirklichkeit die selektive Aufmerksamkeit, die dazu geführt hat, dass kurz vor dem Unfall die *falschen* Details wahrgenommen wurden.

Die **fokussierte Aufmerksamkeit** ist eng mit der selektiven Aufmerksamkeit verwandt. Während letztere unsere Fähigkeit beschreibt, bewusst bestimmte Aspekte unserer Umgebung wahrzunehmen, beschreibt erstere den Prozess der unbewussten Konzentration, die wir manchmal trotz Anstrengung nicht unterdrücken können. Fokussierte Aufmerksamkeit tritt beispielsweise auf, wenn sich beim Lernen unsere Aufmerksamkeit statt auf das vor uns liegende Buch auf das im Hintergrund stattfindende Gespräch fokussiert. Fokussierte Aufmerksamkeit ist daher stärker durch die Umgebung determiniert, z.B. durch die Art und Weise wie diese gestaltet ist. Eine grafische Benutzerschnittstelle mit vielen kleinen visuellen Elementen und einem großen, wird zwangsweise zu einer fokussierten Aufmerksamkeit auf das größere Element führen. Wird dieses z.B. durch Bewegung noch weiter hervorgehoben verstärkt sich der Effekt. Hier erkennt man die Problematik der fokussierten Aufmerksamkeit: wird sie sorgsam eingesetzt, kann sie als effizientes Kommunikationsmittel verwendet werden. Wird zuviel Gebrauch von ihr gemacht, wird die Benutzerschnittstelle schnell als unübersichtlich und verwirrend wahrgenommen.

#### Exkurs: Der Cocktail-Party-Effekt

Die Fähigkeit des Menschen seine Aufmerksamkeit auf eine bestimmte Schallquelle zu richten und diese auch in lauten Umgebungen mit vielen Schallquellen zu verstehen, wird als **selektives Hören** bezeichnet. Diese Fähigkeit ist eng verknüpft mit dem räumlichen Hören, welches uns erlaubt eine Schallquelle im Raum zu lokalisieren. Ist uns dies gelungen, so werden andere Schallquellen in ihrer Lautstärke gedämpft und weniger laut wahrgenommen. Ein gutes Beispiel einer solchen Situation ist die Cocktail-Party auf der sich viele Gäste miteinander unterhalten und man den Gesprächspartner gut verstehen kann, obwohl sich viele andere Partygäste in unmittelbarer Nähe befinden. Die wahrgenommene Lautstärke der anderen Schallquellen wird im Verhältnis zur physikalisch gemessenen Lautstärke um bis zu 15dB gedämpft, d.h. wir nehmen unseren Gesprächspartner zwei- bis dreimal lauter wahr als andere Gäste.



Die dritte Kategorie ist die **geteilte Aufmerksamkeit**. Diese beschreibt die Fähigkeit des Menschen, seine Aufmerksamkeit auf verschiedene Dinge gleichzeitig zu verteilen. Ist diese Fähigkeit gestört, können viele Aufgaben nicht zufriedenstellend gelöst werden, mit teilweise verheerenden Folgen. Gelingt es z.B. beim Autofahren nicht, gleichzeitig das Navigationsgerät und die Straße im Auge zu behalten, verfährt man sich im besten Fall nur. Die Einschränkungen bei der Fähigkeit die Aufmerksamkeit zu teilen, führen auch bei Benutzerschnittstellen zu suboptimalen Ergebnissen, z.B. wenn zur Lösung einer Aufgabe mehrere Anwendungen gleichzeitig verwendet werden müssen. Häufig

wird die geteilte Aufmerksamkeit mit unserer Fähigkeit in Verbindung gebracht, unsere kognitiven Fähigkeit zeitlich auf mehrere Aufgaben zu verteilen und die unterschiedlichen Ergebnisse miteinander zu verknüpfen. Beim Autofahren beschreibt dies z.B. die Fähigkeit, kurz auf die Instrumente zu blicken und dann wieder auf die Straße, um zu entscheiden ob man zu schnell fährt. Dies muss dann permanent wiederholt werden, um ohne Geschwindigkeitsübertretung durch die Stadt zu fahren. Eine nähere Betrachtung auf Basis **kognitiver Ressourcen** folgt in Abschnitt 3.5.2.

Eine gängige Metapher, die in der kognitiven Psychologie verwendet wird ist die des Scheinwerfers. Ähnlich wie einen Scheinwerfer richten wir unsere Aufmerksamkeit auf bestimmte Elemente. Es zeigt sich in Untersuchungen, dass die Richtung und die Breite der Aufmerksamkeit (wie bei einem Scheinwerfer) gesteuert werden kann [70]. Auch wenn häufig im Zusammenhang mit Aufmerksamkeit die visuelle Aufmerksamkeit gemeint ist, ist das Konzept im Wesentlichen unabhängig von der betrachteten Modalität. Bedingt durch die Unterschiede unserer Sinne, ergeben sich naturgemäß unterschiedliche Ausprägungen von Aufmerksamkeit. Aber selbst beim auditiven Sinn gibt es klare Hinweise der selektiven und fokussierten Aufmerksamkeit, z.B. beim **Cocktail-Party-Effekt** [23] (siehe Exkurs auf Seite 34).

## 3.5 Kognitive Belastung

Die Beschränkung unseres kognitiven Apparates wird uns schnell bewusst, wenn wir vor einer schwierigen Aufgabe stehen, die wir unter Zeitdruck lösen müssen, wenn wir mehrere schwere Aufgaben gleichzeitig bewältigen wollen, oder unter emotionaler Belastung Entscheidungen treffen müssen. Aus diesem Grund wurde der **kognitiven Belastung** bereits früh in der psychologischen Forschung große Aufmerksamkeit gewidmet. Auf Miller und seinen einflussreichen Aufsatz aus den 60er Jahren, geht die Daumenregel zurück, dass das Arbeitsgedächtnis fünf bis neun Gedächtnisinhalte ( $7 \pm 2$ ) zur gleichen Zeit vorhalten kann [84]. Während inzwischen davon ausgegangen wird, dass diese Regel so pauschal nicht zu halten ist, ist unstrittig, dass unser Arbeitsgedächtnis sehr stark limitiert ist.

Neben dieser generellen **Arbeitsgedächtnisbelastung** manifestiert sich die kognitive Belastung insbesondere bei **Mehrfachaufgaben**, wie sie bei jeder etwas komplexeren Tätigkeit auftreten. Ein vertrautes Beispiel ist wieder das Autofahren, wobei gleichzeitig Lenkrad und Pedal bedient sowie die Straßensituation beobachtet und ausgewertet werden muss. Daher ist es von entscheidender Bedeutung, die kognitive Belastung auch quantitativ messen und bewerten zu können. Aus Sicht der Mensch-Maschine-Interaktion eröffnen sich dann Möglichkeiten der Bewertung von Benutzerschnittstellen, bzw. ein systematischer Ansatz, Benutzerschnittstellen von vornherein so zu entwickeln, dass die durch sie entstehende kognitive Belastung minimiert wird.

Im Folgenden diskutieren wir zunächst den allgemeinen Begriff der **Arbeitsgedächtnisbelastung**, wenden uns dann dem Themenkomplex der **Mehrfachaufgaben** zu, um abschliessend einige Messverfahren für kognitive Belastung vorzustellen.



### 3.5.1 Arbeitsgedächtnisbelastung

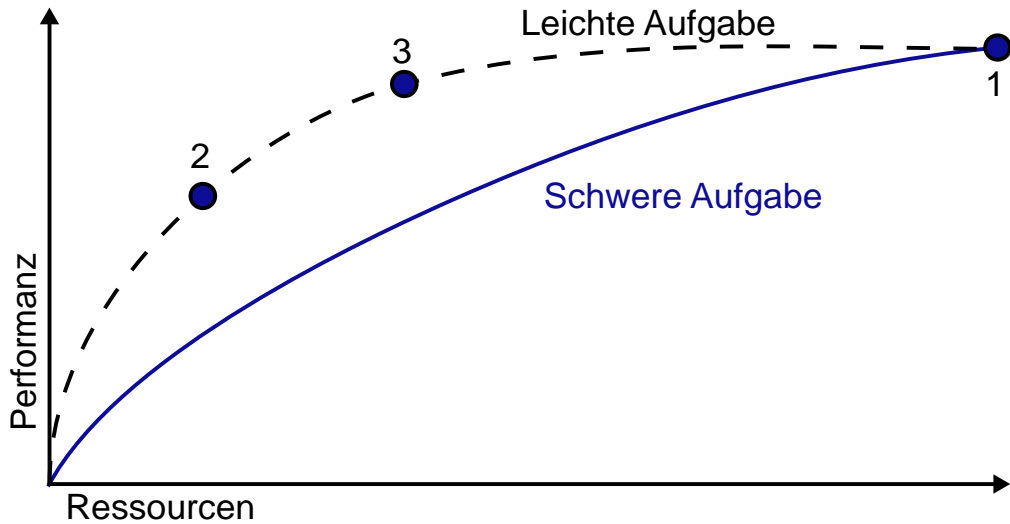
Gerade beim Lernen neuer Aufgaben entstehen hohe Belastungen für das Arbeitsgedächtnis (vergl. dazu auch Abschnitt 3.2). Zur Beschreibung dieser Art von Belastung wurden in der Vergangenheit mehrere Theorien entwickelt, unter denen die prominenteste, die **Cognitive Load Theory** von Sweller et al. [110] ist. Sie vermittelt ein Verständnis davon, wie Information während des Lernprozesses zu gefestigtem Wissen wird, d.h. wie Information dauerhaft vom Kurzzeitgedächtnis ins Langzeitgedächtnis transferiert wird. Sweller hat während seiner Untersuchungen z.B. festgestellt, dass viele Lernmethoden ineffektiv sind, weil sie die Arbeitsgedächtnisbelastung nur unzureichend berücksichtigen. Eine wichtige Erkenntnis der *Cognitive Load Theory* ist, dass beim Präsentieren von Lernmaterialien durch die richtige Wahl komplementärer Modalitäten (z.B. die Verknüpfung von Text und Bildern in einer Bedienungsanleitung) die Arbeitsgedächtnisbelastung reduziert werden kann, da unterschiedliche Typen von Arbeitsgedächtnis (in diesem Fall das visuelle und das verbale Arbeitsgedächtnis) parallel verwendet werden können. Wird die gleiche Information z.B. als reiner Text präsentiert, kann dies zu einer Überlastung des Arbeitsgedächtnisses, und in Konsequenz zu einem verlangsamten Lernprozess führen [81]. Es stellte sich weiterhin heraus, dass neben der Verwendung mehrerer Modalitäten darauf geachtet werden muss, dass die Informationen eng miteinander verknüpft werden. So sollten Grafiken in der Nähe der ergänzenden Textstellen stehen, da sonst die vorteilhaften Effekte auf die Arbeitsgedächtnisbelastung durch zusätzlichen Integrationsaufwand der beiden Informationsarten zunichte gemacht werden könnten. Dieser negative Effekt tritt z.B. nachweislich auf, wenn Grafiken in einem Buch auf einer anderen Seite als der referenzierende Text platziert wurden<sup>1</sup>.

