



Leseprobe aus: Bittrich/Blankenberger, Experimentelle Psychologie, ISBN 978-3-621-27802-7

© 2011 Beltz Verlag, Weinheim Basel

<http://www.beltz.de/de/nc/verlagsgruppe-beltz/gesamtprogramm.html?isbn=978-3-621-27802-7>

## 2 Methode

### 2.1 Versuchspersonen

#### 2.1.1 Stichprobe

Können nicht alle Individuen der relevanten Population untersucht werden, so muss man eine Stichprobe ziehen. Insbesondere in der Soziologie und Politologie sind aufwändige Ziehungsverfahren gebräuchlich (z. B. Schichten- und Quotenstichproben), in der Psychologie beschränkt man sich jedoch meist auf einfache und verlässliche Verfahren.

**Uneingeschränkte Zufallsauswahl.** Im Idealfall hat man eine Liste aller Populationsmitglieder, aus der per Zufall die Untersuchungsteilnehmer gezogen werden. Solch eine Stichprobe ist eine *repräsentative* Stichprobe. Leider ist solch ein Vorgehen kaum durchführbar, da man in den seltensten Fällen eine Liste aller Populationsmitglieder hat, geschweige denn gewährleisten kann, dass alle ausgewählten Personen an der Untersuchung teilnehmen.

Ein beliebter Fehler bei der Stichprobenziehung besteht darin, als »Liste« beispielsweise das Telefonbuch zu verwenden und anschließend von einer Zufallsauswahl zu sprechen. Dies ist jedoch nicht zutreffend, da das Telefonbuch unvollständig ist und die fehlenden Personen keine Zufallsauswahl darstellen, sondern systematisch fehlen. Ebenso ungeschickt ist es, sich zu einer bestimmten Zeit an einen bestimmten Ort, beispielsweise die Fußgängerzone, zu begeben und dort Passanten zu befragen. Macht man das tagsüber, dann befragt man vornehmlich Rentner, Arbeitslose, Schüler, Studenten und Touristen – keine wirklich repräsentative Stichprobe.

**Eingeschränkte Zufallsauswahl.** Hat man keine Liste aller Populationsmitglieder, sondern nur eines Teils (z. B. der eigenen Stadt oder Universität), spricht man von einer eingeschränkten Zufallsauswahl. In diesem Falle ist die Durchführung zwar praktikabler, dafür ist die Generalisierbarkeit ein wenig eingeschränkt.

**Anfallende Stichprobe.** In den meisten Fällen untersucht man einfach diejenigen Personen, die sich freiwillig zur Teilnahme an einem Experiment melden, etwa im Institut auf einen Aushang hin. Dies wäre sicher völlig unangemessen zur Vorhersage der nächsten Bundestagswahlergebnisse. Jedoch ist man in der Experimentellen Psychologie an allgemeingültigen Gesetzmäßigkeiten interessiert, die für nahezu alle Menschen gelten. In diesem Fall stellt die anfallende Stichprobe eine praktikable Wahl dar.

## 2.1.2 Ausschluss von Versuchspersonen

Es ist ein schwerer Verstoß gegen die wissenschaftliche Ethik, Versuchspersonen von der Auswertung auszuschließen, weil ihre Ergebnisse nicht hypothesenkonform sind. Allerdings gibt es Situationen, in denen der Ausschluss von Teilnehmern vom Experiment (bzw. deren Daten von der Auswertung) durchaus zulässig und notwendig ist.

Erfüllt eine Versuchsperson nicht alle relevanten Voraussetzungen für das Experiment, dann darf/kann sie nicht mitmachen. Beispielsweise ist eine farbfeldsichtige Person sicher keine gute Versuchsperson für ein Experiment zur Farbwahrnehmung. Ebenso verlangen Experimente mit lexikalischer Entscheidungsaufgabe nach Teilnehmern mit muttersprachlichen Kenntnissen. Auch in anderen Fällen müssen Versuchspersonen ausgeschlossen werden; weiß eine Person etwa, dass es in einem Experiment zum inzidentellen Lernen später einen Behaltenstest geben wird, dann wäre es für diese Person eben kein inzidentelles Lernen mehr. In all diesen Fällen kann man die Versuchsperson schon *vor* der Teilnahme ausschließen.

Stellt man erst während des Experimentes fest, dass die Versuchsperson nicht macht, was sie laut Instruktion machen soll, dann muss man diese Versuchsperson gegebenenfalls *nach* der Teilnahme ausschließen. Gibt es beispielsweise in einem Stroop-Experiment die beiden Bedingungen ›Wort lesen‹ und ›Farbe benennen‹, die Versuchsperson liest aber immer das Wort und benennt nie die Farbe, verhält sie sich nicht instruktionsgemäß und ihre Daten werden bei der Auswertung nicht beachtet. Schwieriger ist der Fall, wenn die Versuchsperson nur manchmal nicht macht, was sie laut Instruktion soll. Angenommen, man führt ein Experiment zum Kopfrechnen durch, die Versuchsperson soll alle Additionsaufgaben mit einstelligen Operanden lösen, also etwa  $3 + 4$ ,  $6 + 8$  etc. Nun stellt man nach dem Experiment fest, dass die Versuchsperson eine gewisse Anzahl von Fehlern gemacht hat. Bei dieser Art von Aufgabe können Fehlerprozentätze im Bereich bis ca. 10 % als normal angesehen werden. Hat die Versuchsperson jedoch deutlich mehr Fehler gemacht, dann kann sie entweder nicht richtig rechnen und erfüllt somit nicht die Voraussetzungen, oder sie war gelangweilt, unaufmerksam, böswillig, was auch immer, sie hat jedenfalls nicht gemacht, was sie sollte. In diesem Fall schließt man die Versuchsperson von der Auswertung aus. Wichtig ist, dass das Kriterium zum Ausschluss *vorher* festlegt wurde und nicht erst, nachdem man sich die Daten angeschaut hat.

## 2.1.3 Pflege und Entlohnung von Versuchspersonen

Es bedarf hoffentlich keiner besonderen Erwähnung, dass mit Versuchspersonen höflich und respektvoll umgegangen wird. Dazu gehört auch, dass der Versuchsleiter zur vereinbarten Zeit am vereinbarten Ort ist.

Die Versuchsperson wird vollständig darüber aufgeklärt, was von ihr im Rahmen des Experimentes erwartet wird. Erfordert es die Logik des Experimentes, dass man die Versuchsperson über einen Teil im Unklaren lässt oder sie sogar absichtlich in die Irre

führt, dann ist es zwingend erforderlich, dass sie im Anschluss an das Experiment vollständig aufgeklärt wird – oft mit der Bitte, diese Informationen für sich zu behalten, um nicht andere potenzielle Versuchspersonen schon vor dem Experiment aufzuklären.

Erhält die Versuchsperson eine Entlohnung für die Teilnahme, seien es Versuchspersonenstunden als Studienleistung oder schlicht Geld, dann ist ihr das Zugesagte auch auszuhändigen. Eine besondere Situation entsteht dann, wenn eine Versuchsperson zu mehreren Sitzungen ins Labor kommen muss. Hier bietet es sich an, die Bezahlung erst nach Abschluss des kompletten Experimentes in Aussicht zu stellen. Entscheidet die Versuchsperson nach der ersten Sitzung, dass sie nicht weitermachen will, so wird sie für die erste Sitzung bezahlt. Entscheidet sie nach der ersten Sitzung jedoch, an dem gesamten Experiment teilzunehmen, dann wird sie für alle Sitzungen tatsächlich erst am Ende bezahlt. Dies erscheint als akzeptabler Kompromiss zwischen der Freiheit der Versuchsperson, jederzeit die weitere Teilnahme zu beenden und dem Bedürfnis des Versuchsleiters, vollständige Daten zu erheben.

