

■ **Abb. 4.1.** Konstanzmethode. (Nach Hofstätter, 1957)

von Vergleichsreizen (S_v) kombiniert. Die Untersuchungsteilnehmer müssen bei jedem Paar entscheiden, ob der Vergleichsreiz größer oder kleiner (lauter oder leiser) als der Bezugsreiz ist. Das folgende Beispiel (nach Hofstätter, 1957, S. 241 f.) demonstriert das weitere Vorgehen.

Es soll die Differenzschwelle eines Untersuchungsteilnehmers für die Unterscheidung von Gewichten untersucht werden. Ein **Standardreiz** $S_0=100$ g wird mit einer Reihe von Gewichten (S_v) zwischen 88 g und 108 g kombiniert. Der Untersuchungsteilnehmer erhält die Gewichtspaare in zufälliger Reihenfolge mit der Bitte zu entscheiden, ob der **Vergleichsreiz** größer ($S_v > S_0$) oder kleiner ($S_v < S_0$) als der Standardreiz ist. (Im Beispiel wird auch eine dritte Urteilkategorie »gleich« zugelassen.) Jedes Gewichtspaar muss vom Untersuchungsteilnehmer mehrmals beurteilt werden. ■ Abb. 4.1 gibt in idealisierter Form die prozentualen Häufigkeiten der Paarvergleichsurteile wieder.

Für $S_v > S_0$ und $S_v < S_0$ resultiert jeweils eine S-förmig geschwungene Verteilung, die einem Vorschlag Urbans (1931) folgend als »psychometrische Funktion« bezeichnet wird. (■ Abb. 4.1 ist z. B. zu entnehmen, dass der Urteiler beim Vergleich der Gewichte $S_0=100$ g und $S_v=104$ g in 80% aller Fälle $S_v > S_0$ urteilte.) Sind die Empfindungsstärken für einen Reiz im Sinne Thurstones normal verteilt, folgt die psychometrische Funktion den Gesetzmäßigkeiten einer kumulierten Normalverteilung

(»Ogive«). Der Schnittpunkt der beiden psychometrischen Funktionen markiert den **scheinbaren Gleichwert** (SG); er liegt bei 96,8 g. Offensichtlich neigte der hier urteilende Untersuchungsteilnehmer dazu, das Gewicht des Standardreizes zu unterschätzen. Der Differenzbetrag zum tatsächlichen Gleichgewicht (100 g–96,8 g=3,2 g) wird als **konstanter Fehler** (KF) bezeichnet. Die zur 50%-Ordinate gehörenden Abszissenwerte (94,5 g bzw. 99,2 g) kennzeichnen die Grenzen des sog. Unsicherheitsintervalls (4,7 g). Die Hälfte dieses Unsicherheitsintervalls (2,35 g) stellt die **Unterschiedsschwelle** bzw. den EMU dar.

Betrachtet man nur die Funktion für $S_v > S_0$, entspricht der SG demjenigen Gewicht, bei dem der Untersuchungsteilnehmer in 50% aller Fälle $S_v > S_0$ urteilt (im Beispiel: SG=99 g). Hieraus ergibt sich der konstante Fehler zu $KF=S_0-SG$ ($KF=100$ g–99 g=1 g). Zur Bestimmung des EMU benötigt man diejenigen Vergleichsreize, bei denen in 25% (S_{25}) bzw. in 75% (S_{75}) aller Fälle $S_v > S_0$ geurteilt wurde (im Beispiel: $S_{25}=95$ g; $S_{75}=103$ g). Mit diesen Werten resultiert $EMU=(S_{75}-S_{25})/2$ bzw. im Beispiel $(103$ g–95 g)/2=4 g (vgl. hierzu z. B. Irtel, 1996).

Das »Signalentdeckungsparadigma«

Die Psychophysik samt ihrer Methoden (Überblick z. B. bei Eijkmann, 1979; Geissler & Zabrodin, 1976; Guilford, 1954; Irtel, 1996; Mausfeld, 1994; Stevens, 1951, Kap. 1) ist in ihren modernen Varianten stark durch die von Tanner und Swets (1954) bzw. Green und Swets (1966) in die Human- und Sozialwissenschaften eingeführte Signalentdeckungstheorie (»Signal Detection Theory«, SDT) geprägt. Vertreter der Signalentdeckungstheorie bezweifeln die Existenz sensorischer Schwellen und schlagen stattdessen das Konzept der »Reaktionsschwelle« vor. Es wird explizit zwischen der organisch bedingten Sensitivität des Menschen und seiner Bereitschaft unterschieden, in psychophysischen Experimenten (oder auch in ähnlich strukturierten Alltagssituationen) bestimmte Wahlentscheidungen zu treffen. Die organische **Sensitivität** wird als physiologisch und die **Reaktionsschwelle** (oder Entscheidungsbereitschaft) als psychologisch bedingt angesehen (z. B. durch die Bewertung der Konsequenzen, die mit verschiedenen Entscheidungen verbunden sind). Sensitivität und Reaktionsschwelle sind zwei Einflussgrößen, die den Ausgang einer Wahlentscheidung (z. B. bei einem Paarver-

gleichsurteil) gemeinsam beeinflussen. Herauszufinden, welchen Anteil diese beiden Determinanten bei der Steuerung von Wahlentscheidungen haben, ist Aufgabe von Analysen nach der Signalentdeckungstheorie.

