

## Teil III

# Entwicklung Interaktiver Systeme



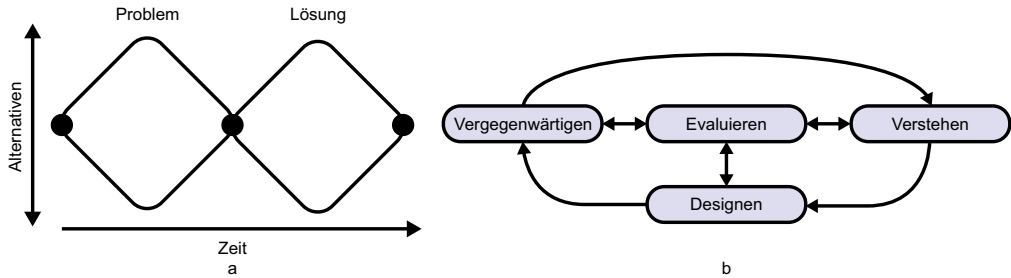
# 10 Grundidee des User Centered Design

In diesem Abschnitt wollen wir uns der Frage widmen, wie ein sinnvoller Prozess zum Entwurf von Benutzerschnittstellen aussehen könnte. Welche Eigenschaften sollte ein solcher Entwurfsprozess besitzen und aus welchen Schritten besteht er? Lange Zeit war der Entwurf von Benutzerschnittstellen geprägt durch die Informatiker und Ingenieure, die die dahinter liegenden Systeme entworfen und implementiert haben. Dass dies nicht immer zu besonders benutzerfreundlichen Systemen führt, haben selbst Weltkonzerne erst in jüngster Zeit erkannt. SAP, einer der Weltmarktführer in betrieblicher Anwendungssoftware, hat sein erstes Usability Labor erst mehr als zwanzig Jahre nach seiner Gründung Mitte der 90er Jahre installiert. Inzwischen bezeichnet der Konzern **Usability** als einen der wichtigsten Faktoren für den Erfolg der SAP Software. Früher wurde die Usability einer Software ausschließlich im Labor getestet, und zwar am fertigen Softwareprodukt mit voll implementierten Benutzerschnittstellen. Diese mussten dann mühselig wieder korrigiert werden, und so wurde die Qualität der Benutzerschnittstelle oft nachrangig behandelt. Ein moderner Entwicklungsansatz integriert die Benutzbarkeit als gleichwertiges, wenn nicht sogar als entscheidendes Element in den Design- und Entwicklungsprozess. Um das besser zu verstehen muss man sich zunächst fragen, was Design überhaupt ist. Dix et al. [26] definieren Design als:

Ziele unter Berücksichtigung von Einschränkungen erreichen.

Die **Ziele** beschreiben, was mit dem Design erreicht werden soll, in welchem Kontext das Design verwendet werden soll, und für welche Benutzer die Software und damit auch die Benutzerschnittstelle entwickelt wird. Es macht schließlich einen großen Unterschied, ob eine Software sich an Jugendliche richtet, z.B. um Bilder im sozialen Netzwerk zu tauschen, oder ob eine neue Benutzerschnittstelle für ein Seniorentelefon entwickelt werden soll.

Design ist immer auch Beschränkungen unterworfen. Daher ist es sehr wichtig diese Beschränkungen zu kennen. Dabei kann es sich zunächst um materielle Beschränkungen handeln. So hat z.B. bei elektronischen tragbaren Geräten die Größe der Batterie häufig einen entscheidenden Einfluss auf die endgültige Form des Geräts. Die Entwicklung neuer Materialien oder technologischer Fortschritt führt oft zu neuen Designalternativen, wie im Fahrzeugbau und in der Architektur regelmäßig zu beobachten ist. Häufig müssen Standards beim Designprozess berücksichtigt werden und beschränken die möglichen Alternativen für das Design einer Benutzerschnittstelle. Die große Anzahl von Beschränkungen führt dazu, dass ein *optimales* Designergebnis meist nicht zu erreichen ist. Der Designer muss daher immer Kompromisse eingehen. Die Kunst dabei ist



**Abbildung 10.1:** Links der Double-Diamond-Prozess, Rechts die Hauptelemente des User-Centered-Design: Evaluieren, Vergegenwärtigen, Verstehen und Designen (nach Norman [86] und Benyon [10])

zu entscheiden wie die Beschränkungen priorisiert werden können, welche abgeschwächt und welche eventuell fallen gelassen werden sollten, um ein akzeptables Designergebnis erzielen zu können. Entscheidend im Entwurfsprozess ist dabei das Wissen, um die Bedürfnisse und **Ziele** der Benutzer. Diese können sich sehr von den ursprünglichen Annahmen unterscheiden, die der Designer einer Benutzerschnittstelle zu Beginn der Arbeit vornimmt. Oberstes Ziel des Designprozesses ist daher die frühe und umfassende Integration der zukünftigen Benutzer.

Der Designprozess teilt sich in zwei Phasen: in das *Verständnis der Problemdomäne* und das *Erarbeiten von Lösungen*. Dazu werden jeweils zu Beginn der beiden Phasen Alternativen exploriert, um diese dann anschließend zu konsolidieren. Wird, wie in Abbildung 10.1a zu sehen, die Anzahl der betrachteten Alternativen auf der einen und der zeitliche Ablauf des Designprozesses auf der anderen Bildachse abgebildet, so ergibt sich die charakteristische Gestalt des **Double-Diamond**, nach dem der Prozess benannt wurde [86]. Die Exploration und die Konsolidierung der beiden Hälften des Double-Diamond folgen dabei dem Prozess in 10.1b, welcher als **User Centered Design (UCD)** bezeichnet wird.

Die Hauptelemente des UCD sind *Verstehen*, *Vergegenwärtigen*, *Evaluieren* und *Designen*. Die Pfeile in der Abbildung zeigen den Ablauf des UCD und die zentrale Stellung der Evaluation. Der UCD Prozess kann mit jedem seiner Elemente beginnen und sollte im besten Fall immer durch mindestens einen Evaluationsschritt überprüft werden. Die Ergebnisse der Evaluation führen dann schrittweise zu einer Verbesserung der Resultate. So kann z.B. ein erster Designentwurf als Startpunkt dienen, welcher verbessert werden soll. Häufig stehen allerdings das Verständnis von (neuen) Benutzerbedürfnissen oder die Möglichkeiten einer neuen Technologie am Anfang des UCD Prozesses, welche dann evaluiert und in einem ersten Design umgesetzt werden. Im Folgenden wollen wir die vier Aspekte des UCD näher beleuchten.