#### 2.1.4 Wie viele Versuchspersonen?

Die Frage nach der Menge der Versuchspersonen ist eine der beliebtesten im Rahmen des Experimentalpsychologischen Praktikums. Die Standardantwort darauf lautet: Es kommt darauf an. Die Zahl der benötigten Versuchspersonen hängt von der sogenannten Irrtumswahrscheinlichkeit  $\alpha$ , der gewünschten Teststärke (definiert als  $1 - \beta$ ) und der Effektgröße  $ab$  (für statistische Grundkonzepte siehe Abschnitt 3.2). Sind diese Werte bekannt bzw. hat man sie entsprechend gewählt, so lässt sich die Zahl der benötigten Versuchspersonen einfach bestimmen, entweder mit Hilfe eines Tabellenbuches (Cohen, 1988) oder mit einem Statistikprogramm (z. B. G\*Power).

#### 2.1.5 Viele Versuchspersonen oder viele Replikationen?

In der empirischen Psychologie lässt sich leider der Trend feststellen, dass eine große Zahl von Versuchspersonen als Qualitätskriterium angesehen wird: Je mehr Versuchspersonen, desto besser war die Untersuchung. Wirft man jedoch einen Blick in qualitativ hochwertige Publikationen, so stellt man fest, dass dort oft mit wenigen, hochgeübten Versuchspersonen gearbeitet wird. Es stellt sich die Frage, welches Vorgehen das bessere ist. Um die Antwort vorwegzunehmen: Viele Replikationen pro Versuchsperson, und damit stabile Lokationsschätzer pro Versuchsperson, sind die bessere Wahl. Die folgenden Abschnitte werden dies begründen. Erneut sei für statistische Grundkonzepte auf Abschnitt 3.2 bzw. ein geeignetes Statistikbuch verwiesen.

**Mittelwertbildung.** Bei fast allen Publikationen werden nicht die Daten der einzelnen Versuchspersonen, sondern über Versuchspersonen hinweg gemittelte Werte (oder andere Maße der zentralen Tendenz) berichtet. Mitteln ist unter allgemeinpsychologischen Gesichtspunkten nur dann gerechtfertigt, wenn die Ergebnisse für alle Personen

einen ähnlichen Verlauf zeigen. Um dies feststellen zu können, benötigt man verlässliche Daten pro Person. Erhebt man beispielsweise Reaktionszeiten, so ist eine einzelne Reaktionszeit pro experimenteller Bedingung kein verlässliches Datum. Aus diesem Grunde wird man in jeder experimentellen Bedingung pro Person viele Reaktionszeiten erheben, um auf diese Weise zu zuverlässigen Daten zu gelangen. Hat man diese zuverlässigen Daten pro Person, kann man beurteilen, ob über Personen hinweg vergleichbare Reaktionszeitverläufe zu beobachten sind. Ist das der Fall, so kann man über Personen mitteln. Hieraus folgt zwangsläufig, dass man viele Beobachtungen pro Person benötigt.

**Teststärke.** Häufig wird als Argument für »viele Versuchspersonen« (und gleichzeitig wenig Replikationen pro Person und Bedingung) die Teststärke  $(1 - \beta)$  angeführt: Mit vielen Versuchspersonen erhalte man eine größere Teststärke. Das ist zwar im Prinzip korrekt, es wird dabei jedoch der Einfluss der Zahl der Replikationen auf die Teststärke übersehen. Um dies zu veranschaulichen stelle man sich eine einfache Situation vor: In jedem *trial* wird entweder ein Licht oder ein Ton präsentiert. Die Aufgabe der Versuchsperson besteht darin, so schnell wie möglich eine Taste zu drücken (Einfachreaktion). Erhoben wird die Reaktionszeit vom Beginn der Reizpräsentation bis zum Drücken der Taste.

Unterstellt man nun bestimmte Effektgrößen und wählt  $\alpha$  und  $\beta$ , so kann man leicht ausrechnen, wie viele Versuchspersonen benötigt werden, um mit der gewählten Power einen entsprechenden Effekt entdecken zu können. Bei Gültigkeit der Normalverteilungsannahme würden die Daten mittels *t*-Test für abhängige Messungen ausgewertet werden. Variiert man die Effektgrößen und die Zahl der Replikationen (*r*) pro Versuchsperson und Bedingung, ergeben sich die in Tabelle 2.1 dargestellten Werte.

Wie man sieht, nimmt mit zunehmender Anzahl von Replikationen der notwendige Stichprobenumfang drastisch ab. Dies ergibt sich aus der Tatsache, dass die gemittelten

Tabelle 2.1

*Notwendige Anzahl von Versuchspersonen in Abhängigkeit von der Anzahl der Replikationen (r), um mit Power  $1 - \beta = 0.95$  einen Effekt der Größe  $\frac{\mu - \mu_0}{\sigma_x}$  zu entdecken,  $\alpha = 0.05$*

Replikationen	Effektgröße $\frac{\mu - \mu_0}{\sigma_x}$						
	1.0	0.75	0.5	0.4	0.3	0.2	0.1
<i>r</i> = 1	16	26	54	84	147	327	1300
<i>r</i> = 2	9	14	28	43	75	165	652
<i>r</i> = 5	5	7	13	19	31	67	262
<i>r</i> = 10	4	5	8	11	17	35	132
<i>r</i> = 20	4	4	5	7	10	19	67
<i>r</i> = 50	3	3	4	4	6	9	29

Replikationen pro Bedingung und Versuchsperson eine viel geringere Streuung besitzen, nämlich  $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{r}}$ . Durch die verringerte Streuung steigen die Effektgrößen. Vergleicht man etwa bei  $\frac{\mu - \mu_0}{\sigma_x} = 0.5$  die Fälle  $r = 1$  (ein Wertepaar pro Person) mit  $r = 10$  (10 Wertepaare pro Person), so wird deutlich, dass man statt 54 nur noch 8 Versuchspersonen benötigt.

**Übungseffekte.** Die letzten Abschnitte haben deutlich gemacht, dass viele Replikationen pro Person und Bedingung wünschenswert sind. Allerdings könnte man einwenden, dass – um das letzte Beispiel erneut aufzugreifen – 54 Versuchspersonen mit je einem Wertepaar effizienter seien als acht Versuchspersonen mit je zehn Wertepaaren, da insgesamt weniger Daten erhoben werden müssen. Abgesehen von dem Aufwand, den das Einwerben und Instruieren vieler Versuchspersonen darstellt, treten bei mehrfachen Replikationen erwünschte Übungseffekte auf: Die Versuchspersonen werden nicht nur schneller (kürzere Reaktionszeiten), sondern die Variabilität der Reaktionszeiten nimmt auch ab. Dies wirkt sich positiv auf die Teststärke aus. Ein Beispiel soll dies illustrieren. Es soll im Folgenden die Situation  $r = 10$  Replikationen pro Bedingung aus Tabelle 2.1 betrachtet werden. Relativ schnell wird die Versuchsperson im Laufe dieser Replikationen besser, mit sinkenden Reaktionszeiten und kleiner werdender Varianz. Eine konservative Annahme unterstellt eine Reduktion der Standardabweichung von den ersten beiden zu den letzten acht Replikationen auf  $\frac{3}{4}$ . Bildet man nun den Mittelwert lediglich über die Replikationen 3–10, so ergibt sich  $\sigma_{\bar{x}} = \frac{0.75\sigma_x}{\sqrt{8}}$ . Berechnet man erneut die notwendigen Stichprobenumfänge, so erhält man die in Tabelle 2.2 angegebenen Werte. Es wird offensichtlich, dass Replikationen nicht nur die notwendigen Stichprobenumfänge deutlich verringern, sondern bei realistisch kleinen Effektgrößen die notwendige Gesamtzahl von Beobachtungen kleiner ausfällt als ohne Replikationen.