Was mit dem Begriff »Reaktionsschwelle« gemeint ist, soll ein kleines Beispiel verdeutlichen: Ein Schüler klagt über Bauchschmerzen und muss zum Arzt. Dieser tastet die Bauchhöhle ab und fragt, ob es weh tut. Man kann ziemlich sicher sein, dass die Entscheidung des Schülers, Schmerzen zu bekunden, davon abhängt, ob z. B. am nächsten Tage eine schwere Klassenarbeit bevorsteht oder ob auf Klassenfahrt gegangen wird. Unabhängig davon, ob die tatsächlichen Empfindungen (Sensitivität) diesseits oder jenseits der physiologischen Schmerzschwelle liegen, wird der Schüler in Erwartung der Klassenarbeit über stärkere Schmerzen klagen als in Erwartung der Klassenfahrt (Reaktionsschwelle).

! **Sensitivität und Reaktionsschwelle sind in klassischen, psychophysischen Untersuchungen konfundiert. Die auf der Signalentdeckungstheorie basierenden Methoden machen eine Trennung dieser beiden Reaktionsaspekte möglich.**

Terminologie. Die Signalentdeckungstheorie geht auf die statistische Entscheidungstheorie (vor allem auf den Ansatz von Neyman & Pearson, 1928, ► S. 491) zurück und hat zum großen Teil das dort übliche Vokabular übernommen. Die objektiv vorgegebenen Reize werden als Input und die Reaktionen der Untersuchungsteilnehmer als Output bezeichnet.

Allgemein setzt die Anwendung des Signalentdeckungsansatzes eine Reihe von (Input-)Reizen voraus, bei denen das untersuchte Merkmal zunehmend stärker ausgeprägt ist ($S_0, S_1, S_2, \dots, S_k$). Diesen Reizen zugeordnet sind Einschätzungen der Merkmalsausprägungen (Empfindungsstärken) $z_0, z_1, z_2, \dots, z_k$ durch die Untersuchungsteilnehmer (Output). Werden Merkmalsausprägungen S_1, S_2, \dots, S_k mit $S_0=0$ verglichen, handelt es sich um die Untersuchung der absoluten Sensitivität (**Absolutschwelle**) bzw. um die Ermittlung der minimalen Reizintensität, die eine gerade eben merkliche Empfindung auslöst. Paarvergleiche von Merkmalsintensitäten $S_i > 0$ können zur Bestimmung der differenziellen Sensitivität (**Differenzschwelle**) genutzt werden.

Ein Vierfelderschema (► Tab. 4.7) verdeutlicht, wie die Reaktionen eines Untersuchungsteilnehmers in Sig-