## 10.1 Verstehen

Das Verständnis der Beschränkungen des Designprozesses ist von großer Bedeutung für die Entwicklung. Dies gilt nicht nur im besonderen Fall des Entwurfs und der Implementierung von Benutzerschnittstellen, sondern sehr allgemein für das Design von Produkten überhaupt. Bei der Mensch-Maschine-Interaktion steht in diesem Zusammenhang das Verständnis der technischen und menschlichen Beschränkungen im Vordergrund, üblicherweise in Bezug auf eine Software (z.B. eine Textverarbeitung) oder ein interaktives Produkt (z.B. eine Spielekonsole). Im UCD stehen Designer vor der Aufgabe, möglichst genau zu erkennen, was die Bedürfnisse der zukünftigen Benutzer sind, sowie in welchem Kontext das Endprodukt verwendet wird. Die Abwägung, welche Beschränkungen berücksichtigt, und welche ignoriert werden können, ist ein wichtiger Aspekt des Verständnisprozesses und hat großen Anteil am Erfolg der endgültigen Benutzerschnittstelle [26]. In Abschnitt 18.2 wird dies am Beispiel mobiler Interaktion ausführlich diskutiert. Technische Nebenbedingungen müssen genauso berücksichtigt werden wie die zu identifizierenden Personengruppen, die direkt oder indirekt mit dem Endprodukt interagieren werden oder von seiner Einführung betroffen sind. In Abschnitt 11.1 wird näher auf unterschiedliche Kategorien von Benutzergruppen eingegangen und wie diese in den Designprozess mit einbezogen werden sollten. Ebenso sollten die Aktivitäten analysiert werden, die von der Einführung des neuen Produkts oder der Benutzerschnittstelle betroffen sind. Häufig müssen Arbeitsprozesse angepasst werden, wenn eine neue Software eingeführt wird. Das Verständnis der Konsequenzen dieser Anpassung ist wichtig, um Fehler zu Beginn des Designprozesses zu vermeiden. Schließlich müssen die technischen Nebenbedingungen analysiert und bewertet werden. Daraus lässt sich dann ableiten, welche Realisierungen technisch unmöglich sind, bzw. finanziell aus dem Rahmen fallen. Findet diese Analyse früh genug im Designprozess statt, werden im späteren Verlauf Kosten und Zeit gespart. Methoden zur Erkennung von Benutzerbedürfnissen und -kontexten werden in Kapitel 11 besprochen.

## 10.2 Designen

Im Designschritt des UCD wird ein konkreter Vorschlag entwickelt, in den die Erfahrungen des bisherigen Prozesses einfließen. Er besteht aus einem abstrakten Schritt, dem **konzeptuellen Design** und einem konkreteren, dem **physischen Design**. Das konzeptuelle Design mündet in ein **konzeptuelles Modell** (siehe Abschnitt 5.1), welches auf abstrakte Art und Weise beschreibt, welche Aktivitäten und Objekte im Zusammenhang mit dem Entwurf der Benutzerschnittstelle stehen (siehe dazu auch Abschnitt 8.4). Es kann aus einer einfachen Auflistung von Aktivitäten der Benutzer mit realen oder digitalen Objekten bestehen oder formaler durch ein **Entity-Relationship Modell** beschrieben werden, welches im Software-Design häufig Verwendung findet. Der Informationsfluss und der Interaktionsablauf werden im konzeptuellen Modell häufig durch Flussdiagramme repräsentiert. Das konzeptuelle Modell des Designers steht im direkten Zusammenhang mit dem **mentalenden Modell** des Benutzers (siehe Abschnitt 5.1). Entscheidend ist dabei, dass das konzeptuelle Modell abstrakt genug bleibt und das konkrete physikalische Design noch offen lässt. Es geht im konzeptuellen Design insofern um die generelle Frage, *was* entwickelt werden soll, *wer* davon beeinflusst wird

und in *welchem* Kontext das Design steht. Es geht nicht um die Frage, *wie* die Umsetzung des Designs erfolgen sollte. Diese Variante des Designs spielt insbesondere beim Verständnis der Problemdomäne (linke Seite des Double-Diamond) eine wichtige Rolle.

Die Frage der Umsetzung wird im **physischen Design** adressiert. Dieses beschäftigt sich mit der konkreten Ausprägung der Benutzerschnittstelle, beispielsweise der konkreten grafischen Umsetzung und der genauen Benutzerführung. Es spezifiziert, wie unterschiedliche Geräte unterstützt werden sollen (z.B. im Fall des Designs einer Webseite, welche Layouts und Interaktionsmöglichkeiten für stationäre und mobile Geräte vorgesehen werden). Das physische Design greift also das abstrakte konzeptuelle Design auf und transformiert es in einen konkreten Designvorschlag. Viele weitere Aspekte müssen im physischen Design berücksichtigt werden, z.B. welchen Eindruck das Produkt beim Benutzer hinterlässt. Die emotionale Wirkung eines Produktes und die Art und Weise, wie Benutzerbedürfnisse von dem Produkt erfüllt werden, hat eine nicht unerhebliche Bedeutung für den Erfolg eines Produktes. In Kapitel 14 wird dieser Aspekt ausführlicher behandelt. Der resultierende physische Designvorschlag ist häufig nicht endgültig, dient aber im Fortlauf des UCD der Vergegenwärtigung des bisherigen Designprozesses und wird beim Erarbeiten von Lösungen (rechte Seite des Double-Diamond) benötigt.

## 10.3 Vergegenwärtigen

Ein weitere wichtiger Baustein im Designprozess ist das Vergegenwärtigen von Zwischenlösungen oder Designalternativen. Visualisierungsmethoden bilden häufig die notwendige Diskussionsgrundlage, um im UCD weiter fortschreiten zu können. Methodisch sollten dabei keine Grenzen gesetzt werden. Häufig reicht eine einfache Skizze mit kurzer Beschreibung, um die Kernideen eines Designentwurfes zu kommunizieren. Manchmal müssen auch aufwändige 3D-Modelle konstruiert werden, um die Wirkung des Designs besser abschätzen zu können, da eine reine Visualisierung wenig über die haptischen Qualitäten (wie z.B. die Textur oder das Gewicht) aussagen. Inzwischen steht eine Vielzahl von Werkzeugen bereit, die es Designern erlauben **Prototypen** zu entwerfen, die dann als Diskussionsgrundlage im weiteren Verlauf des UCD eingesetzt werden können. In Kapitel 12 werden Prototypen detailliert besprochen und unterschiedlichste Werkzeuge vorgestellt, mit denen Prototypen erzeugt werden können.