Wann immer es also möglich ist, sollte mit vielen Replikationen pro Versuchsperson und Bedingung gearbeitet werden. Dies ist nicht nur vom Gesamtaufwand her effizienter, sondern die Teststärke ist bei realistischen Annahmen auch größer.

Tabelle 2.2

*Notwendige Anzahl von Versuchspersonen, um mit Power  $1 - \beta = 0.95$  einen Effekt der Größe  $\frac{\mu - \mu_0}{\sigma_x}$  zu entdecken,  $\alpha = 0.05$ . Von  $r = 10$  Replikationen werden nur die Replikationen 3 bis 10 berücksichtigt, deren Standardabweichung soll  $0.75\sigma_x$  betragen*

Replikationen	Effektgröße $\frac{\mu - \mu_0}{\sigma_x}$						
	1.0	0.75	0.5	0.4	0.3	0.2	0.1
$r = 1$	16	26	54	84	147	327	1300
$r = 8$ von 10	4	4	6	8	13	25	94

## 2.1.6 Stichproben parallelisieren

Gelegentlich möchte man parallele Stichproben untersuchen. Damit ist gemeint, dass sich die Personen in zwei (oder mehr) Gruppen hinsichtlich eines oder mehrerer Merkmale ähnlicher sind als per Zufallsauswahl. Ein natürlicher Fall paralleler Stichproben liegt vor, wenn »Illinge« eineiiger Zwillingspaare zwei getrennten Gruppen zugewiesen werden. Somit hat jede Person in Gruppe 1 eine zugehörige Person in Gruppe 2, nämlich ihren Zwillingspartner. Nun sind Zwillinge nicht die üblichen Versuchspersonen und man muss deshalb meist selbst für geeignete Paare (oder Tripel etc.) sorgen. In aller Regel bildet man solche parallelen Stichproben, um die Wirkung möglicher Störvariablen auszuschalten. Angenommen, man wollte ein neues Lehrmodul evaluieren, dann könnte man zwei Stichproben ziehen, Gruppe 1 wird mit dem bisherigen Modul unterrichtet, Gruppe 2 wird mit dem neuen Modul unterrichtet und man vergleicht im Anschluss daran den Lernerfolg. Hat man die Vermutung, dass die Intelligenz der Versuchspersonen eine moderierende Wirkung auf den Lernerfolg hat, so könnte man die beiden Stichproben nach der Intelligenz parallelisieren. Hierfür erhebt man noch vor einer Gruppenzuordnung mit einem geeigneten Instrument die Intelligenz von allen Versuchspersonen. Danach werden die Personen nach Intelligenztestwerten in eine Rangreihe gebracht, man erhält also eine nach IQ-Punkten geordnete Stichprobe mit  $x_{(1)}, x_{(2)}, \dots, x_{(n)}$  (der Index in Klammern spiegelt hier die Ordnung wider). Diese Personen werden nun abwechselnd auf die beiden Gruppen aufgeteilt, sodass jeweils Paarlinge entstehen *und* die mittlere Intelligenz in beiden Gruppen etwa gleich groß ist. Das Vorgehen bei der Aufteilung ist in Tabelle 2.3 dargestellt.

Man beachte, dass nicht einfach die Personen 1, 3, 5, ... in Gruppe 1 und die Personen 2, 4, 6, ... in Gruppe 2 gesteckt wurden, denn das hätte einen systematischen Mittelwertsunterschied zur Folge. Weiterhin ist zu berücksichtigen, dass bei parallelen Stichproben inferenzstatistische Verfahren für abhängige Messungen zu verwenden sind.

Tabelle 2.3

*Parallelisieren einer Stichprobe*

	Personenindex nach IQ-Rangreihe							
Gruppe 1	1	4	5	8	9	12	13	...
Gruppe 2	2	3	6	7	10	11	14	...

## 2.2 Geräte

Von den »Standardgeräten« Papier, Bleistift und Stoppuhr einmal abgesehen ist der PC heutzutage wohl das Experimentierequipment schlechthin. Externe *timer*, Tachistoskop oder Diaprojektor – *state of the art* noch bis in die 80er Jahre des letzten Jahrhun-

derts – sind in experimentalpsychologischen Laboren fast völlig ausgestorben. Der Umstand der Allgegenwärtigkeit des PCs führt leider dazu, dass viele dessen ›Haken und Ösen‹ nicht kennen.

## 2.2.1 Der Monitor

Visuelle Reize, seien es Wörter, Zahlen oder Bilder, werden auf einem Monitor dargestellt – logisch. Heute findet man üblicherweise zwei Monitortypen, die herkömmlichen Röhrenmonitore (CRT, *cathode ray tube*) und die moderneren Flachbildschirme, meist LC-Displays (*liquid crystal display*). Da die technischen Unterschiede für das Experimentieren entscheidende Konsequenzen haben, soll im Folgenden etwas genauer darauf eingegangen werden.

**Der Röhrenmonitor.** Bei einem Röhrenmonitor handelt es sich im Prinzip um eine Braun'sche Kathodenstrahlröhre. Der Kathodenstrahl trifft auf die Rückseite des phosphorbeschichteten Monitorglases und bringt diese Beschichtung zum Leuchten. Bei Farbmonitoren werden drei unterschiedliche Phosphortypen verwendet, die Emissionsmaxima im roten, grünen und blauen Bereich des sichtbaren Spektrums haben. Hat ein Monitor eine Auflösung von z. B.  $1024 \times 768$  Pixel, so ›besitzt‹ er quasi 768 Zeilen, in jeder Zeile gibt es 1024 Pixel. Der Kathodenstrahl beginnt oben links in der ersten Zeile, durchläuft sie, springt zurück in die nächste Zeile, durchläuft sie, bis zur letzten Zeile des Bildschirms. Nun folgt der *vertical retrace*, also der Rücksprung von rechts unten nach links oben. Hat der Monitor eine Bildwiederholrate von z. B. 100 Hz, so geschieht dies 100-mal pro Sekunde, pro Wiederholung werden also 10 ms benötigt. Daraus ergibt sich, dass ein Bildschirminhalt nicht kürzer als eben diese 10 ms (bei 100 Hz Bildwiederholrate) dargeboten werden kann. Aus der Bildwiederholrate ergibt sich außerdem, dass Darbietungszeiten immer nur Vielfache des Frequenzkehrwertes sein können, bei einem 100 Hz Monitor also 10, 20, 30, ... ms. Andere Darbietungszeiten sind technisch nicht möglich.

Im Detail ist die Sache noch etwas komplizierter. Jedes einzelne Pixel wird vom Kathodenstrahl nur für den Bruchteil einer Millisekunde beschossen. Bleibt man bei einer Auflösung von  $1024 \times 768$  Pixel und einer Bildwiederholrate von 100 Hz, so werden während 10 ms  $1024 \times 768$  Pixel beschossen. Für jedes einzelne Pixel bleiben also nur  $\frac{0.01s}{1024 \times 768} \approx 0.013 \mu s$ . Ist das gesamte Bild weiß, so wird das erste Pixel oben links für  $0.013 \mu s$  beschossen. Seine Leuchtdauer wiederum hängt vom verwendeten Phosphor ab. Phosphor P<sub>31</sub> etwa erreicht innerhalb von  $100 \mu s$  sein Emissionsmaximum, um dann in weiteren  $600 \mu s$  auf 10 % abzufallen (Westheimer, 1993). Der in handelsüblichen Monitoren verwendete Phosphor P<sub>22</sub> ist etwas langsamer, insbesondere ist der Abfall für die drei Phosphortypen (rot, grün, blau) unterschiedlich schnell. Physikalisch ist demnach das erste Pixel links oben schon längst wieder ›dunkel‹ bevor das letzte Pixel rechts unten leuchtet. Dennoch nimmt man eine homogene weiße Fläche wahr,

da die Bildwiederholrate in diesem Beispiel deutlich über der Flimmerverschmelzungsfrequenz des menschlichen visuellen Systems liegt (Sperling, 1971). Reduziert man die Bildwiederholrate auf unter 85 Hz, so wird ein Flimmern des Monitors deutlich sichtbar. Dies muss unbedingt vermieden werden.