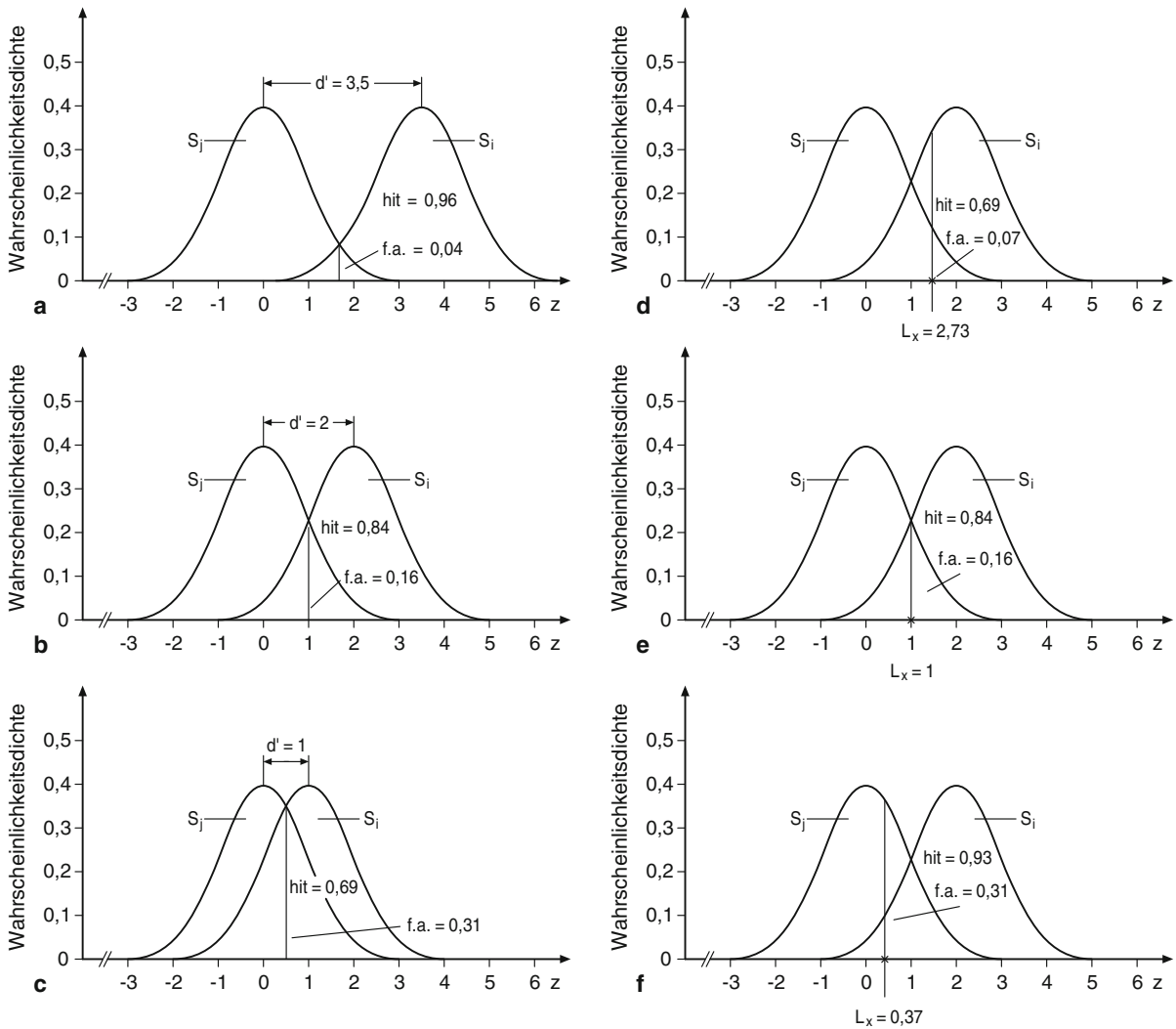
Tab. 4.7. Reaktionsklassifikation in einem Signalentdeckungsexperiment

Input: $S_i > S_0$?	Output: $z_i > z_0$?	
	Ja	Nein
Ja	Hit	Miss
Nein	False Alarm	Correct Rejection

nalentdeckungsexperimenten zu klassifizieren sind. Die Bezeichnung der vier Felder bezieht sich auf die »klassische« Versuchsanordnung eines Signalentdeckungsexperiments, bei der ein energieschwaches Signal unter vielen Störsignalen (Noise, Rauschen) zu identifizieren ist (z. B. die Identifikation feindlicher Flugzeuge auf dem Radarschirm). Der Input wäre in dieser Situation mit den Reizintensitäten $S_i > 0$ und $S_0 = 0$ zu charakterisieren.

Die Identifikation eines tatsächlich vorhandenen Signales wird als »Hit« bezeichnet. Das Übersehen eines Signales führt zu einem »Miss«. Wird ein Störsignal (bzw. ein nicht vorhandenes Signal oder »Noise«) als Signal gedeutet, bezeichnet man dies als »False Alarm« und die korrekte Reaktion auf ein nicht vorhandenes Signal als »Correct Rejection«. Diese Bezeichnungen werden üblicherweise auch angewendet, wenn zwei unterschiedlich stark ausgeprägte Reize zu vergleichen sind ($S_i > 0; S_j > 0; S_i \neq S_j$). Man beachte, dass – wie bei der Konstanzmethode (aber anders als bei dem Law of Comparative Judgement) – die objektiven Größer-kleiner-Relationen der Inputreize bekannt sein müssen.

Die Informationen, die für die Bestimmung von Sensitivität und Reaktionsschwelle benötigt werden, sind die Wahrscheinlichkeiten (geschätzt durch relative Häufigkeiten) für eine Hit-Reaktion und für eine False-Alarm-Reaktion. (Die Wahrscheinlichkeiten für eine Miss-Reaktion bzw. eine Correct-Rejection-Reaktion enthalten keine zusätzlichen Informationen, da sie zu den oben genannten Wahrscheinlichkeiten komplementär sind. Die Wahrscheinlichkeiten für Hit und Miss bzw. für False Alarm und Correct Rejection addieren sich jeweils zu 1.) Die Bestimmung dieser Wahrscheinlichkeiten stellt eine erhebliche untersuchungstechnische Schwierigkeit dar, die der praktischen Anwendbarkeit der Signalentdeckungsmethoden Grenzen setzt (► unten). Die Verrechnung der Paarvergleichsurteile nach dem Signalentdeckungsparadigma zielt darauf ab, den



▣ **Abb. 4.2a-f.** Hit- und False-Alarm-Raten in Beziehung zur Sensitivität d' (a-c) und zur Reaktionsschwelle L_x (d-f) (Erläuterungen ► Text)

Sensitivitätsparameter (d') eines Urteilers sowie dessen **Reaktionsschwelle** (Response Bias – Parameter L_x oder β) zu bestimmen. Diesem an sich einfachen Rechengang (der auf ► S. 168 f. beschrieben wird) liegt eine relativ komplizierte Theorie zugrunde, die im Folgenden kurz dargestellt wird.