### 3.5.2 Belastung durch Mehrfachaufgaben

Aufgaben des alltäglichen Lebens sind häufig durch die Ausführungen mehrerer Aufgaben oder Tätigkeiten zur gleichen Zeit geprägt. So führen wir z.B. Telefonate über unser Mobiltelefon während wir gehen, d.h. in Bewegung sind. Beim Autofahren beobachten wir die Straße und bedienen gleichzeitig Steuer und Pedale des Fahrzeugs. Offensichtlich ist der Mensch in der Lage, solche komplexen **Mehrfachaufgaben** zu bewältigen. Dies gilt allerdings nur für bestimmte Kombinationen von Aufgaben. Wenn wir z.B. versuchen mit der rechten und der linken Hand gleichzeitig unterschiedliche Texte zu Papier zu bringen, fällt uns das in der Regel sehr schwer und ist für die meisten von uns praktisch unmöglich. Mehrfachaufgaben spielen auch bei Benutzerschnittstellen eine wichtige Rolle. Ähnlich wie bei der geteilten Aufmerksamkeit (siehe Abschnitt 3.4) müssen wir bei Mehrfachaufgaben unseren **Ressourceneinsatz** (insbesondere motorische und Gedächtnisleistungen) auf mehrere Aufgaben verteilen. Dabei kommt es ganz darauf an, auf welches Wissen wir bei diesen Aufgaben zurückgreifen müssen. Handelt es sich um gefestigtes prozedurales Wissen (siehe Abschnitt 3.1.2), so kann dieses automatisch mit wenig Ressourceneinsatz abgerufen werden (z.B. eine natürliche Gehbewegung). Erfordert der Wissensabruf unsere Aufmerksamkeit, z.B. beim Lösen einer komplizierten Rechenaufgabe, muss der Ressourceneinsatz erhöht werden, um die Aufgabe erfolgreich zu beenden.

<sup>1</sup>Sollte dies in unserem Buch an der einen oder anderen Stelle passiert sein, so bitten wir vorsorglich um Verzeihung und verweisen auf die Beschränkungen des Seitenlayouts.





**Abbildung 3.1:** Die PRF einer leichten und einer schweren Aufgabe. Bei der leichten Aufgabe steigt die Performanz schnell an und flacht dann ab. Bei der schweren Aufgabe ist der initiale Anstieg der Performanz deutlich langsamer und kontinuierlicher (vgl. auch Wickens und Hollands [128, S. 441]).



Der Effekt des Ressourceneinsatzes in solchen Situationen kann mithilfe der **Durchführungs-Ressourcen-Funktion** beschrieben werden [87] (engl. **PRF**, oder **Performance Resource Function**). Diese Funktion setzt den Ressourceneinsatz in Bezug zur dadurch gewonnenen Verbesserung der Performanz der Aufgabe, bzw. dem Erfolg ihrer Durchführung. Je nach dem, wie gut eine Aufgabe vorher gelernt und wie stark sie automatisiert werden konnte, verläuft diese Kurve entweder steiler oder flacher. In der Regel besitzt die Kurve immer einen Punkt, ab dem weiterer Ressourceneinsatz zu keinem oder nur zu einem geringem weiteren Anstieg der Performanz führt. Dies wird in Abbildung 3.1 verdeutlicht. Bis Punkt zwei erfolgt ein schneller Anstieg der Performanz, der sich dann bis Punkt drei verlangsamt und schließlich stagniert. Es lohnt sich daher kaum noch, weitere Ressourcen in die Aufgabe zu investieren, auch wenn noch das theoretische Maximum der Performanz (Punkt eins) erreicht werden könnte. Die Eigenschaften der Kurve zu kennen ist insbesondere dann hilfreich, wenn parallel eine zweite Aufgabe durchgeführt werden muss, die ebenfalls Ressourcen benötigt. Die optimale Ressourcenverteilung zwischen mehreren Aufgaben zu bestimmen ist häufig sehr schwierig und wir verwenden Heuristiken, um die Verteilung abzuschätzen.

### 3.5.3 Messen der kognitiven Belastung

Die psychologische Untersuchung von Mehrfachaufgaben hat sich als eine Methode etabliert, die kognitive Belastung einer Aufgabe zu messen, indem diese parallel zu einer

zweiten Aufgabe ausgeführt wird. Die grundlegende Idee ist, neben der sogenannten **Primäraufgabe**, die noch zur Verfügung stehenden Ressourcen durch die Leistung bei der Bewältigung einer **Sekundäraufgabe** zu messen (vgl. dazu auch Ogden et al. [88]). Bei dieser Messung der kognitiven Belastung wird ein Benutzer gebeten, die Primäraufgabe so effektiv und so effizient wie möglich zu lösen und gleichzeitig die Sekundäraufgabe zu bearbeiten. Steigen nun die Anforderungen der Primäraufgabe graduell an, jedoch so dass der Benutzer die Aufgabe noch erfolgreich bearbeiten kann, erwartet man gleichzeitig ein Absinken der Leistung in der Sekundäraufgabe. Dieses Leistungsdefizit beschreibt die Zunahme der kognitiven Belastung durch die Primäraufgabe. Neben der Messung der Sekundäraufgabe existiert eine ganze Reihe von weiteren Methoden die kognitive Belastung zu messen, z.B. die Methode der **Selbsteinschätzung** der kognitiven Belastung. Naturgemäß ist eine solche Einschätzung sehr subjektiv und kann insofern nur einen groben Anhaltspunkt für die kognitive Belastung liefern. Etwas vergleichbarer wird die Messung jedoch durch die Verwendung etablierter Fragebögen wie des NASA Task Load Index oder **TLX** [19].

### 3.6 Entscheidungsfindung und -zeiten

Schon seit mehr als 150 Jahren beschäftigen sich kognitive Psychologen mit der Frage, wie menschliches Handeln und die dazugehörigen Entscheidungen getroffen werden. Im 19. Jahrhundert entdeckte Merkel [83], dass die Zeit, die zur Entscheidungsfindung benötigt wird, abhängig von der Anzahl der möglichen Alternativen ist. Unter der **Entscheidungszeit** versteht man die Zeit, die benötigt wird, um auf ein externes Ereignis mit einer Reaktion zu antworten, z.B. die Zeit, die der Autofahrer benötigt, um auf ein Aufleuchten der Bremsleuchten des vorausfahrenden Fahrzeugs mit der Betätigung der Bremse zu reagieren. Die Entscheidungszeit besteht in der Regel aus einem konstanten Teil, der unabhängig von der Anzahl der möglichen Alternativen ist, und einem variablen Teil, der sich mit der Anzahl der Alternativen verändert. Bei seinen Versuchen fand Merkel heraus, dass die Entscheidungszeit logarithmisch zur Anzahl der Alternativen wächst, d.h. jede zusätzliche Alternative verlängert die Entscheidungsfindung zwar, allerdings in einem geringeren Maße mit wachsender Anzahl der Alternativen.

Hick und Hyman untersuchten dieses Phänomen unter Berücksichtigung der Informationstheorie (vgl. Shannon und Weaver [102]) fast 100 Jahre später erneut [46, 53] und konnten zeigen, dass die Entscheidungszeit linear zum Informationsgehalt der Alternativenmenge ist. Wenn also  $N$  gleich wahrscheinliche Alternativen möglich sind, entspricht dies einem Informationsgehalt von  $\log_2(N)$  Bit. Man kann auch sagen, dass die Entscheidungszeit bei einer Verdopplung der Alternativen um einen konstanten Betrag steigt. Informationstheoretisch kann man dann äquivalent feststellen, dass sich die Entscheidungszeit um einen konstanten Betrag erhöht, sobald sich der Informationsgehalt der Alternativenmenge um ein Bit erhöht. Dies führt zu der mathematischen Formulierung des **Hick-Hyman Gesetzes**:

$$EZ = k + z * H_s = k + z * \log_2(N)$$

Hierbei bezeichnet  $EZ$  die Entscheidungszeit, die Konstante  $k$  repräsentiert den Teil der

Entscheidungszeit, der unabhängig von der Anzahl der Alternativen ist,  $z$  den konstanten Zuwachs der Entscheidungszeit pro zusätzlichem Bit an Information und  $H_s$  den durchschnittlichen Informationsgehalt der Alternativenmenge. Dieser lineare Zusammenhang gilt nicht nur, wenn alle Alternativen gleich wahrscheinlich sind (und damit informationstheoretisch keine Redundanz vorliegt), sondern auch im realistischeren Fall, in dem Alternativen unterschiedlich wahrscheinlich sind. In diesem Fall ist  $H_s$  geringer und trotzdem konnte in empirischen Versuchen nachgewiesen werden, dass die Entscheidungszeiten weiterhin durch das Hick-Hyman Gesetz vorhergesagbar sind. Analysiert man nun die Eingabe- und Ausgabemöglichkeiten von Benutzerschnittstellen in Bezug auf den Informationsgehalt (d.h. unter Berücksichtigung der Alternativen bei der Ein- und Ausgabe im informationstheoretischen Sinne), kann das Hick-Hyman Gesetz dazu verwendet werden, die Entscheidungszeit von Benutzern zu modellieren und beim Entwurf der Benutzerschnittstelle zu berücksichtigen oder zu optimieren.