Obwohl ein einzelnes Pixel also nur für einige Hundert  $\mu\text{s}$  bis maximal ein bis zwei ms leuchtet<sup>1</sup>, ist es nicht möglich, dieses früher als nach 10 ms erneut zum Leuchten anzuregen. Beispielsweise kann ein Farbwechsel eines Pixels von blau auf grün bei einer Bildwiederholrate von 100 Hz frühestens nach 10 ms erfolgen.

**Das LC-Display.** Die Technik von LC-Displays ist grundsätzlich verschieden von Röhrenmonitoren. Jedes Pixel wird quasi direkt angesteuert und leuchtet – oder eben nicht. Dies klingt erst einmal sehr verlockend für uns, könnte man doch so zu fast beliebigen Zeitpunkten beliebige Pixel an- und ausschalten. Leider erlaubt die PC-Technik das nicht. Zwar gibt es bei LC-Displays keinen Kathodenstrahl, der Zeile für Zeile über den Bildschirm läuft, dennoch wird das Bild zeilenweise aufgebaut. Beim LC-Display, ebenso wie beim Röhrenmonitor, liegt die Information, was auf dem Bildschirm erscheinen soll, im Graphikspeicher des Computers. Diese Information wird nun entsprechend der Bildwiederholrate des Monitors ausgelesen. LC-Displays arbeiten gewöhnlich mit nur 60 oder 70 Hz, also wird die Information 60 oder 70 mal pro Sekunde aus dem Graphikspeicher ausgelesen und in ein sichtbares Bild umgewandelt. Im Vergleich zu Röhrenmonitoren sind die minimalen Darbietungszeiten von LC-Displays daher deutlich länger. Weiterhin haben die Flüssigkristalle der LC-Displays minimale Schaltzeiten, was schnelle Bildwechsel erschwert. Schnell bewegte Objekte hinterlassen auf LC-Displays häufig einen Schweif oder Schlieren. Bei aktuellen Displays ist dieses Problem inzwischen etwas geringer geworden. Um zu extrem kurzen Schaltzeiten zu gelangen, haben sich die Displayhersteller eine neue Technik (Overdrive) ausgedacht, bei der an jedem Pixel eine Überspannung angelegt wird. Allerdings muss man dazu vor dem Schaltvorgang wissen, welche Überspannung benötigt wird. Um diese Informationen zu erhalten, werden zukünftige Bildinformationen verrechnet. Dies hat zur Folge, dass die Ausgabe um ein bis drei Displayzyklen, bei den üblichen 60 Hz also bis zu 50 ms, verzögert wird. D. h. die Grafikkarte liefert das Bildsignal, aber der Monitor stellt diese Information erst 50 ms später dar. Diese Verzögerung ist bei normalem Arbeiten unproblematisch, der Monitor ›hinkt‹ der Grafikkarte um eine feste Zeit hinterher. Erfasst man jedoch Reaktionszeiten, schlägt sich diese Verzögerung systematisch nieder: Die Reaktionszeiten werden um diese feste Zeit fehlgemessen. Das mag so lange akzeptabel sein, wie man keine quantitativen Modelle anpasst und durch die systematische Fehlmessung zu unsinnigen Parameterschätzungen kommt. Besonders problematisch wird es, wenn man z. B. visuelle und auditive Reize zeitsynchronisiert darbieten möchte. Statt der vorgesehenen simultanen ergibt sich eine nicht-intendierte zeitversetzte Präsentation – mit gegebenenfalls fatalen Folgen für die Ergebnisinterpretation.

---

<sup>1</sup>Wir haben einen Monitor mit Phosphor P22 vermessen, ein weißes Pixel ist nach ca. 0,5–1,0 ms wieder dunkel.

Schließlich leiden die meisten LC-Displays, insbesondere von Notebooks, auch daran, dass sie einen kleineren Gamut<sup>2</sup> als Röhrenmonitore besitzen, was sie für Farbwahrnehmungsexperimente unbrauchbar macht. Allerdings werden neuerdings immer öfter sogenannte Wide-Color-Gamut und RGB-LED-Backlight Bildschirme angeboten. Diese LC-Displays weisen einen Farbraum auf, der sogar deutlich größer als der Adobe-RGB-Farbraum ist (Porteck & Wirtgen, 2009). In Verbindung mit einer durchzuführenden Kalibrierung und Profilierung wären diese Bildschirme ideal, wenn sie nicht gleichzeitig auch über die Overdrive-Technik verfügten und höhere Bildwiederholraten hätten.

LC-Displays mögen praktisch für den Schreibtisch und mittlerweile sogar für die Bildbearbeitung sein, für zeitkritische Experimente sind sie problematisch bis hin zu unbrauchbar.

**Synchronisation mit dem vertical retrace.** Bei der Darbietung visueller Reize ändert sich im Laufe eines *trials* der Bildschirminhalt. Ein Durchgang beginnt etwa mit einem leeren Bildschirm, dann wird der Hinweisreiz präsentiert, danach erfolgt die Darbietung des imperativen Reizes und die Versuchsperson muss reagieren. Anschließend ist der Bildschirm wieder leer. Solch eine Abfolge von Bildschirminhalten muss nicht nur zeitlich exakt sein, die Bildschirmwechsel müssen auch mit dem *vertical retrace* synchronisiert werden. Nur während des *vertical retrace* lassen sich Bildschirminhalte ändern, ohne dass auf dem Monitor ein kurzes Flackern entsteht. Verwendet man ein kommerzielles Experimentalsteuersystem, dann ist zu hoffen, dass die Synchronisation mit dem *vertical retrace* implementiert ist. Programmiert man seine Experimente von Hand, muss man selbst darauf achten, diese Synchronisation zu realisieren.

**Farbkalibrierung.** Für viele experimentelle Situationen sind die Standardeinstellungen eines Monitors hinsichtlich Helligkeit, Kontrast, Gamma und Farbraum völlig ausreichend, man achtet per Augenmaß darauf, dass ein »angenehmes« Bild geliefert wird. Dazu gehört, dass Kontrast und Helligkeit so gewählt werden, dass es in abgedunkelten Laborräumen nicht zur Blendung führt und alle Reize gut sichtbar sind. Weiterhin ist die Farbtemperatur so zu wählen, dass ein neutrales Grau gegeben ist. Präsentiert man beispielsweise Wörter in schwarz auf weißem Hintergrund, dann sollte der Hintergrund tatsächlich weiß und nicht rosa erscheinen.

Arbeitet man jedoch im Bereich der Farbwahrnehmung oder auch nur mit farbigen Reizen, ist eine regelmäßige und exakte Kalibrierung des Monitors nötig. Professionelle Graphikmonitore aus den Bereichen Bild- und Videoverarbeitung bringen entsprechende Testgeräte und Software schon mit, z. T. ist die Elektronik sogar in den Monitor integriert. Für Standardmonitore gibt es handelsübliche Kolorimeter inkl. Software zu erschwinglichen Preisen, eine kleine Übersicht findet sich bei Porteck (2009). Im Wesentlichen geht es darum, den Farbraum zu vermessen, das Gamma für den Rot-, Grün-

---

<sup>2</sup>Der Gamut ist der Körper im Farbraum, welcher mit dem Monitor durch innere Farbmischung nachgestellt werden kann.

und Blaukanal zu bestimmen und die Farbtemperatur des Weißpunktes festzulegen. Diese Informationen werden dann dem Graphikkartentreiber in Form von Farbprofilen zur Verfügung gestellt. Auf identisch kalibrierten Monitoren erscheint eine rote Tomate (annähernd) gleich rot, sofern der Farbton innerhalb des Farbraumes liegt.