## 10.4 Evaluieren

Die **Evaluation** steht in der Abbildung 10.1 im Mittelpunkt und ist von zentraler Bedeutung im UCD. Jeder der bisher beschriebenen Schritte des UCD sollte durch einen Evaluationsschritt abgesichert werden. Zu diesem Zweck steht eine große Bandbreite von Evaluationsmethoden zur Verfügung, die in den Kapitel 11 und 13 im Detail vorgestellt werden. Häufig reicht es allerdings aus, wenn der Designer nach einem Bearbeitungsschritt eine der Methoden für sich selbst anwendet, um die Plausibilität des Schrittes zu überprüfen, oder eine kurze Diskussion mit dem Auftraggeber oder einem Kollegen anzustoßen. Einige Schritte erfordern allerdings eine umfangreiche Bestandsaufnahme und Evaluation, die z.B. im Rahmen von groß angelegten Benutzerstudien umgesetzt

werden. Entscheidend bei der Wahl der richtigen Methode ist die konkrete Fragestellung und der vorangehende Schritt im UCD. So werden z.B. Prototypen häufig in Benutzerstudien überprüft, während ein konzeptuelles Design in vielen Fällen zunächst nur mit dem Auftraggeber diskutiert wird.

## 10.5 Iteratives Design

Die Fülle an Beschränkungen, die es beim UCD zu berücksichtigen gilt, führt in der Regel dazu, dass der erste Entwurf einer Benutzerschnittstelle verbessert werden muss und erst nach mehreren Iterationen des Designs ein zufriedenstellendes Endprodukt vorliegt. Zwischenergebnisse sollten, wie im vorherigen Abschnitt beschrieben, evaluiert werden. Es stellt sich aber die grundlegende Frage, wann dies geschehen sollte. In der traditionellen Softwareentwicklung fand die Evaluation der Benutzerschnittstelle am Ende des Entwicklungsprozesses statt. Kosmetische Änderungen am Design der Benutzerschnittstelle waren so zwar noch möglich, grundlegende Änderungen der Struktur oder Funktionalität waren aber nur unter großen Anstrengungen und Kosten zu realisieren. Im UCD ist der Entwurf der Funktionalität einer Software oder eines interaktiven Produktes hingegen eng verzahnt mit dem Design und der Evaluation der Benutzerschnittstelle. Die Benutzerschnittstelle ist damit ein zentrales Element des Produktes, welches nicht einfach von der Funktionalität zu trennen ist. Die Iteration des Designs der Benutzerschnittstelle findet in möglichst kleinen Schritten statt, denen jeweils ein Evaluationsschritt folgt. Dies kann parallel für verschiedene Designvorschläge der Benutzerschnittstelle durchgeführt werden, so dass mehrere Alternativen gegeneinander abgewogen werden können.

## 10.6 Implementierung

Am Ende jedes Designprozesses steht die Implementierung bzw. Umsetzung des letzten Designvorschlags. Auch wenn bei der bisherigen Entwicklung der Prototypen schon Programmcode implementiert wurde, muss in der Implementierungsphase üblicherweise mit weiterem (nicht unerheblichem) Programmieraufwand gerechnet werden. Eventuell müssen Software-Prototypen komplett neu programmiert werden, da der Code der Prototypen nicht zur Wiederverwendung geeignet ist. Neben der Software wird häufig ein nicht unerheblicher Hardwareanteil in der Implementierungsphase realisiert. Dokumentationen müssen ebenfalls geschrieben und geprüft werden. Schließlich sollte eine Abstimmung aller Komponenten erfolgen und diese mit dem Vertrieb besprochen werden, bevor das Produkt in den Handel und zum Benutzer kommen kann. Der Aufwand der Implementierungsphase ist nicht zu unterschätzen und kann bei komplexen Produkten ein mehrfaches der Kosten des UCD Prozesses verschlingen.



### Übungsaufgaben:

1. Stellen Sie sich vor, Sie hätten die Gelegenheit Ihren Arbeitsplatz neu zu gestalten. Überlegen Sie sich hierzu die verschiedenen Aspekte und Prozesse, die Sie aus Sicht des UCD berücksichtigen müssten, um ein möglichst gutes Ergebnis zu erzielen. Überlegen Sie sich insbesondere, wie das konzeptuelle Design, das physische Design, die Beschränkungen und Anforderungen an den Arbeitsplatz aussehen müssten. Formulieren Sie im Sinne des UCD (Abbildung 10.1) Ihre Vorgehensweise, und wie Ihre Zwischenergebnisse evaluiert werden könnten.
2. Machen Sie ein Foto von einem Verkaufsautomaten Ihrer Wahl. Diskutieren Sie das konzeptuelle Modell des Automaten, d.h. beschreiben Sie die Objekte und Aktivitäten, die beim Kaufvorgang eine Rolle spielen mit Hilfe des Fotos. Diskutieren Sie anhand dieses Modells die Vor- und Nachteile des Automaten aus Sicht des UCD.



# 11 Benutzeranforderungen erheben und verstehen

Im vorhergehenden Kapitel haben wir uns eingehend mit dem Prozess des User-Centered-Design (UCD) beschäftigt. Eines der wichtigen Elemente des Prozesses ist das Verständnis der **Benutzeranforderungen**, d.h. welche Bedürfnisse Benutzer bezüglich einer Benutzerschnittstelle haben und in welchem **Benutzungskontext** die Benutzer mit der Benutzerschnittstelle interagieren. Für den Entwickler einer Benutzerschnittstelle stellt sich dabei die Frage, wie diese Benutzeranforderungen erhoben und interpretiert werden können ohne die Benutzerschnittstelle vollständig implementieren zu müssen. Diese Frage stellt sich insbesondere zu Beginn des UCD – in einer Phase, in der noch keine detaillierten Prototypen zum Testen zur Verfügung stehen. Dieses Kapitel geht damit der Frage nach, wie man Wissen zu den Benutzeranforderungen initial erheben und wie dieses Wissen dann Entwickeln von Benutzerschnittstellen zur Verfügung gestellt werden kann.