Theorie. Die Signalentdeckungstheorie basiert – wie auch das Thurstone'sche Law of Comparative Judgment – auf der Annahme, dass die Empfindungsstärke für einen Reiz eine normal verteilte Zufallsvariable darstellt. **Empfindungsstärkeverteilungen** für verschiedene

Reize unterscheiden sich in ihren Mittelwerten, aber nicht in ihrer Streuung. Werden zwei Reize mit unterschiedlichen Merkmalsausprägungen verglichen, ist bei nicht allzu großen Reizunterschieden mit einer Überschneidung der beiden Empfindungsstärkeverteilungen zu rechnen (▣ Abb. 4.2). Entsprechendes gilt für den Vergleich »Reiz« vs. »Noise«, bei dem ebenfalls unterstellt wird, dass sich die »Noiseverteilung« und die »Reizverteilung« überschneiden.

Das Paarvergleichsurteil geht modellhaft folgendermaßen vonstatten: Zu zwei zu vergleichenden Reizen S_i und S_j ($S_i > S_j$) gehören zwei Verteilungen von Empfin-

dungsstärken, von denen wir zunächst annehmen, sie seien dem Urteiler bekannt. Die in einem Versuch durch Reiz S_i ausgelöste Empfindungsstärke z_i wird mit beiden Verteilungen verglichen. Ist die Wahrscheinlichkeit dafür, dass diese Empfindungsstärke zur Empfindungsstärkenverteilung von S_i gehört, größer als die Wahrscheinlichkeit der Zugehörigkeit zu der Empfindungsstärkenverteilung für S_j , entscheidet ein perfekter Urteiler $S_i > S_j$. Sind die Wahrscheinlichkeitsverhältnisse umgekehrt, lautet die Entscheidung $S_j > S_i$. Das Entscheidungskriterium (in [Abb. 4.2](#) durch einen senkrechten Strich verdeutlicht) für die Alternativen $S_i > S_j$ und $S_i < S_j$ liegt bei einer Empfindungsstärke, die in beiden Verteilungen die gleiche Dichte (=Höhe der Ordinate) aufweist. Rechts von diesem Entscheidungskriterium sind die Dichten für die S_i -Verteilung größer als für die S_j -Verteilung, d. h., hier müsste $S_i > S_j$ geurteilt werden.

In [Abb. 4.2a](#) unterscheiden sich die beiden Reize um $d' = 3,5$ Einheiten der z-normierten Empfindungsstärkenskala. (Der arbiträre Nullpunkt dieser Skala wurde beim Mittelwert der S_j -Verteilung angenommen.) Offensichtlich kann dieser Urteiler die beiden Reize S_i und S_j recht gut unterscheiden. Seine Sensitivität bzw. Diskriminationsfähigkeit (d') ist hoch. Bei diesem deutlich unterscheidbaren Reizpaar kommt es mit einer Wahrscheinlichkeit von 96% zu einem Hit (S_i wird korrekterweise für größer als S_j gehalten) und mit einer Wahrscheinlichkeit von nur 4% zu einer f.a.- bzw. False-Alarm-Reaktion (S_i wird fälschlicherweise für größer als S_j gehalten). Mit geringer werdendem Abstand der beiden Reize (bzw. mit abnehmendem d') sinkt die Hit-Rate und steigt die False-Alarm-Rate. Bei einem Abstand von einer Empfindungsstärkeeinheit ($d' = 1$, vgl. [Abb. 4.2c](#)) beträgt die Hit-Wahrscheinlichkeit 69% und die False-Alarm-Wahrscheinlichkeit 31%. Der Parameter d' charakterisiert die Sensitivität bzw. die sensorische Diskriminationsfähigkeit eines Urteilers.

Signalentdeckungstheoretiker vermuten nun, dass ein Urteiler selten so perfekt urteilt wie in den [Abb. 4.2a–c](#). Bedingt durch psychologische Umstände kann unabhängig vom »objektiven« Entscheidungskriterium bevorzugt $S_i > S_j$ (oder $S_i < S_j$) geurteilt werden. Dies wird in den [Abb. 4.2d–f](#) deutlich. Fällt beispielsweise die in einem Versuch durch Reiz S_i ausgelöste Empfindungsstärke in den Bereich 1 bis 1,5, ordnet der Urteiler diese Empfindungsstärke dem Reiz S_j zu, ob-

wohl in diesem Bereich die Wahrscheinlichkeitsdichte für die S_i -Verteilung größer ist als für die S_j -Verteilung (vgl. [Abb. 4.2d](#)). Das **Entscheidungskriterium** (L_x) oder die **Reaktionsschwelle** ist nach rechts versetzt. Dadurch wird die Wahrscheinlichkeit einer False-Alarm-Reaktion (es wird fälschlicherweise $S_i > S_j$ behauptet) zwar geringer; gleichzeitig sinkt jedoch auch die Wahrscheinlichkeit eines Hits. (Man vergleiche hierzu die [Abb. 4.2b](#) und [d](#) mit $d' = 2$.)