## 2.2.2 Reaktionserfassung

Meist möchte man nicht nur Reize präsentieren, sondern auch die Reaktion der Versuchsperson erfassen. In vielen Fällen soll die Versuchsperson eine (von mehreren) Taste(n) drücken. In Abhängigkeit von der verwendeten abhängigen Variablen ergeben sich hierbei verschiedene Probleme.

**Qualität der Reaktion.** Angenommen, in einem Gedächtnisexperiment werden in der Lernphase 20 Bilder präsentiert, nach dem Behaltensintervall folgt die Testphase. In der Testphase werden 20 alte und 20 neue Bilder präsentiert, insgesamt also 40. Bei jedem Bild soll die Versuchsperson per Tastendruck entscheiden, ob sie das präsentierte Bild für alt oder neu hält. Es stehen also zwei unterschiedliche Tasten zur Verfügung und man ist nur daran interessiert, ob die Versuchsperson die eine oder andere Taste gedrückt hat. In diesem Fall ist es fast egal, welche Art von Tasten man verwendet, am einfachsten wären zwei unterschiedliche Tasten auf der normalen Computer-Tastatur oder auch die beiden Tasten einer herkömmlichen Computer-Maus.

**Reaktionszeiterfassung.** Eine völlig andere Situation ergibt sich, wenn man an der Erfassung der Reaktionszeiten interessiert ist. Bezogen auf das obige Gedächtnisexperiment könnte neben der Qualität der Antwort zusätzlich der Zeitbedarf der ›alt‹ bzw. ›neu‹ Antworten von Interesse sein. Die naheliegende Lösung, auch hier zwei Tasten der normalen Tastatur oder die Tasten einer Computer-Maus zu verwenden, ist leider schlecht. Der Druck auf eine Taste der Tastatur oder der Maus wird nicht sofort an den Computer weitergeleitet bzw. von diesem ausgewertet, sondern es vergeht eine gewisse Zeit. Diese Zeitverzögerung ist unglücklicherweise nicht konstant und außerdem von den technischen Eigenschaften der Peripherieanbindung abhängig (Beringer, 1992b; Crosbie, 1990; Segalowitz & Graves, 1990). Kurz gesagt: Tastatur und Maus sind unbrauchbar.

Als Alternative bieten sich externe Tastaturen an, die über die serielle oder parallele Schnittstelle eingebunden werden. In diesen externen Tastaturen kann man Mikroschalter mit Entprellung verwenden und die Abfrage der Schnittstelle kann mit hoher zeitlicher Auflösung erfolgen. Solchen Eingabemedien ist in jedem Fall der Vorzug zu geben, wenn Reaktionszeiten erfasst werden sollen (vgl. Voss, Leonhart & Stahl, 2007).

Es soll nicht verschwiegen werden, dass zukünftig vermutlich immer mehr Computer ohne parallele und/oder serielle Schnittstelle ausgeliefert werden. Dies stellt ein ernsthaftes Problem dar: Die Ansteuerung des USB-Ports ist bei weitem komplizierter

als die Ansteuerung der seriellen oder parallelen Schnittstelle. Alternative Experimentierboards mit mehreren analogen und digitalen Eingängen, die via USB an den Rechner angebunden werden, sind häufig viel zu langsam, als dass man sie für ernsthaftes Experimentieren einsetzen könnte – Ausnahmen bilden hochpreisige Modelle.

Neben dem geeigneten Eingabegerät zur Reaktionserfassung benötigt man zweifellos eine Uhr. Jeder Computer stellt Funktionen zur Zeitmessung zur Verfügung – leider in ungenügender Genauigkeit (siehe aber Ulrich & Giray, 1989). Zu Zeiten von MS-DOS hat man deshalb den Timerbaustein 8253 so umprogrammiert, dass er eine millisekundengenaue Zeitmessung ermöglichte (vgl. Beringer, 1992a; Graves & Bradley, 1987, 1991). Nach der Einführung des Pentium-Prozessors hatte man die Möglichkeit, die Prozessorticks auszulesen. Dies erlaubte eine Messgenauigkeit weit jenseits dessen, was man als Experimentalpsychologe benötigt, erforderte aber auch Assembler-Kenntnisse. In der Windows-Welt steht einem ein hochpräziser *timer* mit den Funktionen `QUERYPERFORMANCEFREQUENCY()` und `QUERYPERFORMANCECOUNTER()` zur Verfügung, ohne dass man sich mit Assembler in die Tiefen des Timerbausteins oder Prozessors begeben muss. Allerdings bedeutet ein präziser *timer* unter Windows noch keine präzise Zeitmessung (siehe Abschnitt 2.2.3).

**Spracherfassung.** Gelegentlich will man die Sprache oder andere Äußerungen einer Versuchsperson erfassen. Geht es tatsächlich darum, die Äußerungen oder den Gesang aufzuzeichnen, dann benötigt man ein qualitativ hochwertiges Mikrofon und einen entsprechenden Mikrofoneingang am Sound-Subsystem des Computers. Etwas komplizierter wird es, wenn man durch den Beginn des Sprachsignals eine Reaktionszeituhr stoppen möchte. Beispielsweise soll die Versuchsperson in einem Experiment einfache Kopfrechenaufgaben der Art  $3+4$  oder  $6 \times 7$  lösen. Die korrekte Antwort soll ausgesprochen werden und man ist an der Lösungszeit, gemessen von der Aufgabenpräsentation bis zum Beginn der Antwortaussprache, interessiert. In diesem Fall ist die Methode der Wahl die Verwendung eines Stimmschlüssels (*voicekey*).

**Stimmschlüssel.** Bei einem Stimmschlüssel handelt es sich um einen Schwellwertschalter. Die Versuchsperson spricht in ein Mikrofon, dieses ist an den Stimmschlüssel angeschlossen, welcher seinerseits über die serielle oder parallele Schnittstelle mit dem Computer verbunden ist. Überschreitet der Schalldruckpegel des Signals die gewählte Schwelle, schaltet der Stimmschlüssel von Null auf Eins (oder umgekehrt).

Dieser Wechsel des Signalpegels wird vom Computer detektiert und die Reaktionszeituhr kann gestoppt werden. Da es sich bei dem akustischen Signal um eine Schwingung handelt, ändert sich das analoge Signal des Mikrofons ständig. Deshalb versieht man den Stimmschlüssel mit einem hysteresebefahenen Schaltverhalten, damit kein fortwährender Pegelwechsel entsteht (vgl. Abbildung 2.1).

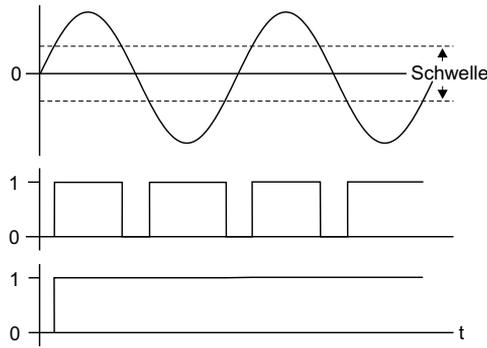


Abbildung 2.1. Oben: Signal und gewählte Schwelle. Mitte: Schaltverhalten eines einfachen Schwellwert Schalters. Unten: Schaltverhalten eines hysteresesehefteten Schalters.