## 11.1 Stakeholder

Bevor man versucht, die Benutzeranforderungen zu erheben und zu verstehen, muss zunächst einmal geklärt werden, wer die Benutzer eigentlich sind, die später das Produkt und die Benutzerschnittstellen verwenden werden. Diese Frage zu beantworten ist gar nicht so einfach wie man vielleicht denken mag. Nehmen wir als Beispiel eine *intelligente* Zahnbürste, die Kinder dabei unterstützt, regelmäßig und richtig ihre Zähne zu putzen<sup>1</sup>. Offensichtlich richtet sich das Produkt an Kinder, die damit eindeutig als Benutzergruppe zu identifizieren wären. Allerdings sind diese nicht die alleinigen Benutzer des Produkts. Die Eltern der Kinder sind in der Regel diejenigen, die das Produkt kaufen und eventuell auch die Auswertungen der Zahnbürste erhalten. Sie sind also ebenfalls direkt betroffen. Schließlich kann das Produkt auch einen Einfluss auf die Tätigkeit von Zahnärzten haben, so dass auch diese in den Entwurfsprozess mit eingebunden werden sollten. Allgemein bezeichnet man im UCD die Personen, die durch die Einführung und den Einsatz eines Produktes betroffen sind, als **Stakeholder** (dt. Mitglied einer Interessensgruppe). Man erkennt an diesem Beispiel, dass die Gruppe der Stakeholder sehr weitreichend sein kann und manchmal schwer abzugrenzen ist. Im UCD betrachtet man grundsätzliche zunächst alle Personen, die durch ein Produkt direkt oder indirekt betroffen werden als Stakeholder [106].

---

<sup>1</sup>Eine entsprechendes Produkt wird z.B. von der Firma Kolibree angeboten, siehe <http://www.kolibree.com/>. Die Zahnbürste sendet über eine Bluetooth-Schnittstelle Daten zum Putzverhalten an ein Smartphone. Diese werden dort ausgewertet, um dem Benutzer Tipps zur besseren Zahnpflege geben zu können.

Diese Personengruppen werden dann, in Abhängigkeit von ihrer Beziehung zu dem neuen Produkt unterschiedlich priorisiert. Als nützlich hat sich in der Praxis die Einteilung in drei Gruppen erwiesen: die **primären, sekundären, tertiären** Stakeholder [29]. *Primäre* Stakeholder sind die Benutzer, die mit großer Wahrscheinlichkeit direkt mit dem Produkt in Kontakt kommen und dieses täglich benutzen. Im Beispiel der Zahnbürste sind dies die Kinder, die sich damit die Zähne putzen. *Sekundäre* Stakeholder sind Benutzer, die durch die Einführung des Produkts indirekt betroffen sind, es also nicht regelmäßig direkt nutzen, z.B. die Eltern der Kinder. *Tertiäre* Stakeholder sind im weitesten Sinne von der Einführung des Produkts betroffen: Neben den Zahnärzten und ihren Sprechstundenhilfen, könnten das auch Mitarbeiter von Elektronikläden sein, die Eltern über die neuen Möglichkeiten der Zahnbürste informieren.

Die Gruppe aller Benutzer ist also oft viel größer als ursprünglich angenommen und es kann schnell unübersichtlich werden, welche der Gruppen betrachtet werden sollten und welche nicht. Stakeholder lassen sich in einem **Zwiebelschalen-Modell** organisieren. Im Mittelpunkt des Modells steht das Produkt mit der Benutzerschnittstelle. In der ersten inneren Schale befinden sich die primären Stakeholder, in der zweiten die sekundären und in der dritten die tertiären. Um zu verstehen, welche Benutzer in welcher Form in den Entwicklungsprozess mit einbezogen werden sollen, ist es wichtig, die Stakeholder zu Beginn des UCD zu identifizieren und bei den weiteren Prozessschritten zu involvieren. Üblicherweise werden zunächst die primären, dann die sekundären und schließlich die tertiären Stakeholder berücksichtigt.

Hat man die relevanten Stakeholder und damit die potenziellen direkten und indirekten Benutzer identifiziert, müssen im nächsten Schritt die konkreten Bedürfnisse und Benutzerkontexte ermittelt werden. Im Folgenden werden einige Erhebungsmethoden besprochen, um dieses Ziel zu erreichen.

## 11.2 Interviewtechniken

Die direkte **Befragung** der Benutzer durch unterschiedliche Interviewtechniken ist eine der gängigsten und auch zuverlässigsten Methoden, um mehr über Benutzerbedürfnisse zu erfahren und die Kontexte und Nebenbedingungen zu verstehen, in denen eine Benutzerschnittstelle eingesetzt werden könnte. Es können jedoch nicht alle Benutzerbedürfnisse durch direkte Befragung erhoben werden. Henry Ford, der Begründer der amerikanischen Automarke, wird sinngemäß mit dem Ausspruch in Verbindung gebracht: *Hätte ich meine Kunden nach Ihren Bedürfnissen gefragt, hätten Sie schnellere Pferde gewollt.* Damit wird die häufig anzutreffende Unfähigkeit von Benutzern angesprochen, neue Bedürfnisse und Innovationen überhaupt aktiv zu benennen. Interviewtechniken können aber sehr gut geeignet sein, um erste Eindrücke zu sammeln oder spezifische Fragen zu beantworten. Grundsätzlich wird ein Interview immer von einem Interviewer oder Moderator geleitet. Je nach Fragestellung und Kontext können unterschiedliche Typen von Interviews durchgeführt werden. Man unterscheidet dabei **unstrukturierte, strukturierte, semi-strukturierte** und **Gruppeninterviews** [35]. *Unstrukturierte* Interviews werden häufig in einer sehr frühen Phase des UCD geführt, um ein grundlegendes Verständnis von Benutzerkontexten und -bedürfnissen zu gewinnen (linke Seite des Double-Diamond, Abbildung 10.1). Der Interviewer gibt dabei grob ein Thema vor

und es entspannt sich in der Regel eine Diskussion mit dem Befragten, die zu unvorhergesehenen Themen führen kann. So ergibt sich ein reiches, wertvolles und komplexes Meinungsbild zu einem Themenbereich. Ergebnisse unstrukturierter Interviews mit mehreren Befragten lassen sich allerdings häufig schwer miteinander vergleichen, und die Analyse kann oft nicht standardisiert durchgeführt werden.