[Abb. 4.2f](#) zeigt ein zu weit nach links versetztes Entscheidungskriterium (L_x). Hier werden im Bereich 0,5 bis 1 die Empfindungsstärken, die durch Reiz S_j ausgelöst werden, der S_i -Verteilung zugeordnet, obwohl die Wahrscheinlichkeitsdichten für die S_j -Verteilung größer sind als für die S_i -Verteilung.

Damit erhöht sich zwar die Wahrscheinlichkeit eines Hits (93%); gleichzeitig steigt jedoch die Wahrscheinlichkeit für False Alarm (31%). In [Abb. 4.2e](#) ist das Entscheidungskriterium – wie bereits in [Abb. 4.2a–c](#) – »richtig« plazierte. Beide Wahrscheinlichkeitsdichten (die Wahrscheinlichkeitsdichte für die S_i -Verteilung und die Wahrscheinlichkeitsdichte für die S_j -Verteilung) sind für die Empfindungsstärke, die das Entscheidungskriterium markiert, gleich groß. Das Verhältnis der Wahrscheinlichkeitsdichten lautet $L_x = 1$. (L von Likelihood-Ratio: Wahrscheinlichkeitsdichte in der S_i -Verteilung dividiert durch die Wahrscheinlichkeitsdichte in der S_j -Verteilung. Für L_x wird in der Literatur gelegentlich auch der Buchstabe β verwendet.) Für [Abb. 4.2d](#) lautet der Wert $L_x = 2,73$: Das Entscheidungskriterium (mit $z_i = -0,5$ und $z_j = 1,5$) hat in der S_i -Verteilung eine höhere Dichte als in der S_j -Verteilung (0,352 für S_i und 0,129 für S_j). In [Abb. 4.2f](#) resultiert für den Quotienten ein Wert unter 1 ($L_x = 0,37$). Das Entscheidungskriterium (mit $z_i = -1,5$ und $z_j = 0,5$) hat in der S_j -Verteilung eine größere Dichte (0,352) als in der S_i -Verteilung (0,129).

L_x -Werte charakterisieren die Reaktionsschwelle eines Urteilers. Werte über 1 sprechen für eine »**konservative**« oder ängstliche Entscheidungsstrategie: False-Alarm-Entscheidungen werden möglichst vermieden, bei gleichzeitigem Risiko, dabei die Hit-Rate zu reduzieren. Umgekehrt weisen L_x -Werte unter 1 eher auf eine »**progressive**« oder mutige Entscheidungsstrategie hin: Die Hit-Rate soll möglichst hoch sein bei gleichzeitig erhöhtem False-Alarm-Risiko. Generell gilt, dass Werte $L_x \neq 1$ für eine Reaktionsverzerrung (Response Bias) sprechen.

Man beachte, dass in den **Abb. 4.2a–c** L_x konstant ($L_x=1$) und d' variabel und in den **Abb. 4.2d–f** d' konstant ($d'=2$) und L_x variabel ist. Hiermit wird eine wichtige Eigenschaft der Signalentdeckungsparameter deutlich:

! Die sensorische Diskriminationsfähigkeit d' ist unabhängig von der Reaktionsschwelle L_x .

Traditionelle psychophysische Methoden nutzen lediglich die Hit-Rate, um die Differenzschwelle zu bestimmen. Danach wäre der Urteiler in **Abb. 4.2f** mit einer Hit-Rate von 93% äußerst sensitiv (niedrige Differenzschwelle) und der Urteiler in **Abb. 4.2b** mit einer Hit-Rate von 84% weniger sensitiv (hohe Differenzschwelle). Tatsächlich verfügen nach der Signalentdeckungstheorie beide Urteiler über die gleiche Sensitivität ($d'=2$); der Unterschied in den Hit-Raten ist auf verschiedene Reaktionsschwellen (L_x) und nicht auf unterschiedliche Sensitivitäten zurückzuführen.

Neuere Entwicklungen zur (verteilungsfreien) Bestimmung von d' und L_x hat Balakrishnan (1998) vorgestellt.

Praktisches Vorgehen. Bei den bisherigen Ausführungen zur Theorie der Signalentdeckung wurde davon ausgegangen, dass die Empfindungsstärkeverteilungen, die d' - bzw. L_x -Werte und damit auch die Hit-Wahrscheinlichkeiten und False-Alarm-Wahrscheinlichkeiten bekannt seien. Dies ist normalerweise jedoch nicht der Fall. In der Praxis werden – in umgekehrter Reihenfolge – zunächst die Hit-Wahrscheinlichkeiten und die False-Alarm-Wahrscheinlichkeiten und danach erst d' für die Sensitivität bzw. L_x für die Reaktionsschwelle ermittelt.