Diese Überlegungen setzen allerdings alle voraus, dass dem Benutzer die verfügbaren Alternativen a priori bekannt sind. Ist dies nicht der Fall, beispielsweise in einem langen, unbekannten und unsortierten Menü, dann muss der Benutzer so lange die Alternativen lesen, bis er die gewünschte gefunden hat, also durchschnittlich die Hälfte der Einträge. Die Zeit steigt in diesem Fall linear mit der Anzahl der Alternativen. Sortiert man die Liste beispielsweise alphabetisch (wie in einem Adressbuch oder im Index dieses Buchs), dann kann der Benutzer eine Suchstrategie wie z.B. Intervallteilung anwenden, und die Zeit wächst nur noch logarithmisch.



### Übungsaufgaben:

1. Überlegen Sie sich, wie deklaratives und prozedurales Wissen bei der Mensch-Maschine-Interaktion zum Einsatz kommen. Nennen Sie jeweils zwei Beispiele und diskutieren Sie anhand dieser die Vor- und Nachteile beider Wissensarten.
2. Entwerfen Sie ein kleines Experiment, um ihr Kurzzeitgedächtnis zu testen. Führen Sie das Experiment mit Freunden und Bekannten durch und vergleichen sie die Ergebnisse miteinander. Diskutieren Sie Ihre Erkenntnisse unter Berücksichtigung der Aussagen, die zum Kurzzeitgedächtnis in Abschnitt 3.1.1 gemacht wurden.
3. Sie sind für die Gestaltung einer Benutzeroberfläche mit 50 Alternativen verantwortlich und überlegen, ob sie 50 Schaltflächen gleichzeitig darstellen sollten (flache Struktur), die Alternativen in fünf Menüs als Zehnergruppen zugänglich machen (tiefere Menüstruktur), oder fünf Menüs mit je fünf Untermenüs zu je 2 Möglichkeiten vorsehen (ganz tiefe Menüstruktur). Hilft ihnen das Hick-Hyman Gesetz aus Abschnitt 3.6, um zwischen diesen Designalternativen zu entscheiden? Was fällt Ihnen auf? Erläutern Sie die grundsätzliche Aussage des Gesetzes zur Verwendung von Menüstrukturen. Welche anderen Faktoren haben hier einen Einfluss auf die Ausführungszeit?
4. Skizzieren sie die in Abschnitt 3.5.2 beschriebene *Durchführungs-Ressourcen-Funktion* für eine schwere und eine einfache Aufgabe Ihrer Wahl. Erläutern Sie den jeweiligen Kurvenverlauf und schätzen Sie die Lage der Punkte zwei und drei (vergleiche Abbildung 3.1) in den von Ihnen gefundenen Kurvenverläufen.

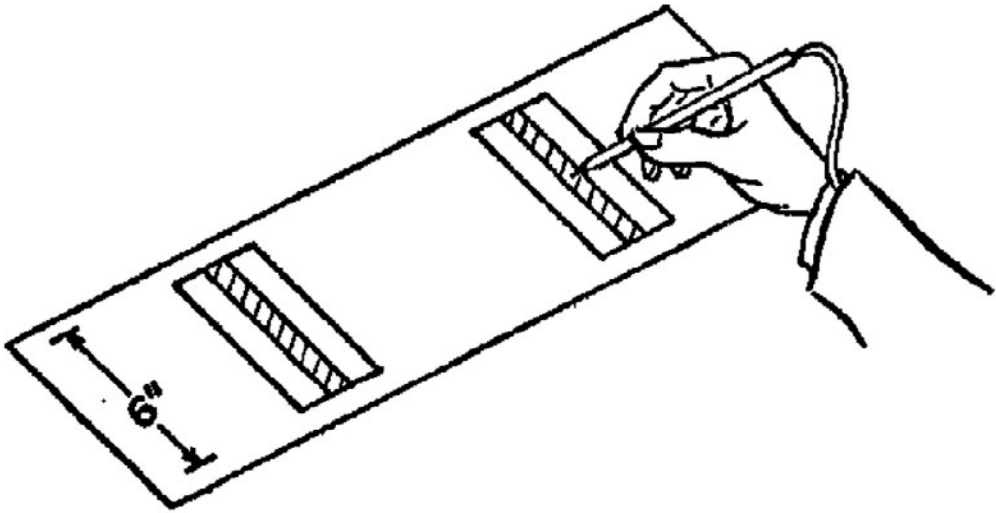
## 4 Motorik

Nach der menschlichen Wahrnehmung und Informationsverarbeitung wollen wir in diesem Kapitel die menschliche **Motorik** betrachten, da sie unser wichtigster Kanal zur Steuerung interaktiver Systeme ist. Das Wort Motorik leitet sich vom lateinischen *motor* = Beweger ab, und bezeichnet zunächst alle Bewegungen des Menschen, die dieser mithilfe seiner Muskulatur ausführt. Zur Interaktion mit dem Computer verwenden wir dabei vor allem die Muskulatur der Finger, Hände und Arme, um entweder Eingaben auf **Touchscreens** durch Berührung vorzunehmen oder **Eingabegeräte** wie Maus, Trackball oder Tastatur zu bedienen. In anderen Fällen verwenden wir auch die gesamte Körpermuskulatur, beispielsweise bei der Bewegung in VR-Umgebungen wie dem **CAVE** oder bei der Steuerung von Anwendungen oder Spielen durch **Gesteninteraktion**, wie an neueren Spielekonsolen. Wir beschränken uns in diesem Kapitel jedoch auf die motorischen Aspekte der Interaktion mit einer oder zwei Händen.

Dabei sind die motorischen Bewegungen des Körpers meistens durch eine Regelschleife gesteuert: Ein Wahrnehmungskanal (z.B. Sehsinn oder Propriozeption, siehe Kapitel 2) misst dabei die tatsächlich ausgeführte Bewegung, wodurch diese exakt gesteuert werden kann. Durch die Koordination von Hand und Auge werden wir beispielsweise in die Lage versetzt, kleine Bewegungen räumlich exakt auszuführen. Grundsätzlich gibt es in der Motorik immer eine Wechselbeziehung zwischen Geschwindigkeit und Genauigkeit: Je schneller wir eine Bewegung ausführen, desto ungenauer wird sie, und je genauer wir sie ausführen, desto langsamer wird sie normalerweise. Diese Wechselbeziehung hat ganz konkrete Auswirkungen auf die Gestaltung grafischer Benutzerschnittstellen, wie wir im kommenden Abschnitt sehen werden. Sie wird außerdem beeinflusst durch individuelle Unterschiede zwischen Benutzern, wie Alter, Übung, Fitness, oder Gesundheitszustand.

### 4.1 Fitts' Law

Wenn wir eine Benutzerschnittstelle entwerfen, bei der wir Dinge durch Zeigen auswählen müssen, beispielsweise mit einem Stift auf einem interaktiven Tisch, dann interessiert uns normalerweise, wie wir diese Interaktion so schnell wie möglich gestalten können. Es wäre interessant, den offensichtlichen Zusammenhang *weiter* = *langsamer* genauer fassen zu können, um daraus auch Vorhersagen über die Geschwindigkeit ableiten zu können. Gleiches gilt für die Bedienung mit Zeigegeräten wie **Maus**, **Trackball**, **Touchpad** oder **Lichtgriffel**. Bereits in den 1950er Jahren führte der amerikanische Psychologe **Paul Fitts** Experimente zur menschlichen Motorik durch, aus denen er ein Gesetz ableitete, das für eine eindimensionale Bewegung die benötigte Zeit in Abhängigkeit vom zurückgelegten Weg und der Größe des zu treffenden Ziels beschreibt [34]. Abbildung 4.1 zeigt seinen Versuchsaufbau. Das gefundene Gesetz wurde durch ande-



mmibuch.de/a/4.1

**Abbildung 4.1:** Der ursprüngliche Versuchsaufbau zur Ermittlung von Fitts' Law [34]: Ziel war es, abwechselnd die im Bild schraffierten Kontaktplatten zu treffen, ohne dabei die benachbarten weißen Platten zu berühren.

re Kollegen und in nachfolgenden Publikationen mehrfach modifiziert<sup>1</sup> und hat unter dem Namen **Fitts' Law** breite Akzeptanz im Bereich der Mensch-Maschine Interaktion gefunden. Durch dieses Gesetz sind auch viele Zeige- und Menütechniken inspiriert, die heute in Desktop-Betriebssystemen zu finden sind (siehe Kapitel 15). Die heute am häufigsten verwendete Formel geht auf MacKenzie [78] zurück und berechnet die Bewegungszeit (engl. *movement time*, MT) wie folgt:

$$MT = a + b * ID = a + b * \log_2\left(\frac{D}{W} + 1\right)$$

Dabei entspricht  $D$  der Distanz vom Startpunkt bis zur Mitte des zu treffenden Ziels und  $W$  der Breite des Ziels *in Bewegungsrichtung*. Der logarithmische Term wird insgesamt auch als **Index of Difficulty** ( $ID$ ) bezeichnet und die Konstante  $+1$  in der Klammer garantiert, dass  $ID$  nur positive Werte annehmen kann. Die Konstanten  $a$  und  $b$  variieren von Situation zu Situation, wobei  $a$  so etwas wie eine Reaktionszeit beschreibt, und  $b$  einen allgemeinen Faktor für die Geschwindigkeit, der beispielsweise mit der Skalierung der Bewegung zusammenhängt. Das Experiment wurde mit einem Stift auf einer Tischplatte ausgeführt und kann daher sofort auf unser Beispiel vom Anfang des Abschnitts angewendet werden. Aus der Formel können wir lesen, dass die Zeit, um ein bestimmtes Ziel anzuwählen, logarithmisch mit dem Abstand  $D$  vom Startpunkt zum Ziel wächst. Genauso sinkt sie umgekehrt logarithmisch mit der Breite  $W$