Je konkreter im weiteren Verlauf des UCD die Fragestellungen werden, desto strukturierter kann das Interview geführt werden. *Strukturierte* Interviews bestehen aus einer Menge vorher definierter Fragen mit einer vorgegebenen Antwortauswahl, einer sogenannten **geschlossenen Fragestellung**. Typische Antworten auf geschlossen Fragen sind auf die Menge (*ja/nein/weiß nicht*) beschränkt. Im Gegensatz zu unstrukturierten Interviews sind alle Fragen an alle Befragten gleich zu stellen. Hierdurch wird eine Vergleichbarkeit der Antworten gewährleistet. *Semi-strukturierte Interviews* kombinieren Elemente von strukturierten und unstrukturierten Interviews und bilden somit einen Kompromiss zwischen beiden Interviewformen. Bei einem semi-strukturierten Interview geht der Interviewer nach einem Frageskript vor, welches das Interview strukturiert. Die Antworten des Befragten hingegen sind frei. Entscheidend dabei ist, dass der Interviewer die Fragen so stellt, dass die Antwort des Befragten nicht beeinflusst wird. Insbesondere sollten Fragen offen formuliert werden, so dass einfache Antworten wie in strukturierten Interviews nicht möglich sind.

In Gruppeninterviews wird schließlich nicht der Einzelne befragt, sondern die Erkenntnisse mithilfe von Gruppengesprächen und -arbeiten gemeinsam mit einem Moderator gewonnen. Dies wird in Abschnitt 11.4 weitergehend besprochen.

## 11.3 Fragebögen

Die Verwendung von **Fragebögen** ist ebenfalls ein probates Mittel der **Befragung**, um strukturiert Information von Stakeholdern zu erhalten. Fragebögen werden auch häufig in strukturierten Interviews verwendet und die Vorgehensweise ähnelt dem Interview sehr, mit dem Unterschied, dass Fragebögen ohne Interviewer vom Befragten bearbeitet werden. Insofern ist es sehr wichtig, dass die Fragen sorgfältig und klar formuliert sind, da der Befragte keine Rückfragen stellen kann. Dies kann auch ein Kriterium sein, um zu entscheiden, ob eher ein strukturiertes Interview oder ein Fragebogen als Methode gewählt werden sollte: Nimmt man an, dass der Benutzer sehr motiviert ist und wenig Rückfragen hat, kann der Fragebogen gewählt werden. Ist der Benutzer hingegen wenig motiviert und ist die Wahrscheinlichkeit von Rück- und Klärungsfragen groß, ist es sinnvoll, die Information im Rahmen eines strukturierten Interviews zu erheben.

Ein großer Vorteil des Fragebogens ist die einfache Verteilung an große Gruppen von Stakeholdern und die entsprechend große Menge an Antworten, die man erhalten kann. Fragebögen können herkömmlich auf Papier reproduziert werden oder den Befragten digital auf einem PC oder Tablet präsentiert werden. Letzteres hat den großen Vorteil der automatischen Auswertung. Eine Besonderheit des elektronischen Fragebogens ist der Online Fragebogen, der näher in Abschnitt 11.3.3 besprochen wird.

### 11.3.1 Struktur

Ein Fragebogen beginnt in der Regel mit allgemeinen Fragen zum Hintergrund des Befragten, wie z.B. dem Alter, dem Bildungsniveau und der technische Expertise. Diese Information ist nützlich, um weitere Fragen zu relativieren und insbesondere die Antworten von unterschiedlichen Befragten besser vergleichen zu können. Fragen zu einer bestimmten Funktion einer Smartphone App werden beispielsweise häufig in Abhängigkeit vom Alter oder der technischen Expertise beantwortet, da ältere Leute tendenziell weniger Erfahrung mit Smartphones haben. Nach einem Block von allgemeinen Fragen folgen die spezifischeren Fragen. Dabei sollte man sich von folgenden Prinzipien leiten lassen (nach Rogers et al. [96, S. 238ff]):

- *Leichte vor schweren Fragen.*
- *Beeinflussungseffekte von Fragen vermeiden.* Vorhergehende Fragen können nachfolgende Fragen beeinflussen (**Reihenfolgeeffekt**). Fragen, die leicht zu beeinflussen sind, sollten an den Anfang gestellt werden, Fragen, die zu einer Beeinflussung führen, eher ans Ende. Bei elektronischen Fragebögen sollte die Fragereihenfolge im Zweifelsfall zufällig generiert werden.
- *Verschiedene Versionen des Fragebogens erstellen.* Erwartet man große Unterschiede bei den Antworten der Befragten und lassen sich diese Unterschiede einfach erklären (z.B. durch das Alter der Befragten), so sollten unterschiedliche Fragebögen zum Einsatz kommen, die es erlauben, den Unterschieden weiter auf den Grund zu gehen.
- *Klare Fragen stellen.* Die Fragen sollten unmissverständlich gestellt werden und klar formuliert sein.
- *Fragebogen kompakt gestalten.* Zu viele Fragen und zu viel Leerraum zwischen den Fragen führt zu Ermüdungserscheinungen. Da in der Regel am Ende die anspruchsvolleren Fragen gestellt werden (siehe oben), sollte eine Ermüdung des Befragten möglichst vermieden werden.
- *Nur notwendige Fragen stellen.* Fragen, die mit dem Thema des Fragebogens nichts zu tun haben, sollten unterlassen werden. So wenig Fragen wie möglich stellen.
- *Offene Fragen ans Ende.* Sollten offene Fragen gestellt werden, sollten diese am Ende des Fragebogens platziert werden. Selbst wenn offene Fragen nicht ausgewertet werden können, ist eine allgemeine offene Frage am Ende des Fragebogens sinnvoll, um dem Befragten die Möglichkeit der Rückmeldung zu geben.