Die Bestimmung der Hit- und False-Alarm-Wahrscheinlichkeiten ist eine äußerst zeitaufwendige und für die Untersuchungsteilnehmer häufig bis an die Grenzen ihrer Belastbarkeit gehende Aufgabe. (Dies gilt jedoch nicht nur für Signalentdeckungsaufgaben, sondern z. B. auch für die klassische Konstanzmethode – vgl. **Abb. 4.1** –, bei der ebenfalls für die einzelnen Reizkombinationen Präferenzwahrscheinlichkeiten geschätzt werden müssen.) Um für ein Reizpaar die entsprechenden Wahrscheinlichkeiten schätzen zu können, sollten mindestens 50 Versuche durchgeführt werden, d. h., ein Untersuchungsteilnehmer muss für dasselbe Reizpaar mindestens 50 Mal entscheiden, welcher der beiden Reize das

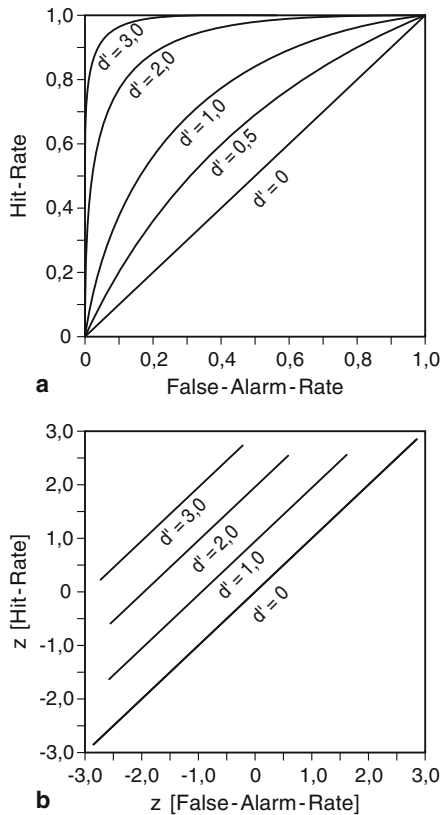
untersuchte Merkmal in stärkerem Maße aufweist. Es besteht die Gefahr, dass bei derartig aufwendigen Versuchsreihen die Ergebnisse durch Ermüdungs- oder Übungseffekte verfälscht werden.

Bei vier Reizen wären sechs Reizpaare in jeweils 50facher Wiederholung zu bewerten, d. h., dem Untersuchungsteilnehmer würden 300 Paarvergleichsurteile abgefordert. Bei mehr als vier Reizen lohnt sich ein vollständiger Paarvergleich nur selten, weil dann Reizpaare auftreten können, die so deutlich voneinander verschieden sind, dass sich die Empfindungsstärkeverteilungen nicht mehr überschneiden.

Repräsentieren mehrere Reize äquidistant ein breiteres, objektiv erfassbares Merkmalskontinuum, erspart es Untersuchungsaufwand, wenn nur benachbarte Reize verglichen bzw. die Reize zunächst nach der Methode der sukzessiven Kategorien geordnet werden. Hit- und False-Alarm-Raten benachbarter Reize basieren dann auf Urteilen, bei denen der objektiv größere Reiz auch für größer gehalten (Hit) bzw. fälschlicherweise als kleiner eingestuft wurde (False Alarm). (Näheres zu dieser methodischen Variante findet sich z. B. bei Velden & Clark, 1979, oder Velden, 1982.)

Nach Ermittlung der Hit- und False-Alarm-Raten gestaltet sich die **Berechnung von d' und L_x** vergleichsweise einfach. Nehmen wir einmal an, ein Untersuchungsteilnehmer hätte in 100 Versuchen mit $S_i > S_j$ 93mal $S_i > S_j$ geurteilt und bei 100 Versuchen mit $S_i < S_j$ 31mal das Urteil $S_i > S_j$ abgegeben. Er hätte damit eine Hit-Rate von 93% und eine False-Alarm-Rate von 31%. Um d' zu ermitteln, werden anhand der Standardnormalverteilungstabelle (**Anhang F**, **Tab. F1**) diejenigen z -Werte bestimmt, die von der Fläche 93% bzw. 31% abschneiden. Diese Werte lauten $z_i=1,50$ und $z_j=-0,50$ (mit gerundeten Flächenanteilen). Die Differenz dieser beiden Werte entspricht d' : $d'=1,50-(-0,50)=2,00$.

Für die Bestimmung von L_x werden die Wahrscheinlichkeitsdichten (Ordinaten) dieser z -Werte in der Standardnormalverteilung benötigt, die ebenfalls in **Tab. F1** aufgeführt sind. Sie lauten in unserem Beispiel 0,129 (für $z_i=1,50$) und 0,352 (für $z_j=-0,50$). Als Quotient resultiert der Wert $L_x=0,37$. Diese Werte entsprechen den Verhältnissen in **Abb. 4.2f** (der Likelihood-Quotient und andere Maße zur Beschreibung von »Response-Bias« werden bei MacMillan & Creelman, 1990, vergleichend diskutiert).