### 2.2.3 Betriebssystem

War bislang vom PC die Rede, dann war damit in aller Regel ein Standard-PC gemeint, also ein gewöhnlicher Rechner, auf dem kein Apfel-Logo klebt. Lässt man die Mac-Welt einmal außer Acht, dann hat man beim Standard-PC immer noch die Wahl des Betriebssystems. Die beiden verbreiteten Varianten Linux und Windows (NT, 2000, XP, Vista, Windows 7, alle vorherigen Varianten sind indiskutabel) sind im Prinzip zur Experimentalsteuerung denkbar schlecht geeignet. Bei beiden handelt es sich um Multitasking-Betriebssysteme, bei denen eine Vielzahl von Prozessen im Hintergrund laufen, die nicht abzuschalten sind. Erfordert das Experiment jedoch ein präzises *timing* der Reizdarbietung und Reaktionserfassung, dann ist es ausgesprochen ungünstig, wenn das Betriebssystem zwischenzeitlich beschließt, mal eben andere Aufgaben abzuarbeiten. Eigentlich bräuchte man zur sauberen Experimentalsteuerung ein Echtzeit-Betriebssystem.

Vor noch nicht allzu langer Zeit war es üblich, Experimente nicht unter Windows bzw. Linux laufen zu lassen, sondern MS-DOS zu verwenden. In nicht wenigen Labors wird auch heute noch unter DOS experimentiert, so hat man die Maschine problemlos im Griff, das Betriebssystem funkt einem nicht dazwischen und die Programmierung ist relativ einfach. Allerdings ist die DOS-Welt wirklich nicht mehr auf der Höhe der Zeit hinsichtlich Gerätetreiber, Soundausgabe, USB-Schnittstelle etc. Wer noch eine funktionierende ältere DOS-Maschine zum Experimentieren benutzt, der kann sie guten Gewissens weiterhin betreiben. Bei neuen Rechnern muss man sich mit Windows (oder Linux) arrangieren. Dies macht einige zusätzliche Kontrollen nötig. Arbeitet man beispielsweise unter Windows, so genügt es nicht, das gesamte *trial-timing* vorzugeben, sondern man muss jedes Ereignis auch zeitlich erfassen, um nach einem *trial* entscheiden zu können, ob die Zeitvorgaben eingehalten wurden. Die Chance, dass dies der Fall ist, ist mit schnellen Rechnern deutlich besser, d. h. ein aktueller Rechner (2009/2010) ist unter Windows XP bzw. 7 (aber nicht Vista) hinreichend schnell, um damit experi-

mentieren zu können. Bei einem in die Jahre gekommenen Computer muss dies nicht gegeben sein (vgl. Myors, 1999).

## 2.2.4 Programmsysteme

Zum Verfassen eines Textes verwendet man ein geeignetes Textprogramm, zur Datenauswertung ein gutes Statistikprogramm. Auch zur Experimentalsteuerung existieren entsprechende Programmsysteme, es gibt jedoch Gründe, die gegen deren Verwendung sprechen. Ein einfacher Grund lautet: Die guten Experimentalsteuerprogramme sind teuer. Ein zweiter ausschlaggebender Grund gegen die Verwendung solcher Baukastensysteme ist, dass man unvorhergesehen an deren Grenzen stößt.

Ein typisches Problem bei der Experimentalsteuerung besteht darin, die Abfolge der experimentellen Bedingungen zu kontrollieren. Die Varianten ›in vorgegebener Reihenfolge‹ oder ›in zufälliger Reihenfolge‹ lassen sich bei allen Programmsystemen recht einfach realisieren. Die Standardsituation beim Experimentieren ist jedoch eine Zufallsabfolge mit Einschränkung. Beispielsweise will man bei einem Experiment zum Kopfrechnen alle einstelligen Additionsaufgaben verwenden. Diese Aufgaben sollen in zufälliger Reihenfolge präsentiert werden, ohne dass aufeinanderfolgende Aufgaben gemeinsame Operanden und Ergebnisse besitzen. Auf die Aufgabe  $3 + 4$  darf weder  $4 + 5$  oder  $8 + 3$  (gemeinsamer Operand), noch  $7 + 2$  ( $7$  ist das Ergebnis von  $3 + 4$ ), noch  $1 + 2$  (das Ergebnis entspricht einem vorherigen Operanden) folgen. Solche Einschränkungen in Zufallsabfolgen lassen sich mit Experimentalsteuerprogrammen kaum realisieren. Das bedeutet aber, dass man die Bedingungsabfolge extern generieren und der Experimentalsteuerung zuweisen muss. Werden dafür eigene Programme geschrieben, ist der Schritt zur vollständigen Programmierung des Experimentes nicht mehr allzu groß (siehe auch Abschnitt 7.2).

## 2.3 Material

### 2.3.1 Grundsätzliches

Das in einem Experiment eingesetzte Material ist von entscheidender Bedeutung, insbesondere im Hinblick auf eine adäquate Operationalisierung der unabhängigen Variablen. Triviale Fehler lassen sich problemlos vermeiden. Will man in einem Experiment zur visuellen Suche z. B. rote und blaue Rechtecke auf dem Computer-Monitor darbieten, genügt eine einfache Betrachtung, um festzustellen, ob die Rechtecke tatsächlich rot und blau sind. Verwendet man die Reize hingegen im Rahmen eines Experimentes zur Farbwahrnehmung, ist sehr viel mehr Aufwand bei der Beurteilung der Reizqualität zu treiben. Ähnlich verhält es sich bei der Darbietung akustischer Reize. Handelt es sich um einen Hinweisreiz, dann ist es vermutlich egal, ob er 440 oder 445 Hz aufweist. Ist die Frequenz von entscheidender Bedeutung, muss man tatsächlich prüfen, ob der

Letztlich gilt es zu klären, wie viel Zeit von der Reaktion der Versuchsperson bis zum Beginn des nächsten Durchgangs vergehen soll (*inter-trial-interval*, ITI). Das ITI sollte so gewählt werden, dass der gesamte Ablauf für die Versuchsperson angenehm ist. Ist er zu schnell, dann wird das Experiment unnötig anstrengend, im umgekehrten Fall langweilt sich die Versuchsperson sehr schnell. Beides gilt es zu vermeiden. In den meisten Experimenten sind 1–2 Sekunden für das *inter-trial-interval* geeignet. Auch hier stellt sich die Frage nach konstanter vs. variabler Dauer. Ist die Vorperiode schon zeitlich variabel, dann kann man das ITI problemlos konstant wählen. Ist die Vorperiodendauer jedoch konstant, dann sollte das ITI variabel gewählt werden. In Abbildung 2.10 ist der vollständige zeitliche Ablauf eines Durchgangs im Einfachreaktionsparadigma dargestellt.

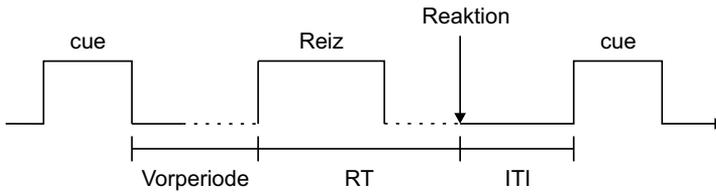


Abbildung 2.10. Zeitlicher Verlauf eines vollständigen Durchgangs mit Hinweisreiz, variabler Vorperiode, konstanter Reizdauer und *inter-trial-interval* im Einfachreaktionsparadigma.

## 2.7.2 Wahlreaktion

Beim Einfachreaktionsparadigma gibt es zwar unterschiedliche Reize bzw. experimentelle Bedingungen, aber nur eine einzige mögliche Reaktion. Deshalb kann eine Versuchsperson auch keinen Fehler begehen – vom Nicht-Drücken einmal abgesehen. Vernünftigerweise interessiert man sich für den Zeitbedarf der Reaktionen. Im Falle des Wahlreaktionsparadigmas gibt es ebenfalls mindestens zwei unterschiedliche Reize bzw. experimentelle Bedingungen, aber mindestens zwei unterschiedliche Antwortmöglichkeiten. Damit können als abhängige Variable sowohl die Reaktionszeit als auch die Qualität der Antwort betrachtet werden. Das Wahlreaktionsparadigma ist das mit Abstand häufigste Paradigma der Experimentellen Psychologie. Da viele Sachverhalte auch auf andere Paradigmen angewendet werden können, soll erneut etwas ausführlicher auf die verschiedenen Aspekte eingegangen werden.