Es ist eine gute Idee, einen Fragebogen vorab mit einer kleinen Gruppen von Befragten auf Ermüdungserscheinungen und Reihenfolgeeffekte zu testen (**Pilotstudie**). Ein solcher Test ermöglicht auch eine Abschätzung des zeitlichen Aufwands, mit dem der Befragte rechnen muss. Dieser sollte vorab kommuniziert werden.

**Es gibt viel zu wenige gute Science-Fiction Romane**

**Numerische Skala**  
 (1 repräsentiert starke Zustimmung, 5 starke Ablehnung)

1    2    3    4    5

☐   ☐   ☐   ☐   ☐

**Verbale Skala**

Starke                      Starke  
 Zustimmung   Zustimmung   Neutral   Ablehnung   Ablehnung

☐       ☐       ☐       ☐       ☐

Abbildung 11.1: Beispiel einer Likert-Skala. Oben wird eine numerische Skala verwendet. Unten, zur Beantwortung der gleichen Frage, eine verbale Skala.



mmibuch.de/s/11.1

## 11.3.2 Antwortformen

Antworten folgen in Fragebögen unterschiedlichen Formaten. Neben der offenen Frage, die eine beliebige Antwort erlaubt, finden sich häufig geschlossene Fragen, die nur einige wenige Antwortmöglichkeiten zulassen. Bei quantitativen Angaben ist es eventuell sinnvoll, Intervalle vorzusehen, so dass die Befragten beispielsweise nicht ihr genaues Alter angeben müssen. **Checkboxen** werden durch den Befragten angekreuzt (bei Papierfragebögen) oder angeklickt (bei Online Fragebögen). Je nach Fragetypus können mehrere Checkboxen zu Antwort vorgesehen werden. Dabei gibt es Fragen, die normalerweise *genau eine* mögliche Antwort haben, z.B. die Frage nach dem Geschlecht oder der Anrede. Andere Fragen erlauben es, mehrere Checkboxen auszuwählen, z.B. die Frage, an welchen Wochentagen man seinem Hobby nachgeht. Checkboxen sind sehr gut geeignet, um schnell und präzise Antworten auf genau spezifizierte Fragen zu erhalten.

Häufig sollen Befragte ihre Einschätzung zu bestimmten Sachverhalten im Fragebogen ausdrücken. In solchen Fällen kommen **Bewertungsskalen** zum Einsatz, die es erlauben unterschiedlichen Antworten miteinander zu vergleichen. Die **Likert-Skala** erfragt die persönliche Einschätzung zu einem bestimmten Sachverhalt. Dieser Sachverhalt wird in einer positiven oder negativen Aussage festgehalten, wie z.B. der Aussage *Es gibt viel zu wenige gute Science-Fiction Romane*. Durch die Antwort wird der Grad der Zustimmung zu dieser Aussage gemessen, entweder als Zahl auf einer Skala oder durch Ankreuzen einer von mehreren vorgegebenen Antworten (siehe Abbildung 11.1).

Üblicherweise finden sich in einem Fragebogen mehrere Fragen, die mit Likert-Skalen beantworten werden sollen. In diesem Fall ist es ratsam, alle Aussagen entweder positiv oder negativ zu formulieren, um so Fehler bei der Beantwortung zu reduzieren. Eine

**QR-Codes in einem Buch finde ich....**

Attraktiv	<table border="1" style="border-collapse: collapse; width: 100%;"> <tr><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td></tr> </table>								Hässlich
Klar	<table border="1" style="border-collapse: collapse; width: 100%;"> <tr><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td></tr> </table>								Konfus
Fade	<table border="1" style="border-collapse: collapse; width: 100%;"> <tr><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td></tr> </table>								Lebhaft
Begeistert	<table border="1" style="border-collapse: collapse; width: 100%;"> <tr><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td></tr> </table>								Langweilig
Hilfreich	<table border="1" style="border-collapse: collapse; width: 100%;"> <tr><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td></tr> </table>								Störend
Überflüssig	<table border="1" style="border-collapse: collapse; width: 100%;"> <tr><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td><td style="width: 12.5%; height: 20px;"></td></tr> </table>								Sinnvoll



mmibuch.de/a/11.2

**Abbildung 11.2:** Beispiel eines semantischen Differenzials. Eine Aussage wird mithilfe von bipolaren Adjektivpaaren auf einer Skala bewertet.

wichtige Entscheidung, die ebenfalls getroffen werden sollte, ist die Anzahl der Antwortmöglichkeiten (in Abbildung 11.1 sind es z.B. jeweils fünf). Diese Anzahl hängt von der Frage ab, aber auch von der gewünschten Granularität der Antwort. So kann es in einigen Fällen sinnvoll sein, nur drei Antwortmöglichkeiten vorzugeben (*Zustimmung, Neutral, Ablehnung*), um ein schnelles Meinungsbild einzuholen. Größere Anzahlen können nachträglich noch gruppiert werden in positiv/neutral/negativ. In Abbildung 13.3 auf Seite 144 wird dies durch die gestrichelte bzw. durchgezogene Umrandung der Untergruppen angedeutet.

In der Praxis werden häufig zwischen fünf und acht Antwortmöglichkeiten bei Likert-Skalen verwendet. Eine ungerade Anzahl hat den Vorteil eines klaren neutralen Mittelpunkts, aber den Nachteil, dass bei Unsicherheit dieser Mittelpunkt häufig ausgewählt wird. Eine gerade Anzahl hingegen veranlasst die Befragten, sich auf eine der beiden Seiten zu schlagen und entsprechend Farbe zu bekennen.

Ein alternatives Instrument zur Befragung ist das aus der Psychologie bekannte **Semantische Differenzial**. Hier wird die Einschätzung des Benutzers nicht über die Bewertung einer Aussage erfasst, sondern mithilfe paarweise angeordneter bipolarer Adjektive. Das Adjektiv-Paar repräsentiert jeweils die beiden Enden eines Spektrums (z.B. *attraktiv* versus *hässlich*) und der Befragte muss seine Bewertung durch Ankreuzen zwischen den jeweiligen Polen ausdrücken (siehe Abbildung 11.2). Die Reihenfolge der Pole sollte ausbalanciert werden, so dass sich negative und positive Adjektive auf jeder Seite abwechseln. Der Gesamtwert für jedes bipolare Paar errechnet sich bei der Auswertung als Durchschnitt der Antwortwerte, wobei der höchste Antwortwert am positiven Pol liegt. In der Praxis haben sich sieben Zwischenwerte für das semantische Differenzial durchgesetzt.