■ **Abb. 4.3.** ROC-Kurven für unterschiedliche d' -Werte mit Hit- und False-Alarm-Raten (a) bzw. z-Werten (b) als Koordinaten. (Nach MacMillan, 1993, S. 25)

In den ■ Abb. 4.2d–f variieren die L_x -Werte bei konstantem $d' = 2$. Da d' als Differenz zweier z-Werte berechnet wird ($d' = z_i - z_j$), gibt es theoretisch unendlich viele z-Wert-Paare, die der Bedingung $d' = 2$ genügen (z. B. $z_i = 0,60$ und $z_j = -1,40$; $z_i = 0,85$ und $z_j = -1,15$ etc.). Zu jedem dieser z-Wert-Paare gehört eine False-Alarm- und eine Hit-Rate bzw. ein spezifischer L_x -Wert. Trägt man nun in ein Koordinatensystem mit der False-Alarm-Rate als Abszisse und der Hit-Rate als Ordinate alle Paare von Hit- und False-Alarm-Raten ein, die zu einem identischen d' -Wert führen, erhält man die sog. **Receiver Operating Characteristic** oder kurz **ROC-Kurve** (bei Swets, 1973, steht die Abkürzung ROC für »Relative« Operating Characteristic).

Die ■ Abb. 4.3a zeigt ROC-Kurven für unterschiedliche d' -Werte. Transformiert man die Hit- und False-

Alarm-Raten in z-Werte der Standardnormalverteilung, sollten die ROC-Kurven für unterschiedliche d' -Werte idealerweise parallele Geraden sein (■ Abb. 4.3b).

Ein Urteiler ohne Diskriminationsvermögen (mit $d' = 0$) urteilt insoweit zufällig, als die Hit-Rate immer der False-Alarm-Rate entspricht. Mit größer werdendem d' verändert sich das Verhältnis Hit zu False Alarm zugunsten der Hit-Rate. Links vom Scheitelpunkt der Kurven (für $d' > 0$) fallen die Entscheidungen konservativ ($L_x > 1$) und rechts davon progressiv ($L_x < 1$) aus.

L_x und d' sind dimensionslose Zahlen und sagen nichts über die tatsächliche Diskriminationsfähigkeit bzw. über die Lokalisierung der Reaktionsschwelle auf dem Merkmalskontinuum aus. Clark (1974) zeigt in einer Untersuchung über Schmerzreaktionen auf unterschiedlich intensive Thermalreize, wie Sensitivitäts- und Reaktionsschwellenparameter in Einheiten des untersuchten Merkmals transformiert werden können.

Risikofreie Entscheidungen, die beispielsweise beim Paarvergleich verschiedener Gewichte im Rahmen einer wissenschaftlichen Untersuchung zu treffen sind, weisen in der Regel nur geringfügige Reaktionsverzerrungen auf ($L_x \approx 1$). Die meisten alltäglichen Entscheidungen dürften jedoch insoweit riskant sein, als sie bestimmte Konsequenzen nach sich ziehen, die vom Urteiler mehr oder weniger negativ bewertet werden. In derartigen Fällen ist mit deutlichen Reaktionsverzerrungen zu rechnen ($L_x \neq 1$).

Wie man Konfidenzintervalle für d' und L_x berechnen kann (bzw. wie man Parameterdifferenzen auf Signifikanz testet), wird bei Kadlec (1999) beschrieben.

Anwendungen. Der Einsatz der Signalentdeckungsmethodik empfiehlt sich generell, wenn die vier verschiedenen, mit einer Entscheidungssituation verbundenen Ausgänge (vgl. ■ Tab. 4.7) unterschiedlich bewertete Konsequenzen nach sich ziehen oder – in Termini der Signalentdeckungstheorie – unterschiedliche Auszahlungen oder »Pay offs« aufweisen. Ärzte unterscheiden sich beispielsweise darin, ob sie bei Verdacht auf Krebs eher bereit sind, eine Miss-Reaktion oder eine False-Alarm-Reaktion zu riskieren. Bei einem Miss würde ein tatsächlich vorhandener Krebs übersehen werden. Das Risiko besteht, dass sich die Krankheit weiterentwickelt und zu einem späteren Zeitpunkt nicht mehr erfolgreich operiert werden kann. Bei einem False Alarm

hingegen riskiert man eine unnötige Operation mit allen damit verbundenen Folgen.