<sup>1</sup>Für eine Diskussion der verschiedenen Formeln siehe Drewes [27]

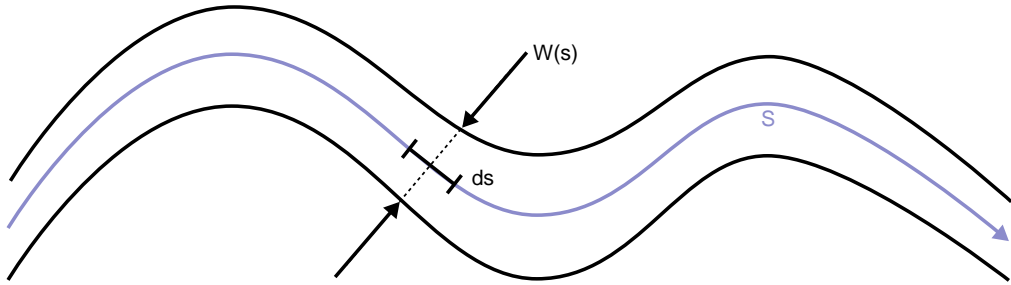
des Ziels. Diese wird wohlgemerkt *in Bewegungsrichtung* betrachtet. Ein Button, der breiter als hoch ist, ist also prinzipiell horizontal leichter zu treffen als vertikal. Das gilt allerdings nur, solange er nicht zu schmal wird, denn dann greift irgendwann das **Steering Law** (siehe Abschnitt 4.2). Bei der Umsetzung einer Benutzerschnittstelle ist es daher immer sinnvoll, Ziele möglichst groß zu machen, wenn sie schnell ausgewählt werden sollen. Bei einem grafischen **Button** oder Listeneintrag sollte also nicht nur der Text der Beschriftung sensitiv sein, sondern die gesamte Fläche des Buttons oder Eintrags, gegebenenfalls sogar noch ein Gebiet um den Button herum, sofern dies nicht bereits näher an einem benachbarten Button liegt. Insbesondere auf Webseiten wird diese Grundregel immer wieder verletzt, da es in HTML eben einfacher ist, nur den Text selbst sensitiv zu machen, als die gesamte umgebende Fläche.

Darüber hinaus gibt es insbesondere bei der Bedienung mit der **Maus** bevorzugte Regionen des Bildschirms an dessen Rand, die auch in Kapitel 15 nochmals angesprochen werden: Bewegt man die Maus beispielsweise nach oben bis zum Bildschirmrand und darüber hinaus, dann bleibt der Mauszeiger in vertikaler Richtung am Bildschirmrand hängen. Bewegt man sie in eine Ecke, dann bleibt er in beiden Richtungen hängen. Da es also unerheblich ist, wie weit man die Maus danach noch weiter bewegt, wird die Variable  $W$  in der Formel für Fitts' Law beliebig groß. Demnach lassen sich verschiedene Klassen unterschiedlich bevorzugter Bildschirmpixel finden, was in den Übungsaufgaben zu diesem Kapitel vertieft wird.

Betrachtet man die Bewegungen mit Arm oder Zeigegerät noch genauer, so kann man sie in verschiedene Phasen und verschiedene Situationen unterteilen und damit eine noch genauere Vorhersage erreichen. Außerdem wurden bestimmte Typen von Zeigebewegungen identifiziert, für die Fitts' Law – obwohl prinzipiell anwendbar – nicht die bestmögliche Vorhersage liefert. Diese Betrachtungen gehen jedoch über den Umfang eines einführenden Buches hinaus, und Fitts' Law kann in den allermeisten Fällen als angemessene Beschreibung für Zeigebewegungen verwendet werden. Als Einstiegspunkt bei weiterem Interesse seien beispielsweise das kritische Paper von Drewes [27] sowie der Übersichtsartikel von Soukoreff und MacKenzie [107] empfohlen.

## 4.2 Steering Law

Das ursprüngliche Experiment zu Fitts' Law betrachtete eindimensionale Bewegungen, und obwohl der Bildschirm einen zweidimensionalen Interaktionsraum darstellt, sind direkte Bewegungen vom Startpunkt zum Ziel einer Zeigebewegung ihrer Natur nach eindimensional, da sie grob gesehen einer geraden Linie folgen. Damit gilt für sie **Fitts' Law**. Wenn der Weg, den wir mit dem Zeigegerät verfolgen müssen, nun nicht mehr beliebig ist, sondern entlang eines bestimmten Pfades verlaufen muss und diesen nicht weiter als bis zu einem bestimmten Abstand verlassen darf, dann gelten andere Gesetze. Ein praktisches Beispiel hierfür ist die Auswahl in einem geschachtelten Menü (siehe Abbildung 15.3 auf Seite 161) oder die Steuerung einer Spielfigur in **Sidescroller**-Spielen wie **Super Mario Bros.** oder **Defender**. Das gleiche Problem tritt aber auch im Alltag auf, wenn beispielsweise ein Auto ohne Kollision durch einen Tunnel gelenkt werden muss oder ein Skirennfahrer beim Slalom keines der Tore auslassen darf.



**Abbildung 4.2:** Verfolgen eines gebogenen Pfades  $S$ : die Geschwindigkeit und damit die benötigte Zeit hängt von der Breite  $W(s)$  ab [1].

Diese Situation ist in Abbildung 4.2 dargestellt. Da sie etwa der Aufgabe entspricht, ein Auto über eine Straße mit variabler Breite zu steuern, wird das dafür gefundene Gesetz als **Steering Law** bezeichnet. Accot und Zhai [1] leiten dabei für die gesamte Zeit  $T$  zum Zurücklegen des Pfades  $S$  folgende Formel her:

$$T = a + b * \int_S \frac{1}{W(s)} ds$$

Dies ist die allgemeinste Formulierung des Gesetzes und besagt, dass entlang des Pfades  $S$  über den Kehrwert der zulässigen Breite integriert wird. Die Variablen  $a$  und  $b$  sind genau wie bei Fitts' Law von der Situation abhängig und haben die gleiche Funktion.

In praktischen Situationen kann der Pfad oft in wenige Segmente mit konstanter Breite zerlegt werden. Betrachten wir Abbildung 15.3 auf Seite 161, dann sind dies drei Segmente: Das erste Segment erstreckt sich vom Apfelsymbol bis zum Erreichen des Menü-Eintrags erster Ebene. Dieses Segment muss keinem bestimmten Pfad folgen und wird meist recht vertikal ausgeführt: Es gilt Fitts' Law. Das zweite Segment verläuft entlang des gefundenen Menüeintrags bis zum Erreichen des Untermenüs. Dabei darf der erste Menüeintrag nicht verlassen werden und es gilt das Steering Law. Die Breite  $W(s)$  ist hier aber konstant und das Integral lässt sich einfach berechnen als Produkt aus dem horizontal zurückgelegten Weg und dem Kehrwert der zulässigen Breite, also etwa (Breite des Menüs / Höhe eines Eintrags). Für das dritte Segment im Untermenü bis zum dort ausgewählten Eintrag gilt wieder Fitts' Law wie im ersten Segment.

Auf diesem Wege lassen sich die Interaktionszeiten für geschachtelte Menüs abschätzen und als Regel ergibt sich dabei, Menüs nicht unnötig breit zu machen, sowie die Einträge so hoch wie möglich. Bei Computerspielen kommt das **Steering Law** auch umgekehrt zum Tragen: Kann man die Geschwindigkeit selbst bestimmen, wie beispielsweise bei einem Autorennen, dann ergibt sich (bei unfallfreier Fahrt) die benötigte Zeit aus der Enge des Kurses und der persönlichen Spielstärke, die in der Konstanten  $b$  zum Ausdruck kommt. Ist die Geschwindigkeit, wie in manchen **Sidescroller**-Spielen fest vorgegeben,

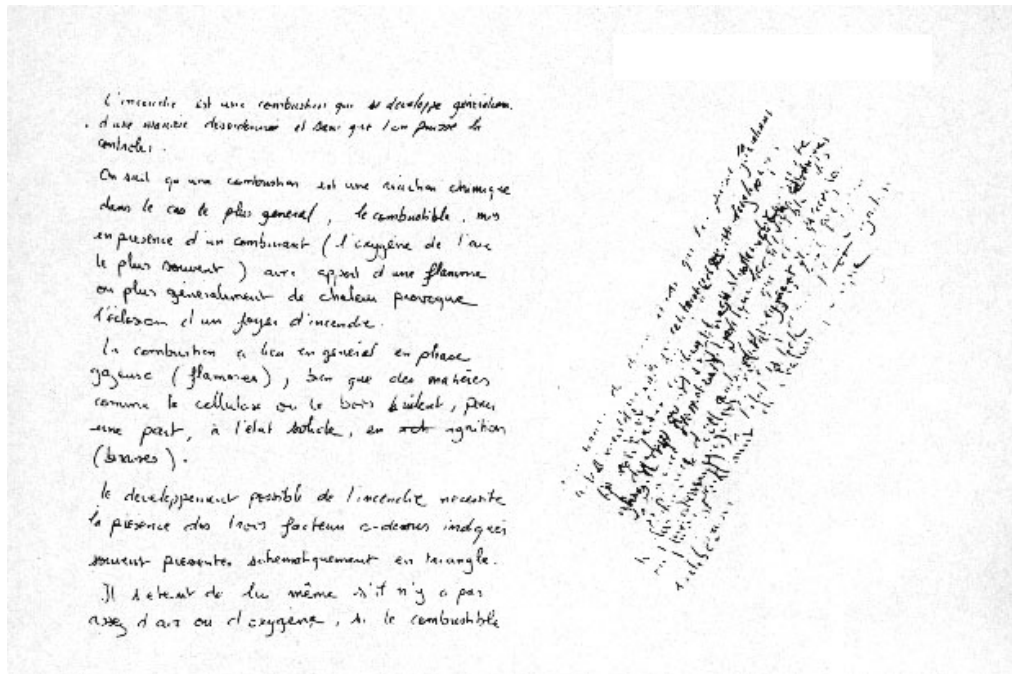


dann ergibt sich aus ihr und der persönlichen Konstante  $b$  eine minimal benötigte Breite  $W(s)$  und sobald diese nicht mehr zur Verfügung steht, kommt es zur Kollision. Die wachsende Übung im Spiel drückt sich dabei in einem kleineren Wert für  $b$  aus.