### 11.3.3 Online Fragebögen

Fragebögen, die über ein **Online-Medium** verbreitet werden, z.B. als Webseite oder in einer E-Mail, erreichen ein sehr großes Publikum und haben damit das Potenzial, sehr große Mengen an Antworten zu liefern. Im Gegensatz zu anderen klassischen Verbreitungsmedien, z.B. per Post oder durch direkte Verteilung an die Befragten, ist die Kontrolle der erreichten Personengruppen geringer und auch die Rücklaufquoten, die bei klassischen Fragebögen und direkter Verteilung bei bis zu 100% liegen, sind mit 1-2% online wesentlich geringer. Es lässt sich auch wesentlich schwieriger überprüfen, ob die Angaben im Fragebogen der Wahrheit entsprechen oder ob ein Befragter mehrere Fragebögen mit widersprüchlicher Information abgegeben hat. Trotzdem erfreuen sich Online Fragebögen großer Beliebtheit, da sie nicht nur einfach zu verbreiten sind, sondern auch automatisch ausgewertet werden können.

Durch die Überprüfung der Plausibilität der Antworten direkt bei der Eingabe lassen sich viele Eingabefehler direkt vermeiden. Darüber hinaus kann ein Online Fragebogen dynamisch aufgebaut sein und nachfolgende Fragen in Abhängigkeit von bereits beantworteten Fragen auswählen. Hat jemand z.B. sein Alter mit 70 Jahren angegeben, können Fragen folgen, die auf die Besonderheiten älterer Benutzer zugeschnitten sind. Bei einem Online Fragebogen können Fragen zufällig gemischt werden, um Reihenfolgeeffekte zu unterbinden. Größter Nachteil des Online Fragebogens ist die Ungewissheit, ob man tatsächlich eine repräsentative Menge von Befragten erreicht. Häufig werden Anfragen für Online Fragebögen über soziale Netzwerke verbreitet und erreichen damit nur eine bestimmte gesellschaftliche Gruppe, wie z.B. Studierende oder Hausfrauen. Verwendet man zur Verbreitung der Fragebögen eine *micro Job* Plattform wie z.B. **Mechanical Turk**<sup>2</sup>, dann erhält man auch gleichzeitig die benötigten Probanden, allerdings mit noch weniger Kontrolle über deren Auswahl und die Qualität der Ergebnisse.

Inzwischen existiert eine Reihe von Webdiensten, die es erlauben, Online Fragebögen einfach und kostengünstig zu erstellen<sup>3</sup> und alle gängigen Antworttypen zur Verfügung stellen, wie z.B. Checkboxes, Likert-Skalen und Semantische Differenziale. Trotzdem sollte man einige Punkte beim Erstellen eines Online Fragebogens beachten (nach Andrews et al. [3]):

- Fragebogen zunächst entsprechend der Regeln aus Abschnitt 11.3.1 entwerfen, und eine erste Papierversion des Fragebogens erstellen.
- Verbreitungsstrategien für den Fragebogen entwerfen. Diese kann z.B. die Verbreitung per E-Mail oder per sozialem Netzwerk beinhalten.
- Aus dem Papierfragebogen wird eine elektronische Version erzeugt, z.B. mithilfe einschlägiger Webdienste (siehe oben). Dabei sollten die besonderen Möglichkeiten des Online Fragebogens berücksichtigt werden, wie z.B. die Überprüfung von Eingaben und die Möglichkeit, zu unterschiedlichen Fragen zu verzweigen.
- Überprüfen der Funktionalität des Fragebogens und der Information, die über die Befragten gespeichert wird. Eventuell die rechtliche Situation im jeweiligen Land

<sup>2</sup><https://www.mturk.com/>

<sup>3</sup>z.B. die Dienste <http://limesurvey.org>, <http://questionpro.com/> und <http://docs.google.com>

klären, z.B. um zu prüfen, ob IP-Adressen gespeichert und verwendet werden dürfen. Falls ein webbasierter Dienst verwendet wird, sollte auch geprüft werden, auf welchem Server die Daten gespeichert und weiterverarbeitet werden.

- Gründliches Testen der Online Fragebogens mit einer kleinen Gruppe von Benutzern oder Experten, um mögliche Probleme zu erkennen, die nach einem Roll-out des Fragebogens nicht mehr zu beheben sind.

Wurde ein Online Fragebogen veröffentlicht, sollten größere Korrekturen nicht mehr vorgenommen werden, da diese das Ergebnis der Umfrage beeinflussen.

## 11.4 Fokusgruppen

Interviews finden in der Regel zwischen einem Interviewer und einem Befragten statt. Manchmal kann es aber sinnvoll sein, eine Befragung mit einer Gruppe durchzuführen. Eine weit verbreitete Form solcher Befragungen sind **Fokusgruppen**. Die Fokusgruppe setzt sich aus verschiedenen Stakeholdern des zukünftigen Systems zusammen und im Rahmen des UCD ist es durchaus sinnvoll, verschiedene Fokusgruppen zu befragen. Ein Beispiel: Im Rahmen der Entwicklung eines neuen Bibliotheks-Ausleihsystems sollten nicht nur die Leser der Bibliothek, sondern auch die Bibliothekare und die Verwaltungsangestellten zu ihren Erwartungen, Bedürfnissen und Wünschen an das System befragt werden, und zwar vorzugsweise getrennt, um gegenseitige Beeinflussung auszuschließen.

In Gruppen kommen häufig Aspekte der Zusammenarbeit zur Sprache, die in Einzelbefragungen eventuell untergehen würden. Gruppenteilnehmer können sich gegenseitig helfen und so gewährleisten, dass alle relevanten Aspekte diskutiert werden. Es besteht allerdings die Gefahr, dass nicht jeder seine tatsächliche Meinung äußert, sondern sich stattdessen dem allgemeinen Gruppenmeinungsbild anschließt. Um solche Effekte zu verringern, werden Fokusgruppen von einem Moderator geleitet.

Einige Hilfswerkzeuge können die Arbeit der Gruppe unterstützen. Ein Beispiel ist das **CARD-System**<sup>4</sup> welches aus verschiedenen Kartentypen besteht, die in der Diskussion eingesetzt werden können, um Aktivitäten und Informationsflüsse in einer Organisation zu visualisieren [116].