Die einfachste Form eines Wahlreaktionsexperimentes ist das *Ein-Intervall-Design*: Innerhalb eines Durchgangs wird genau ein Reiz präsentiert, die Versuchsperson soll instruktionsgemäß eine von mindestens zwei Antworttasten drücken. In vielen Fällen sind die unterschiedlichen Reize bzw. experimentellen Bedingungen klar unterscheidbar und die Aufgabe der Versuchsperson ist es, diese voneinander zu diskriminieren. Solche Experimente werden als *Ja-Nein-Experimente* (auch Diskriminations- oder

Wiedererkennens-Experimente) bezeichnet, selbst wenn die tatsächlichen Antworten nicht »ja« oder »nein« lauten, sondern z. B. »hell« vs. »dunkel«. Handelt es sich jedoch um die Bedingungen »gerade wahrnehmbarer Reiz« vs. »kein Reiz« und die Aufgabe der Versuchsperson besteht darin zu entscheiden, ob ein Reiz präsentiert wurde oder nicht, spricht man von einem *Entdeckungsexperiment*. Allerdings wird auch im Kontext von Entdeckungsexperimenten häufig von Ja-Nein-Experimenten gesprochen (Macmillan & Creelman, 1991).

**Zeitlicher Ablauf.** Der zeitliche Ablauf eines Durchgangs ist in wesentlichen Teilen identisch zum Zeitverlauf in Einfachreaktionsexperimenten. Erneut muss man sich zwischen einer festgelegten Reizdauer und einer durch den Tastendruck der Versuchsperson beendeten Reizdarbietung entscheiden (vgl. Abbildung 2.8). Auch für die *cue*-Eigenschaften ergeben sich die gleichen Empfehlungen wie beim Einfachreaktionsparadigma: Der Hinweisreiz sollte gut sichtbar bzw. hörbar sein, er sollte verschieden vom imperativen Reiz sein und er sollte für eine Dauer von ca. 200–500 ms dargeboten werden. Der wesentliche Unterschied zwischen Einfach- und Wahlreaktionsparadigma besteht in der Konstruktion der Vorperiode. Da die Versuchsperson auf jeden Fall das Erscheinen des imperativen Reizes abwarten muss, um entscheiden zu können, welche Taste zu drücken ist, kann man problemlos eine konstante Vorperiode wählen. Empfehlenswert sind Vorperiodendauern zwischen 200 und 500 ms.

Die mögliche Alternative auf die Vorperiode komplett zu verzichten, ist nicht zu empfehlen. Zum einen entsteht durch das unmittelbare Aufeinanderfolgen von Hinweis- und Zielreiz ein sehr schneller Bildwechsel, der Stress induzieren kann. Zum anderen kann durch diesen Bildwechsel eine Scheinbewegung erzeugt werden, die es auf jeden Fall zu vermeiden gilt.

Gelegentlich werden sogenannte Antwortfenster verwendet, d. h. der Versuchsleiter begrenzt die Zeit, in der die Versuchsperson reagieren kann. Üblicherweise macht man dies in Wahlreaktionsexperimenten, in denen die Fehlerzahl als abhängiges Maß betrachtet wird: Durch die Begrenzung sollen bei der Versuchsperson Zeitdruck erzeugt und somit mehr Fehler provoziert werden (vgl. Abschnitt 3.1.2). Ist man jedoch an Reaktionszeiten interessiert, so ist die Verwendung von Antwortfenstern nicht empfehlenswert. Sie führen zu den gleichen Problemen wie die Anwendung willkürlicher Reaktionszeitintervalle bei der Datenvorauswertung (vgl. Abschnitt 3.1). Diese Kritik gilt im Übrigen in gleichem Maße für das Einfachreaktionsparadigma.

Letztlich muss man sich noch für ein geeignetes ITI entscheiden. Da die Vorperiode konstant ist, sollte das ITI von Durchgang zu Durchgang unterschiedlich lang sein. Wäre dies nicht der Fall, entstünde ein fester Takt, der die Versuchsperson dazu verleiten könnte, getaktet zu reagieren.

In Abbildung 2.11 ist der vollständige zeitliche Verlauf eines Durchgangs im Wahlreaktionsexperiment dargestellt.

**Reaktionszeiten.** Häufig ist man an Reaktionszeiten als abhängigem Maß interessiert. Erhoben wird in diesem Fall die Zeit vom Präsentationsbeginn des Zielreizes bis zum

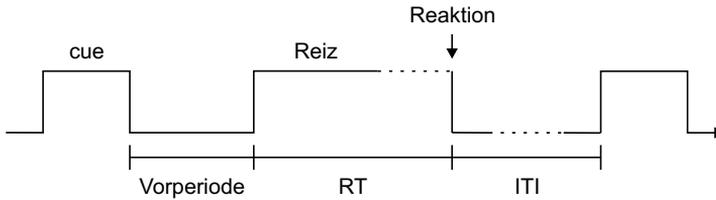


Abbildung 2.11. Vollständiger zeitlicher Verlauf eines Durchgangs im Wahlreaktionsparadigma mit variablem *inter-trial-interval*.

Drücken der entsprechenden Reaktionstaste. Allerdings müssen auch die Fehler betrachtet und ausgewertet werden, um einen eventuellen *speed-accuracy tradeoff* erkennen zu können (vgl. Abschnitt 3.1.3) und um feststellen zu können, ob die Fehlerraten in einem akzeptablen Bereich liegen (vgl. Abschnitt 2.1.2). In die deskriptive und inferenzstatistische Auswertung gehen nur Reaktionszeiten von korrekten Durchgängen ein!

Durch fehlerhafte Durchgänge entstehen fehlende Werte (*missing values*). Dies führt in einem *within-subjects* Design zu Problemen, wenn man pro Versuchsperson und experimenteller Bedingung nur eine Reaktion erhebt. Dann wären die Daten der Versuchsperson unvollständig und sie fiel komplett aus der inferenzstatistischen Standardanalyse heraus. Deshalb ist es wichtig, von jeder Versuchsperson unter allen experimentellen Bedingungen gültige Daten zu erhalten. Daraus ergibt sich ein weiteres Argument für die mehrfache Messung innerhalb einer Person und experimentellen Bedingung, da über die korrekten Replikationen gemittelt werden kann (vgl. Abschnitt 3.1). Hätte man beispielsweise zehn Replikationen pro Bedingung und Versuchsperson, dann gingen in diese Mittelung bei einer Versuchsperson z. B. zehn Werte ein, bei einer anderen aber nur neun, weil sie einen Fehler gemacht hat. Dennoch hat man pro Person und Bedingung letztlich gültige (mittlere) Beobachtungen.