Auch in Benutzerschnittstellen kommt das Gesetz auf verschiedene Art zum Tragen: Neben den ausführlich diskutierten geschachtelten Menüs, kann es allgemein auf das Zeichnen von Kurven angewendet werden, beispielsweise beim Freistellen von Umrissen in einer Bildverarbeitung: Je genauer dies getan wird, desto länger dauert es, und zwar wächst die benötigte Zeit gerade mit dem Kehrwert der zulässigen Abweichung. **Fitts' Law** und das **Steering Law** decken zusammen einen breiten Bereich der Interaktion in grafischen Benutzerschnittstellen ab, solange mit einem einzelnen Punkt als Eingabe gearbeitet wird. Werden zur Interaktion beide Hände verwendet, so stellt sich die Situation wiederum anders dar.

## 4.3 Interaktion mit beiden Händen

In den 1980er Jahren beschrieb **Yves Guiard** die Arbeitsteilung zwischen beiden Händen bei Aktionen, die beidhändig ausgeführt werden und stellte ein formales Modell zur Beschreibung solcher Aktionen auf [40]. Dabei beschreibt er die Gliedmaßen als abstrakte Motoren, die an einem Ende in ihrem jeweiligen Referenzsystem aufgehängt sind (z.B. der Arm an der Schulter) und ihr anderes Ende durch eine Bewegung steuern. Damit lassen sich koordinierte Bewegungen mehrerer Gliedmaßen formal beschreiben. Eine Aussage seines Artikels ist, dass wir Aktionen im Raum eher selten mit zwei gleichberechtigten Händen ausführen, sondern dass sich die Hände die Arbeit meist asymmetrisch teilen: Die nicht-dominante Hand (beim Rechtshänder also die Linke) legt dabei das Referenzsystem fest, in dem die dominante Hand eine Aktion ausführt. Die beiden Arme dazwischen bilden eine motorische Kette. Diese Regel leitet Guiard u.a. aus folgendem Experiment ab (Abbildung 4.3): Eine Testperson schreibt einen handschriftlichen Brief auf einem etwa DIN A4 großen Stück Papier. Auf der Tischoberfläche darunter liegt (von der Testperson unbemerkt) zunächst eine Schicht Kohlepapier, darunter ein weiteres weißes Stück Papier. Während der Brief die gesamte Seite mit Text ausfüllt, ist die Tischfläche, die tatsächlich zum Schreiben benutzt wird, wesentlich kleiner und die Schrift verläuft insbesondere auch nicht waagrecht. Die nicht-dominante (hier: linke) Hand hält das Briefpapier in einer Position, die es für die dominante (hier: rechte) Hand angenehm macht, entlang ihrer natürlichen Bewegungsrichtung (Drehung um Schulter bzw. Ellbogen) und in einem angenehmen Winkel zu schreiben. Nach ein paar Zeilen wird der Brief hochgeschoben und die gleiche Tischfläche wieder benutzt. So schafft die nicht-dominante Hand das (grobe) Referenzsystem für die (feinere) Schreibbewegung der dominanten Hand. Eine ähnliche Arbeitsteilung zwischen beiden Händen ist bei vielen Alltagstätigkeiten zu beobachten: beim Nähen oder Sticken fixiert die Linke den Stoff während die rechte die Nadel führt. Beim Essen fixiert die Linke mit der Gabel das Schnitzel, während die Rechte mit dem Messer ein Stück davon abschneidet. Beim Gitarren- oder Geigenspiel drücken die Finger der Linken Hand die Saiten um Tonhöhe oder Akkord festzulegen, während die Rechte sie zupft oder streicht, um zum richtigen Zeitpunkt einen Ton zu erzeugen. Beim Malen hält die Linke die Palette, während die Rechte den Pinsel führt, der darauf Farben aufnimmt oder mischt.



mmibuch.de/a/4.3

**Abbildung 4.3:** Links: handgeschriebener Brief, Rechts: überlagerte Schriftabdrücke auf darunter liegendem Kohlepapier, aus Guiard [40]

Für die menschliche Interaktion mit dem Computer bedeutet dies, dass wir die menschliche Fähigkeit zu beidhändiger Interaktion am besten unterstützen, indem wir eine ähnliche Arbeitsteilung vorsehen. Dies passiert automatisch beispielsweise bei der Bedienung mobiler Geräte, die meist links gehalten und mit der Rechten bedient werden. Bei der Mausbedienung haben wir gelernt, mit der linken Hand z.B. eine Taste zu drücken und damit in einen anderen Interaktionsmodus zu gelangen, während die Rechte den Mauszeiger und die Maustasten bedient (z.B. Shift-Click). Interessant wird die asymmetrische Arbeitsteilung jedoch insbesondere bei Benutzerschnittstellen für interaktive Oberflächen, die mit grafischen oder physikalischen Objekten gesteuert werden. Beispiele für solche Schnittstellen und eine genauere Diskussion der Auswirkungen finden sich in Kapitel 17.

## Übungsaufgaben:



1. Nach Fitts' Law ist ein Bildschirmpixel umso schneller zu erreichen, je näher es zur aktuellen Mausposition liegt. Darüber hinaus gibt es besondere Pixel, die aus anderen Gründen schnell erreichbar sind. Genauer gesagt kann man auf einem Bildschirm vier Klassen von Pixeln danach unterscheiden, wie schnell und einfach sie erreichbar sind. Welches sind diese vier Klassen, und warum?
2. Drei dieser Klassen werden in aktuellen Desktop-Interfaces auch auf besondere Weise genutzt. Welche sind das und worin besteht die besondere Nutzung?
3. Berechnen Sie die Geschwindigkeit der Mauszeigerbewegung für die in Abbildung 15.3 auf Seite 161 gezeigte Interaktion mit Hilfe der in diesem Kapitel vorgestellten Gesetze. Schätzen Sie die Breite und Höhe der Menüeinträge in der Abbildung und nehmen Sie an, dass  $a = 600ms$  und  $b = 150 \frac{ms}{bit}$  betragen.



## 5 Mentale Modelle und Fehler

Wenn wir mit einem Computer oder einem anderen Gerät oder System interagieren, machen wir uns ein gewisses Bild von dessen Funktionsweise. Wir interpretieren die Ausgaben des Systems und seine Reaktionen auf unsere Eingaben und wir versuchen, Vorhersagen darüber zu machen, welche Eingabe zu welcher Reaktion führen wird. Ein System, das solche Vorhersagen erfüllt und sich so verhält, wie wir es erwarten, erscheint uns logisch, schlüssig und daher einfach zu bedienen. Verhält sich das System anders als wir es erwarten, so erscheint es uns unlogisch und verwirrend. Dies kann entweder daran liegen, dass das System objektiv unlogisch reagiert, oder aber dass unsere Vorhersagen auf einer falschen angenommenen Funktionsweise beruhen.

Die Kognitionswissenschaft und Psychologie beschäftigen sich unter anderem damit, wie wir unsere Umwelt und die darin enthaltenen Dinge und Vorgänge verstehen und verarbeiten. Eine derzeit gängige Annahme ist, dass wir Menschen sogenannte **mentale Modelle** unserer Umgebung konstruieren, die es erlauben, Vorgänge zu erklären und Vorhersagen zu machen. Der Begriff des mentalen Modells wurde durch **Kenneth Craik** [25] 1943 eingeführt und durch **Don Norman** [86, 85] ausgiebig auf den Bereich der Mensch-Maschine-Interaktion angewendet. Das Konzept der mentalen Modelle hilft uns dabei, das Verständnis des Nutzers von einem System zu beschreiben und zu analysieren, wo Fehler vorliegen, bzw. wie ein gutes Interaktionskonzept aussehen könnte.

### 5.1 Verschiedene Modellarten

Das **mentale Modell** ist ein nur in unserer Vorstellung existierendes Abbild eines realen Objektes, Systems oder Vorgangs. Nehmen wir ein einfaches Beispiel: Wenn wir auf einem Kinderspielfeld beobachten, dass zwei verschieden schwere Kinder auf den beiden Seiten einer Wippe jeweils am Ende sitzen, dann ist uns klar, warum die Wippe nicht im Gleichgewicht ist: beide Seiten der Wippe sind gleich lang, die Gewichte aber verschieden. Zum Wippen müssen beide Seiten jedoch im Gleichgewicht sein. Um die Wippe ins Gleichgewicht zu bringen, können wir entweder dem leichteren Kind helfen, indem wir seinen Sitz herunterdrücken (also das Gewicht künstlich erhöhen), oder das schwerere Kind weiter nach innen, näher zum Drehpunkt setzen (also den Hebel verkürzen). Diese Vorhersagen werden durch unser mentales Modell von der Funktion einer Wippe ermöglicht.