So steht man im Rahmen der klinischen Diagnostik oft vor dem Problem, bei stetig verteilten Indikatoren für eine Krankheit (z. B. PSA für Prostatakarzinome) entscheiden zu müssen, ab welchem Wert (**Cutoff-Point**) ein Patient als krank und damit behandlungsbedürftig gelten soll. Wird dieser Wert zu niedrig angesetzt, dann werden zu viele Patienten als »krank« diagnostiziert, d. h., man riskiert (zugunsten einer hohen Hit-Rate) eine hohe False-Alarm-Rate (progressive Entscheidungsstrategie). Bei zu hohem Wert könnten tatsächlich kranke Patienten übersehen werden, was mit einer hohen Miss-Rate (bzw. einer geringeren Hit-Rate) gleichzusetzen wäre (konservative Entscheidungsstrategie).

Wenn nun verschiedene Ärzte unterschiedliche Cutoff-Points verwenden, resultieren arztpezifische Hit- und False-Alarm-Raten, die herangezogen werden können, um eine ROC-Kurve für PSA zu konstruieren (dies setzt natürlich voraus, dass die wahren Verhältnisse – krank oder nicht krank – bekannt sind. Die Hit- und False-Alarm-Raten könnten hier also nur katamnestisch, beispielsweise nach Vorliegen histologischer Befunde, ermittelt werden). Die PSA-ROC-Kurve sollte dem Typus nach einer der in [Abb. 4.3a,b](#) dargestellten Kurven entsprechen, d. h., der d' -Parameter müsste für alle Ärzte konstant sein. Die Überlegenheit eines neuen Indikators für Prostatakarzinome (z. B. Anti-VEGF) wäre durch einen höheren d' -Wert nachzuweisen.

Ein weiterer, häufig verwendeter Kennwert für die Sensitivität ist die Fläche unter der ROC-Kurve. Ausführliche Informationen über Entscheidungskriterien in der Medizin findet man bei Lusted (1968).

Der breite Anwendungsrahmen der Signalentdeckungstheorie lässt sich mühelos durch weitere Anwendungsbeispiele belegen. So wurden beispielsweise subjektive Schmerzbeurteilungen unter medikamentösen Bedingungen von Classen und Netter (1985) untersucht. Rollman (1977) setzt sich kritisch mit Anwendungen der Signalentdeckungstheorie in der Schmerzforschung auseinander; Price (1966) befasst sich mit Anwendungen in der Persönlichkeits- und Wahrnehmungspsychologie; Dykstra und Appel (1974) überprüfen mit diesem Ansatz LSD-Effekte auf die auditive Wahrnehmung; Upmeyer (1981) belegt den heuristischen Wert der Signalentdeckungstheorie für theoretische Konstruktionen

im Bereich der sozialen Urteilsbildung und Pastore und Scheirer (1974) geben generelle Hinweise über die breite Anwendbarkeit dieses entscheidungstheoretischen Ansatzes. Über Probleme bei der Übertragung signalentdeckungstheoretischer Ansätze auf psychophysische Fragestellungen berichtet Vossel (1985). Eine Fülle von Anwendungsbeispielen findet man zudem bei Swets (1986b).

Für eine weiterführende Einarbeitung in die Signalentdeckungstheorie sowie deren methodische Erweiterungen stehen inzwischen zahlreiche Monografien und Aufsätze zur Verfügung wie z. B. das bereits erwähnte Buch von Green und Swets (1966) oder auch Coombs et al. (1970), Egan (1975), Eijkman (1979), Hodos (1970), MacMillan (1993), McNicol (1972), Richards und Thornton (1970), Snodgrass (1972), Swets (1964, 1986a) sowie Velden (1982).

4.2.3 Ähnlichkeitspaarvergleiche

Ähnlichkeitspaarvergleiche erfordern vom Urteiler Angaben über die globale Ähnlichkeit bzw. (seltener) die auf ein bestimmtes Merkmal bezogene Ähnlichkeit von jeweils 2 Objekten. In den meisten Anwendungsfällen ist diese Aufgabe für den Urteiler schwerer als ein Dominanzpaarvergleich, bei dem lediglich angegeben wird, bei welchem von 2 Objekten das untersuchte Merkmal stärker ausgeprägt ist.

Die Instruktion der Untersuchungsteilnehmer könnte in etwa lauten: »Schätzen Sie die Ähnlichkeit der folgenden Objektpaare auf einer 5-stufigen Skala mit den Abstufungen ‚sehr ähnlich – ähnlich – weder ähnlich, noch unähnlich – unähnlich – sehr unähnlich‘ ein.« Ein grafisches Verfahren würde von den Untersuchungsteilnehmern fordern, auf einer durch die Extreme »äußerst unähnlich« und »äußerst ähnlich« begrenzten Strecke die empfundene Ähnlichkeit durch ein Kreuz zu markieren. Die Länge der Strecke zwischen dem Skalenende »äußerst unähnlich« und dem gesetzten Kreuz dient dann als Ähnlichkeitsurteil (vgl. hierzu auch [Box 4.8](#); über weitere Methoden zur Ähnlichkeitsschätzung berichtet Sixtl, 1967, S. 277 ff.).

Dieses »Rohmaterial« kann mit verschiedenen Verfahren ausgewertet werden. (Einen anwendungsbezogenen Überblick geben Nosofsky, 1992, sowie Ashby,