Eine gelegentlich zu lesende Alternativstrategie besteht darin, fehlerhafte Durchgänge am Ende des Experimentes (oder der Sitzung) so oft zu wiederholen, bis die Versuchsperson korrekt geantwortet hat. Auf diese Weise entstehen ebenfalls pro Person und Bedingung gültige Beobachtungen. Dies scheint insbesondere attraktiv für die Fälle, in denen jede Bedingung nur ein einziges Mal realisiert wird. Dieses Verfahren hat allerdings einen beträchtlichen Nachteil: Es entstehen unterschiedlich große Übungseffekte für die unterschiedlichen experimentellen Bedingungen. Angenommen, Bedingung A wäre ›einfach‹, führe zu kurzen Reaktionszeiten und keinen Fehlern und Bedingung B wäre ›schwierig‹, führe zu langen Reaktionszeiten und vielen Fehlern. Die fehlerhaften *trials* aus Bedingung B werden nun so oft wiederholt, bis sie endlich fehlerfrei beantwortet werden. Dadurch entstehen sehr viel mehr Wiederholungen der Bedingung B im Vergleich zu A. Deshalb sind die Übungseffekte in Bedingung B größer. Das führt wiederum dazu, dass die Versuchspersonen schneller werden. Ergo wird der

eigentlich vorhandene Unterschied zwischen A und B künstlich reduziert. Die Wiederholung fehlerhafter Durchgänge sollte also unbedingt vermieden werden.

**Qualität der Antwort.** Gelegentlich betrachtet man als abhängige Variable auch die Qualität der Antwort. Im einfachsten Fall hat man in einem Wahlreaktionsexperiment zwei Reizkategorien/experimentelle Bedingungen und zwei mögliche Reaktionen. Die Versuchsperson kann nun die richtige oder falsche Taste drücken. In diesem Fall sollte die Auswertung mittels Signalentdeckungstheorie (SDT, siehe Abschnitt 3.1.2) erfolgen. Hat man mehr als zwei Reizkategorien bzw. experimentelle Bedingungen und somit auch mehr als zwei mögliche Reaktionen, kann auf die übliche Fehlerauswertung (relative Fehlerhäufigkeit) zurückgegriffen werden (vgl. Abschnitt 3.1.2).

**Tastenzuordnung.** So banal es klingen mag, im Wahlreaktionsparadigma muss man sich auch Gedanken über die Bedienung der Tasten und Zuordnung von Tasten zu korrekten Antworten machen. Hat man, wie in den meisten Fällen, zwei Tasten, so ist es üblich, diese mit den Zeigefingern beider Hände bedienen zu lassen, da in diesem Fall homologe Muskelgruppen für die Reaktionsalternativen innerviert werden. Bei mehr als zwei Tasten sollte man nach dem gleichen Prinzip eine geeignete Effektorenwahl treffen.

Bei der Zuordnung von Tasten zu korrekten Antworten unterscheidet man willkürliche von natürlichen Zuordnungen. Bei letzteren liegt eine Art Kompatibilität zwischen dem Reiz und der Antwort vor, diese kann verschiedener Natur sein (Dehaene et al., 1993; Simon & Rudell, 1967; Wallace, 1971). Präsentiert man beispielsweise einen nach rechts oder links zeigenden Pfeil und die Versuchsperson soll eine rechte oder linke Taste drücken, dann wäre die Zuordnung »rechte Taste« zu »rechter Pfeil« kompatibel.

Häufiger gibt es keine natürliche Zuordnung, d. h. die Tastenbelegung kann vom Versuchsleiter willkürlich festgelegt werden. Beispielsweise werden der Versuchsperson einzelne Zahlen visuell präsentiert und sie hat die Aufgabe zu entscheiden, ob die dargebotene Zahl gerade oder ungerade ist. Für diese Antwortalternativen gibt es zwei Tasten. Entscheidet man sich für die Zuordnungen »gerade – rechte Taste« und »ungerade – linke Taste«, kann es sein, dass die »gerade«-Urteile schneller als die »ungerade«-Urteile erfolgen. Dies hat sehr wahrscheinlich nichts mit der gerade/ungerade-Entscheidung zu tun, sondern liegt vielmehr an der Tatsache, dass die Mehrheit der Versuchspersonen Rechtshänder sind und bei diesen Reaktionen mit der rechten Hand etwas schneller erfolgen.

Man muss also die Tastenzuordnung ausbalancieren. Dies kann man *innerhalb* der Versuchspersonen durchführen, etwa von Block zu Block oder von Sitzung zu Sitzung. Alternativ kann man die Tastenzuordnung *zwischen* den Versuchspersonen ausbalancieren, die Hälfte der Versuchspersonen hat die eine Zuordnung, die andere Hälfte die andere. Hat man die Wahl zwischen beiden Varianten, sollte man sich für das Ausbalancieren *zwischen* Versuchspersonen entscheiden. Wechselt man nämlich innerhalb der Versuchsperson die Tastenzuordnung, so muss diese umlernen, was zu verlangsamten

Reaktionen und erhöhten Fehlerzahlen führt. Zudem wird es von den Versuchspersonen als unangenehm empfunden.

Natürlich gibt es Ausnahmen von dieser Empfehlung. Manchmal benötigt man aus inhaltlichen Gründen sowohl »rechte« als auch »linke« Antworten pro Person und Bedingung (z. B. beim SNARC-Effekt, vgl. Dehaene et al., 1993), dann bleibt nichts weiter übrig, als die Tastenzuordnung block- oder sitzungsweise innerhalb der Person auszubalancieren.

### 2.7.3 Go/no-go

Im Wahlreaktionsexperiment stehen mindestens zwei Tasten für die unterschiedlichen Reaktionen zur Verfügung. Gelegentlich liest man, dass in solch einer Situation genauso gut oder sogar besser eine *go/no-go*-Aufgabe verwendet werden kann. Dabei besteht eine Antwortalternative in der *Nicht-Reaktion*, d. h. die Versuchsperson darf bei bestimmten Reizen bzw. experimentellen Bedingungen keine Taste drücken.

Thorpe et al. (2001) verwendeten solch ein Vorgehen beispielsweise bei einer Bilderklassifikationsaufgabe, bei welcher die Versuchspersonen entscheiden mussten, ob Tiere auf Fotos zu sehen waren oder aber nicht. Sie begründeten das Vorgehen u. a. damit, dass die Reaktionszeiten bei einer *go/no-go*-Aufgabe kürzer seien, weil die Versuchsperson nicht zwischen zwei motorischen Reaktionen zu wählen habe. Ungeachtet der Tatsache, dass diese Begründung nicht sehr plausibel ist – die Versuchsperson muss zwar nicht zwischen zwei unterschiedlichen Reaktionen wählen, dafür aber zwischen »reagieren« und »nicht-reagieren« – bedürfte sie der empirischen Bestätigung. Aber selbst wenn sich dies empirisch bestätigen ließe, gibt es weitere Argumente gegen die Verwendung einer *go/no-go*-Aufgabe: Es lassen sich nur Reaktionszeiten für die *go-trials* erheben, man weiß für die *no-go-trials* nicht, warum die Versuchsperson nicht reagiert hat.

Bei Thorpe et al. (2001) erhielten die Autoren zwar für alle Reize die Qualität der Antwort (richtig/falsch), aber nur für die Hälfte der Reize zugehörige Reaktionszeiten. Auch wurden die Reize nur kurz und mit z. T. großen Exzentrizitäten dargeboten: Die Versuchspersonen waren sich oft unsicher, ob ein Tier auf dem Bild zu sehen war oder nicht. Hat die Versuchsperson in einem *no-go-trial* nun nicht gedrückt, weil sie (sicher) kein Tier gesehen hat oder weil sie sich einfach unsicher war? Außerdem haben Thorpe et al. ein Antwortfenster von 1 s vorgeben. Dies ist für *no-go-trials* unvermeidlich, führt jedoch dazu, dass im Prinzip gültige, aber lange Reaktionszeiten in *go-trials* verunmöglicht werden.

All diese Probleme lassen nur einen Schluss zu: Das *go/no-go*-Verfahren als Alternative zum Wahlreaktionsexperiment hat nur Nachteile und sollte nicht eingesetzt werden.

Gelegentlich wird das *go/no-go*-Paradigma auch als Alternative zur variablen Vorperiode im Einfachreaktionsparadigma verwendet. In diesem Fall werden sogenannte