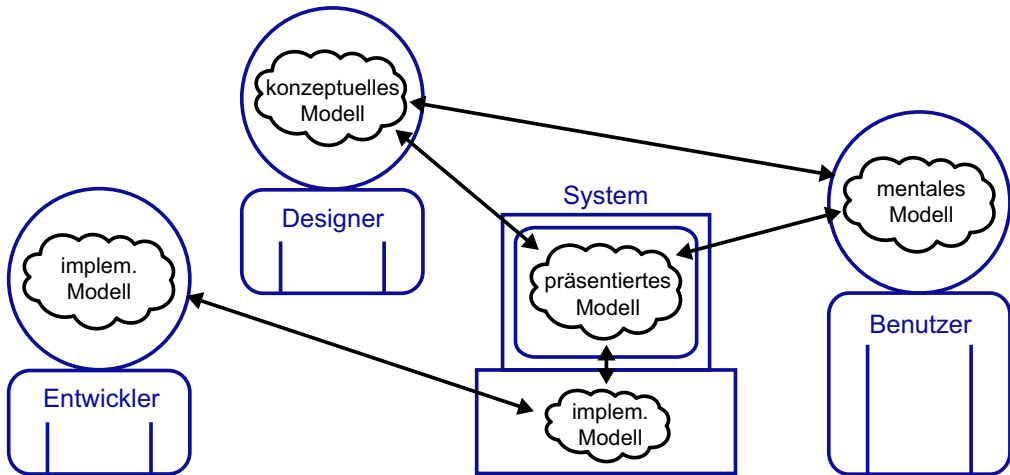
Vielleicht haben wir in der Schule im Physikunterricht auch Genaueres über Schwerkraft und Hebelgesetz gelernt (und auch behalten). Dann wissen wir, dass das jeweils wirkende Drehmoment das Produkt aus der Gewichtskraft und der Hebellänge ist. Ist ein Kind also doppelt so schwer wie das andere, dann müssen wir den Hebelarm halbieren und das Kind in die Mitte zwischen Sitz und Drehpunkt setzen. In diesem Fall ist

unser mentales Modell also detaillierter und erlaubt genauere Vorhersagen. Normalerweise reicht uns jedoch auf dem Spielplatz das ungenauere Modell, da wir die exakten Gewichte fremder Kinder normalerweise nicht wissen, und die Position auf der Schaukel einfacher durch Ausprobieren ermitteln können. Grundsätzlich lässt sich sagen, dass das einfachste mentale Modell, welches das beobachtete Verhalten vollständig erklärt, meist das beste ist. Dies gilt beispielsweise auch für die Newtonsche Physik und die Relativitätstheorie: im Schulunterricht reicht die erstere.

Ein weiteres Beispiel: Unser mentales Modell vom Autofahren besagt, dass wir durch Herunterdrücken des Gaspedals die Geschwindigkeit des Autos erhöhen. Welche genauen physikalischen Vorgänge dazwischen ablaufen ist für die erfolgreiche Bedienung des Gaspedals unerheblich und unser sehr stark vereinfachtes mentales Modell ist völlig ausreichend, solange sich das Auto entsprechend verhält. Ein detailliertes Modell der tatsächlichen Funktionsweise eines Gerätes oder Systems ist das **implementierte Modell**. Im Beispiel der Wippe ist dieses implementierte Modell sehr nahe an unserem mentalen Modell. Im Beispiel des Autos sind sie sehr verschieden. Insbesondere ist es für die Bedienung des Gaspedals auch unerheblich, ob das Fahrzeug durch einen Elektro- oder Verbrennungsmotor angetrieben wird, ob es Räder oder Ketten hat, und wie viele davon. Völlig unterschiedliche implementierte Modelle können also dasselbe mentale Modell bedienen.

Die Art und Weise, wie sich die Funktion eines Systems dem Nutzer darstellt, ist das **präsentierte Modell**. Dieses präsentierte Modell kann vom implementierten Modell wiederum vollständig abweichen oder es kann auch nur eine leichte Vereinfachung davon sein wie im Beispiel der Wippe. Nehmen wir ein anderes Alltagsbeispiel: Beim Festnetz-Telefon haben wir gelernt, dass wir eine Telefonnummer wählen, um eine direkte elektrische Verbindung zu einem anderen Telefon herzustellen. Früher wurden diese Verbindungen tatsächlich in Vermittlungsstellen physikalisch hergestellt. Bei den heutigen digitalen Telefonnetzen werden sie nur noch logisch in Form von Datenströmen hergestellt. Beim Mobiltelefon sind die zugrunde liegenden technischen Abläufe nochmals völlig anders: jedes Telefon bucht sich in eine bestimmte Zelle eines Mobilfunknetzes ein und Datenpakete werden über verschiedene Ebenen eines Netzwerkes zur richtigen Zelle der Gegenstelle weitergeleitet. Trotzdem ist das präsentierte Modell in allen drei Fällen das gleiche: Das Wählen einer Telefonnummer führt dazu, dass eine direkte Verbindung zwischen zwei Telefonen hergestellt wird. Es werden also drei völlig verschiedene implementierte Modelle durch das gleiche präsentierte Modell dargestellt, um das etablierte mentale Modell des Benutzers zu bedienen, so dass dieser kein neues Modell lernen muss.

Wenn Designer ein Gerät oder System entwerfen, dann haben auch sie ein mentales Modell von dessen Bedienung im Kopf. Dieses Modell ist das **konzeptuelle Modell**. Der Entwickler, der das System letztendlich baut, entwickelt das implementierte Modell, das die Funktion auf technischer Ebene umsetzt. Das Zusammenspiel der verschiedenen Modelle ist in Bild 5.1 dargestellt.



**Abbildung 5.1:** Das konzeptuelle Modell im Kopf des Designers und das mentale Modell im Kopf des Benutzers sollten möglichst übereinstimmen. Dazu muss das präsentierte Modell das konzeptuelle bestmöglich vermitteln. Vom implementierten Modell dürfen alle drei beliebig abweichen.



## 5.2 Zusammenspiel der Modelle

Die genannten Modelle interagieren auf verschiedene Weise: Das konzeptuelle Modell des Designers spiegelt sich im präsentierten Modell des Systems wieder. Das präsentierte Modell erzeugt beim Benutzer das mentale Modell. Das präsentierte Modell kann mit dem implementierten Modell identisch sein, muss es aber nicht. Der Entwickler setzt sein Implementierungsmodell durch Bau oder Programmierung um und sorgt so dafür, dass das System nach diesem Modell auf technischer Ebene funktioniert.

### 5.2.1 Transparenz

Eine wünschenswerte Eigenschaft eines Systems ist, dass das konzeptuelle Modell des Designers möglichst identisch mit dem mentalen Modell des Benutzers ist. Dies ist der Fall, wenn das präsentierte Modell das konzeptuelle Modell so gut verständlich macht, dass es der Benutzer vollständig rekonstruieren kann. Man spricht in diesem Fall auch von **Transparenz** in der Benutzerschnittstelle, denn der Benutzer muss dann keine Aspekte des implementierten Modells berücksichtigen, also z.B. Bedienschritte ausführen, die nur der Implementierung geschuldet sind. Stattdessen kann er sich komplett dem eigentlichen Zweck der Interaktion, der Aufgabe selbst widmen.

Leider wird die vollständige Übereinstimmung von konzeptuellem und mentalem Modell oft nicht erreicht, da das präsentierte Modell entweder nicht alle Aspekte wiedergibt und falsch oder nicht vollständig verstanden wird. Der Entwurf eines guten Interaktionskonzeptes beginnt in der Regel mit einer Analyse, welche mentalen Modelle beim Benutzer

bereits existieren oder leicht verstanden werden könnten. Ein darauf basierendes konzeptuelles Modell – angemessen präsentiert – hat gute Chancen, auch vollständig und richtig verstanden zu werden und somit Transparenz herzustellen. Ein Beispiel hierfür ist das vorläufige Löschen einer Datei indem man sie in den Papierkorb bewegt, und das endgültige Löschen, indem der Papierkorb geleert wird. Dieses konzeptuelle Modell existiert identisch in der physikalischen Welt. In Kapitel 10 werden wir uns eingehender mit der Rolle des konzeptuellen Modells im Designprozess von Benutzerschnittstellen beschäftigen.

Hat man andererseits völlige gestalterische Freiheit beim Entwurf des konzeptuellen Modells, weil noch kein passendes mentales Modell der Domäne beim Benutzer existiert, so bietet es sich an, nicht beliebig weit vom implementierten Modell abzuweichen, sondern das konzeptuelle Modell so zu entwerfen, dass es korrekte Schlüsse über die Funktion auf technischer Ebene erlaubt, um damit möglichen späteren Inkonsistenzen mit der technischen Funktionsweise aus dem Weg zu gehen.

Ein klassisches Beispiel für eine solche Inkonsistenz ist die Tatsache in derzeitigen Betriebssystemen für Personal Computer, dass das Bewegen eines Dokumentes in einen anderen Ordner dieses verschiebt, wenn sich beide Ordner auf demselben Laufwerk befinden, aber das Dokument kopiert wird, wenn sich die Ordner auf verschiedenen Laufwerken befinden. In der physikalischen Welt gibt es zwar Dokumente und Ordner, aber keine Entsprechung zu Laufwerken. Der Benutzer muss das Konzept eines Laufwerks aus dem implementierten Modell verstehen, um korrektes Verhalten vorherzusagen. Die Transparenz ist gestört.

### 5.2.2 Flexibilität

Eine weitere wünschenswerte Eigenschaft einer Benutzerschnittstelle ist **Flexibilität**. Dies bedeutet, dass das gleiche Ergebnis innerhalb der Benutzerschnittstelle auf verschiedenen Wegen zu erreichen ist. Ein einfaches Beispiel sind **Tastaturkürzel** bei Desktop-Interfaces: Zum Kopieren einer Datei kann man deren Icon mit der Maus auswählen und dann entweder a) im Menü die Funktionen *Kopieren* und *Einfügen* mit der Maus auswählen, oder b) auf der Tastatur die Kürzel *Ctrl-C* und *Ctrl-V* eingeben.

Flexibilität in einer Benutzerschnittstelle bewirkt, dass verschiedene Arten von Benutzern jeweils die ihnen geläufigste oder zugänglichste Art der Interaktion verwenden können. Im Fall der Tastaturkürzel wird ein Computerneuling oder jemand, der selten den Computer benutzt vermutlich eher den Weg über das Menü wählen, da er sich so keine Tastaturkürzel merken muss, sondern eine sichtbare Funktion auswählen kann (*recognition rather than recall*, siehe auch Abschnitt 3.1.2 und Kapitel 7). Ein fortgeschrittener Benutzer wird sich jedoch leicht die beiden Tastenkombinationen einprägen, wenn er sie häufig benutzt, und so bei jedem Kopiervorgang etwas Zeit sparen, da die Tastaturkürzel schneller ablaufen als das Auswählen der Menüfunktionen. Der fortgeschrittene Benutzer tauscht also Gedächtnisleistung gegen Interaktionsgeschwindigkeit.

Die Tatsache, dass beide Wege in der Benutzerschnittstelle gleichzeitig existieren und die Tastaturkürzel auch hinter dem Menüeintrag vermerkt sind, bewirkt nebenbei eine gute **Erlernbarkeit** der Benutzerschnittstelle: Der Anfänger kann die Kopierfunktion im Menü immer wieder finden und muss sich zunächst nichts merken. Wenn er aber oft ge-



nug den Menüeintrag mit dem zugehörigen Kürzel gesehen hat, wird er sich irgendwann das Kürzel merken und kann beim nächsten Mal den schnelleren Weg nutzen. Durch Flexibilität in der Benutzerschnittstelle ist es sogar möglich, verschiedene konzeptuelle (Teil-) Modelle im gleichen System zu unterstützen. In Desktop-Benutzerschnittstellen gibt es den sogenannten **Scroll-Balken**, eine Leiste am Rand eines langen Dokumentes, mit der man das Dokument auf- und abwärts schieben und verschiedene Ausschnitte betrachten kann. Dabei entspricht die Länge der gesamten Leiste der gesamten Dokumentlänge und die Länge und Position des Reiters zeigt die Länge und Position des aktuell sichtbaren Ausschnitts. Greift man den Reiter nun mit der Maus und bewegt diese nach unten, so verschiebt sich das gedachte Sichtfenster nach unten und das Dokument selbst wandert am Bildschirm scheinbar nach oben. Das zugrunde liegende mentale Modell ist das eines beweglichen Sichtfensters, das über dem Dokument auf und ab geschoben wird. Ein alternativ mögliches mentales Modell ist ein fest stehendes Fenster, unter dem das Dokument selbst auf und ab geschoben wird. Dieses Modell wird oft durch eine symbolische Hand dargestellt, mit der man das Dokument mit der Maus anfassen und verschieben kann. In diesem Fall bewirkt eine Mausbewegung nach unten auch eine Verschiebung des Dokumentes auf dem Bildschirm nach unten, also genau entgegengesetzt zum ersten Modell. In vielen Programmen werden beide Modelle parallel unterstützt und jeder Benutzer kann sich so das für ihn schlüssigere, verständlichere oder bequemere auswählen. In Abschnitt 18.4 werden wir uns etwas allgemeiner mit diesem Phänomen im Zusammenhang mit kleinen mobilen Bildschirmen beschäftigen.

## 5.3 Fehler des Benutzers

Unser Ziel als Designer oder Entwickler ist es (normalerweise), Geräte und Systeme zu bauen, die der menschliche Benutzer möglichst fehlerfrei bedienen kann. Um uns diesem Ziel anzunähern, müssen wir zunächst verstehen, welche Arten von **Fehlern** Menschen machen, und wie diese Fehler entstehen. Für dieses Verständnis wiederum ist es hilfreich, sich anzuschauen, wie wir als Menschen handeln, um ein bestimmtes Ziel zu erreichen.

### 5.3.1 Die Ausführung zielgerichteter Handlungen

Am Beginn einer zielgerichteten **Handlung** steht das **Ziel**, das wir erreichen wollen. Unsere Erfahrung und unser Wissen erlauben uns, eine bestimmte Handlung auszuwählen, von der wir uns versprechen, dass sie dieses Ziel erreichen wird. Diese Handlung besteht wiederum aus einer Folge einzelner Aktionen oder Handlungsschritte, die wir dann ausführen müssen. Damit stoßen wir Vorgänge in der Welt oder in einem System an, die wiederum bestimmte Reaktionen oder Ergebnisse hervorrufen. Diese Reaktionen beobachten wir und interpretieren sie vor dem Hintergrund unseres Wissens und unserer Erfahrung. Diese Interpretation führt zu einer Bewertung, ob das Ziel erreicht wurde und ruft ggf. neue Ziele hervor. Praktisch an allen Stellen dieses Zyklus können nun Probleme auftreten, und diese Probleme führen zu verschiedenen Arten von Fehlern.

Die Kluft zwischen dem angestrebten Ziel und der korrekten Ausführung der richtigen Folge von Aktionen wird als **gulf of execution** bezeichnet. Alles, was die Erkennbar-

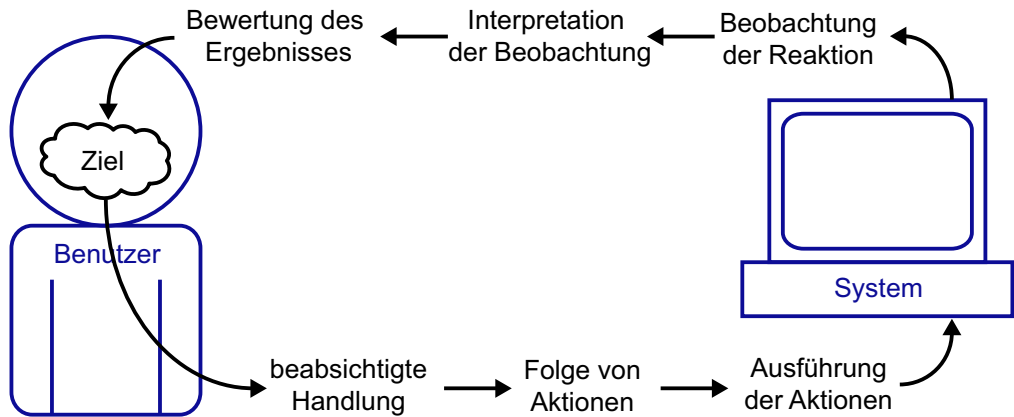


Abbildung 5.2: Die sieben Schritte bei der Ausführung einer zielgerichteten Handlung, adaptiert nach Norman [86]

keit der richtigen Aktionen sowie deren korrekte Ausführung behindert, verbreitert diese Kluft. Gleichmaßen gibt es einen **gulf of evaluation**, eine Kluft, die zwischen den Ausgaben des Systems und deren korrekter Interpretation und Bewertung liegt. Unvollständige oder irreführende Fehler- oder Statusmeldungen sind Beispiele, die diese Kluft verbreitern. Gutes Interaktionsdesign versucht, in beiden Fällen die Breite der Kluft zu minimieren, indem es z.B. immer alle ausführbaren und sinnvollen Aktionen sichtbar macht, und den Status des Systems jederzeit klar und unmissverständlich darstellt.

### 5.3.2 Grundlegende Fehlerarten

Sowohl **Don Norman** [86] als auch **James Reason** [94] unterscheiden zwei grundlegende Arten von Fehlern, nämlich **Irrtümer** (engl. **mistakes**) und **Fehlleistungen** (engl. **slips**). Die echten Irrtümer beruhen darauf, dass zur Erreichung eines Ziels tatsächlich die falsche Handlung ausgewählt wurde. Ein Beispiel dafür wäre, ein Feuer mit einer brennbaren Flüssigkeit wie Spiritus löschen zu wollen. Fehlleistungen entstehen hingegen dadurch, dass eine Handlung zwar richtig geplant, aber falsch ausgeführt wurde. Im obigen Beispiel bedeutet das, dass jemand zwar nach dem Wasser greifen wollte, um das Feuer zu löschen, versehentlich aber nach der Flasche mit Spiritus gegriffen hat.

Irrtümer hängen meist mit einem fehlerhaften mentalen Modell (siehe Kapitel 5) zusammen. Dieses Modell legt dann eben eine falsche Handlung nahe oder bietet gar keine Anhaltspunkte zur Auswahl einer Handlung, so dass zufällig ausgewählt wird. Andere Gründe für echte Irrtümer sind fehlendes Wissen oder fehlende Erfahrung. Echte Irrtümer passieren vor allem im ersten und letzten Schritt des obigen Zyklus, nämlich bei der Auswahl einer konkreten Handlung oder bei der Bewertung des Ergebnisses einer Handlung. Die Fehlleistungen sind wesentlich variantenreicher und passieren vor

allem dadurch, dass (teil-)automatisierte Handlungsabläufe fehlerhaft ausgeführt werden. Beispiele sind (in Anlehnung an Norman [86]):

- Fangfehler: Hier nimmt ein oft ausgeführter und vertrauter Handlungsablauf einen anderen, selten praktizierten, aber gleich beginnenden Ablauf gefangen. Beispiel: ein Benutzer meldet sich auf einer selten besuchten Webseite mit Name und Passwort seines täglich benutzten Rechner-Logins an.
- Beschreibungsfehler treten dann auf, wenn zwei Handlungsabläufe sehr ähnlich sind oder sehr nahe beieinander liegen, wie z.B. das Drücken benachbarter Tasten oder die Eingabe ähnlicher Kommandos.
- Datengesteuerte Fehler treten bei datengesteuerten Handlungen wie z.B. dem Abtippen einer Nummer auf. Beispiel: Anrufen der Faxnummer, die direkt unter der Telefonnummer steht
- Fehler durch assoziierte Aktionen entstehen, wenn die Handlung durch etwas anderes beeinflusst wird, an das man gerade denkt. Ein Beispiel wäre, beim Abheben des Telefons *Herein!* statt *Ja bitte!* zu sagen.
- Fehler durch Aktivierungsverlust entstehen, wenn wir im Verlauf einer Folge von Aktionen den Rest dieser Aktionen vergessen, beispielsweise wenn wir aufstehen um etwas aus der Küche zu holen, und dann in der Küche vergessen haben was wir holen wollten.
- Modusfehler treten dann auf, wenn ein Gerät oder System in verschiedenen Modi sein kann. Ein Beispiel ist das Beenden der falschen Anwendung in Desktop-Umgebungen oder das Treten des vermeintlichen Kupplungspedals in einem Automatik-Fahrzeug.

Praktisch alle Fehler in dieser Liste treten beim zweiten und dritten Schritt des eingangs gezeigten Zyklus auf, nämlich bei der Spezifikation und bei der Ausführung der konkreten Folge von Aktionen. Manche können jedoch auch durch die verschiedenen Schritte von der Beobachtung der Reaktion bis zur Bewertung ausgelöst werden, z.B. Modusfehler durch falsche Interpretation des aktuellen Modus, oder assoziierte Aktionen, die durch Beobachtung oder Interpretation bestimmter Ausgaben entstehen. Manche Webseiten nutzen solche Effekte auch gezielt aus, indem sie Informationen oder Waren versprechen, dann aber an der üblicherweise erwarteten Stelle einen Link zu etwas anderem, wie Werbung oder Schad-Software platzieren. Hier wird darauf spekuliert, dass der Benutzer einen Fang- oder Beschreibungsfehler begeht ohne es zu merken.

### Exkurs: Fehler beim Klettern

Es gibt Tätigkeiten, bei denen Fehler schlimmere Folgen haben. Hierzu gehören **Klettern** und **Bergsteigen**. Die Kletterlegende Kurt Albert beispielsweise starb 2010 in einem einfachen Klettersteig, weil sich die Bandschlinge seines Sicherungskarabiners zufällig ausgehängt hatte<sup>a</sup>. Dem Risiko von Fehlern wird im Bergsport normalerweise durch mehrere Strategien entgegengewirkt:

- **Redundanz:** Sofern irgend möglich, wird die Sicherung eines Kletterers immer durch zwei voneinander unabhängige Systeme hergestellt, beispielsweise durch zwei Fixpunkte beim Bau eines Standplatzes, zwei gegenläufige Karabiner beim Einbinden am Gletscher, oder ein Doppelseil im steinschlaggefährdeten Gelände.
- **Robustheit:** Viele Systeme und Prozesse im Bergsport verhalten sich robust gegen Fehler: bei der Tourenplanung wird ein Sicherheitspuffer für die Rückkehr bei Tageslicht eingeplant, und im alpinen Gelände wird immer ein Biwaksack mitgeführt, sollte doch einmal eine ungeplante Übernachtung notwendig werden. Die Seilkommandos (Stand, Seil ein, Seil aus, Nachkommen) verwenden verschiedene Vokale und Silbenzahlen, um die Chance zu erhöhen, dass sie auch beim Brüllen über 60m Entfernung noch verstanden werden, und am Beginn jeder Klettertour steht der Partnercheck, bei dem beide Kletterer gegenseitig ihre Sicherheitsausrüstung kontrollieren.
- **Einfachheit:** In der Ausbildung werden einfache Verfahren vermittelt, die leichter zu merken sind, statt komplizierterer Verfahren, die zwar theoretisch noch effizienter, im Ernstfall dann aber vergessen sind (z.B. bei der Spaltenbergung mittels loser Rolle).

Trotz aller Vorsichtsmaßnahmen kann aber auch hier das Wirken von Murphy's Gesetz beobachtet werden: Im Sommer 2013 stürzte das zwölfjährige italienische Klettertalent Tito Traversa aus 25m Höhe auf den Boden, weil die Expressschlingen, die er sich ausgeliehen hatte, systematisch falsch montiert waren<sup>b</sup>.

<sup>a</sup><http://www.frankenjura.com/klettern/news/artikel/170>

<sup>b</sup><http://www.bergsteigen.com/news/toedlicher-unfall-wegen-falsch-montierter-express>

### 5.3.3 Murphys Gesetz

Im Zusammenhang mit Fehlern soll hier kurz **Murphys Gesetz**, das oft scherzhaft zitiert wird, sowie seine durchaus ernsten Implikationen für die Mensch-Maschine-Interaktion beschrieben werden. Das Gesetz lautet in seiner kurzen und bekannten Form:

Alles, was schiefgehen kann, wird auch schiefgehen.

Es geht laut Bloch [11] auf den amerikanischen Forscher Edward Murphy zurück, der es nach einem misslungenen Raketen-Experiment formuliert haben soll: Ziel des Experimentes war, die auf den Passagier wirkenden Beschleunigungen während der Fahrt

eines Raketenschlittens zu messen. Hierzu wurden 16 Sensoren in dem sehr kostspieligen Versuchsaufbau verwendet. Jeder dieser Sensoren konnte auf zwei Arten (richtig oder falsch) angeschlossen werden. Als nach Durchführung des Experiments alle Sensorwerte Null waren, stellte sich heraus, dass ein Techniker alle Sensoren systematisch falsch angeschlossen hatte. Daraufhin soll Murphy die Langfassung formuliert haben:

Wenn es mehrere Möglichkeiten gibt, eine Aufgabe zu erledigen, und eine davon in einer Katastrophe endet oder sonstwie unerwünschte Konsequenzen nach sich zieht, dann wird es jemand genau so machen.

Das Gesetz war zunächst recht zynisch gemeint, wurde jedoch kurze Zeit später von einem Vorgesetzten Murphys als Grundlage der Sicherheitsstrategie der U.S. Air Force bei diesen Experimenten zitiert, und hat auch wichtige Implikationen für die Mensch-Maschine-Interaktion. Davon seien hier einige beispielhaft genannt:

- **Bedienfehler:** Jeder mögliche Bedienfehler wird irgendwann einmal gemacht. Das bedeutet umgekehrt: Will man eine fehlerfreie Bedienung sicherstellen, so dürfen keine Bedienfehler möglich sein. Das bedeutet, dass in jedem Modus des Systems nur die dort zulässigen Bedienschritte möglich sind. Ein einfaches Beispiel hierfür ist eine mechanische Sperre der Gangschaltung im Auto bei nicht durchgetretenem Kupplungspedal.
- **Eingabefehler:** Ermöglicht man dem Benutzer freie Eingaben, so wird dieser auch jeden erdenklichen Unsinn, wie z.B. Sonderzeichen oder nicht existierende Kalenderdaten eingeben. Für den Programmierer bedeutet dies, entweder das Programm gegen solche Eingaben robust zu bauen, oder die Eingabe zur Fehlervermeidung einzuschränken, beispielsweise durch Auswahl eines Datums in einem angezeigten Kalender statt freier Eingabe.
- **Dimensionierung:** Egal wie groß man ein System dimensioniert, es wird immer jemand mit oder ohne böse Absicht diesen Rahmen sprengen.

Letzten Endes ist es in der Praxis unmöglich, alle Fehler vorherzusehen, da diese auch außerhalb eines vielleicht beweisbar korrekten oder fehlerfreien Systems liegen können. Murphys Gesetz bekräftigt aber, dass der Mensch als Fehlerquelle in der Mensch-Maschine-Interaktion fest mit eingeplant werden muss und Benutzerfehler nicht etwa unvorhersehbare Unglücke, sondern fest vorhersagbare Ereignisse sind.



### Übungsaufgaben:

1. Ein Bekannter des Autors antwortete nicht auf die in seiner Webseite angegebene dienstliche E-Mail-Adresse. Darauf angesprochen begründete er, sein Rechner im Büro sei durch einen Wasserschaden kürzlich defekt geworden, und damit auch seine Mailbox. Er müsse nun von zuhause E-Mail lesen, und das sei eben eine andere Mailbox. Beschreiben oder skizzieren Sie das mentale Modell, das der Benutzer vom Gesamtsystem E-Mail hat, begründen Sie, woher er es haben könnte, und vergleichen Sie es mit dem tatsächlichen (oder zumindest dem von Ihnen angenommenen) konzeptuellen Modell von E-Mail (egal ob nach dem POP oder iMap Protokoll). Wo sind irreführende Elemente im präsentierten Modell?
2. Ein anderer Bekannter rief eines Tages an, während er ein Haus renovierte. Der Klempner hatte die Wasserhähne im Bad montiert, aber ihre Bedienung fühlte sich eigenartig an. Der Bekannte bat den Klempner, die Drehrichtungen der Ventile umzukehren, aber auch danach schien die Bedienung immer noch eigenartig. Der Klempner rechnete vor, dass es noch 2 weitere Kombinationen gebe und dass er gerne dafür Anfahrt und Montage berechne. Wir alle kennen solche Wasserhähne mit zwei Schraubventilen an den Seiten, einem für kaltes und einem für warmes Wasser. In der Regel ist das warme Wasser links und das kalte rechts. Doch in welche Richtung muss man die beiden Ventile drehen, damit Wasser kommt? Beide im Uhrzeigersinn? Beide dagegen? Beide so, dass sich die Handgelenke nach innen drehen? Oder nach außen? Schauen Sie nicht nach, sondern führen Sie eine kleine Umfrage in Ihrem Bekanntenkreis durch und finden Sie heraus, welche mentalen Modelle über die Funktion eines Wasserhahns existieren, z.B. in Anlehnung an andere Systeme mit runden Regel-Elementen wie Heizkörper, Gasherd oder Stereoanlage.
3. Analysieren Sie anhand der Abbildung 5.2 eine alltägliche Handlung am Bildschirm, z.B. das Umbenennen einer Datei in einer Desktop-Umgebung. Benennen Sie die konkreten Probleme, die in jedem Schritt auftreten können! Vergleichen Sie diese Probleme mit denen beim Umbenennen einer Datei in einer Kommandozeilen-Umgebung.
4. Recherchieren Sie im Web nach Seiten, die die Eingabe von Kalenderdaten mit bestimmten Einschränkungen erfordern (z.B. Flugbuchung, Hinflug vor Rückflug). Wie gehen diese Seiten mit Eingabebefehlern um? Diskutieren Sie den Tradeoff zwischen Programmieraufwand und Benutzerfreundlichkeit.