Die Gespräche mit der Gruppe sollten allerdings nicht zu stark strukturiert werden, um auch Freiräume für die Diskussion völlig neuer Punkte zu lassen. Entscheidend ist, dass die Teilnehmer der Fokusgruppe in die Lage versetzt werden, ihre Meinung im Rahmen des sozialen Gruppenkontexts zu diskutieren. So können bestimmte Aspekte identifiziert werden, die in einem einfachen Interview gegebenenfalls verloren gehen. Üblicherweise werden die Interviews mit Fokusgruppen in Bild und Ton aufgezeichnet, um eine bessere nachträgliche Analyse zu ermöglichen. Es besteht auch die Möglichkeit, nachträglich mit einzelnen Mitgliedern der Fokusgruppen anhand des aufgezeichneten Materials bestimmte Fragestellungen zu vertiefen.

---

<sup>4</sup>Collaborative Analysis of Requirements and Design



## 11.5 Beobachtungen

Einige Aktivitäten lassen sich nicht umfassend durch Interviews und Fragebögen erheben. Insbesondere bei komplexen Abläufen mit heterogenen Strukturen, die sowohl kognitive als auch motorische Leistungen erfordern, fällt es den Betroffenen schwer, die Arbeitsschritte vollständig zu benennen. In solchen Fälle können durch **Beobachtungen** zusätzliche Erkenntnisse gewonnen werden. Beobachtungen können zwar im Labor kontrolliert stattfinden, z.B. indem Stakeholder gebeten werden, bestimmte Abläufe zu simulieren, in der Regel sind jedoch Studien vor Ort im üblichen Umfeld der Stakeholder zielführender. Man spricht in einem solchen Zusammenhang auch von **Feldstudien** (vgl. Abschnitt 13.3.8). Studien im Feld werden als **direkte Beobachtungen** bezeichnet. Im Gegensatz dazu stehen **indirekte Beobachtungen**, die nur die Ergebnisse der Aktivitäten, wie z.B. Tagebucheinträge oder Logbücher nachträglich analysieren. Der Vorteil der direkten Beobachtung liegt in den Erkenntnissen zum Benutzungskontext, den der Beobachter durch persönliche Aufzeichnung und durch die nachträgliche Analyse von aufgezeichnetem Audio- und Videomaterial gewinnt. Nachteilig ist, dass die Präsenz des Beobachters die Gefahr birgt, dass Aktivitäten verfälscht werden. Es muss daher im Vorfeld geklärt werden, ob der Beobachter willkommen ist. Wird innerhalb eines Unternehmens z.B. die Einführung einer neuen Software kontrovers diskutiert, so werden Feldstudien zwangsläufig zu Konflikten führen. Feldstudien sollten gut geplant und sorgfältig durchgeführt werden und häufig wird ein Rahmenwerk zur Strukturierung der Feldstudie eingesetzt, welches sich an folgenden Fragen orientiert (nach Rogers et al. [96, S. 249ff]):

- *Wer* setzt welche Technologien zu welchem Zeitpunkt ein?
- *Wo* werden diese eingesetzt?
- *Was* wird damit erreicht?

Die obigen Fragen nach den Akteuren, den Orten, den Objekten und Zielen können noch weiter verfeinert werden:

- *Orte*: Wie sieht der physikalische Ort aus, wie ist er untergliedert?
- *Akteure*: Wer sind die wichtigsten Personen und ihre Rollen?
- *Aktivitäten*: Welche Aktivitäten werden von den Akteuren ausgeführt und wieso?
- *Objekte*: Welche physikalischen Objekte können beobachtet werden?
- *Handlungen*: Welche konkreten Handlungen nehmen die Akteure vor?
- *Ereignisse*: Welche relevanten Ereignisse können beobachtet werden?
- *Zeit*: In welcher Reihenfolge finden die Ereignisse statt?
- *Ziele*: Welche Ziele werden von den Akteuren verfolgt?
- *Emotionen*: Welche emotionalen Zustände können beobachtet werden?

Eine besondere Form der **direkten Beobachtung** ist die **Ethnografische Studie**. Sie wird erst seit jüngster Zeit im UCD eingesetzt und geht auf die klassische Ethnografie der Anthropologen zu Beginn des 20. Jahrhunderts zurück. Diese entwickelten mit der Ethnografie eine neue Methode zur Erforschung fremder Kulturen und Verhaltensweisen. Der Beobachter, auch als Ethnograf bezeichnet, wird für einen längeren Zeitraum Mitglied der zu beobachtenden Gemeinschaft. Einer der bekanntesten Ethnologen, **Claude Lévi-Strauss** lebte z.B. für seine Studien mehrere Jahre in der Dorfgemeinschaft der Naturvölker des Amazonas und konnte so bahnbrechende Erkenntnisse gewinnen [74]. Ein ähnlicher Ansatz kann im UCD angewendet werden, indem der Beobachter einen längeren Zeitraum intensiv mit Stakeholdern verbringt. Im Gegensatz zur *Feldstudie* sollte dabei zunächst keinerlei Rahmenwerk eingesetzt werden und der Beobachter sollte, ähnlich wie Lévi-Strauss am Amazonas, keine vorherigen Annahmen mitbringen, die die Beobachtung beeinflussen könnten. Der Beobachter wird dann Schritt für Schritt Teil der zu beobachtenden Gruppe und hat so die Möglichkeit, sich besonders gut in die Stakeholder hineinversetzen zu können. Ethnografische Studien können nachträglich mit einem Rahmenwerk oder anderen Beobachtungsverfahren kombiniert werden, um die Erkenntnisse zu dokumentieren.

**Indirekte Beobachtungen** hingegen sind immer dann sinnvoll, wenn entweder die Gefahr besteht, dass der Beobachtete durch direkte Beobachtung beeinträchtigt wird oder eine direkte Beobachtung nicht durchgeführt werden kann oder wenig sinnvoll ist. So ist es z.B. schwierig, Benutzerkontexte über große räumliche Distanzen zu erfassen. Hier müsste dann der Beobachter den Stakeholder wie ein Privatdetektiv verfolgen, was in den meisten Situationen weder zeitlich noch ethisch vertretbar wäre. In solchen Fällen kommen häufig **Tagebuchstudien** mit elektronischen Tagebüchern zum Einsatz, die z.B. über eine Smartphone-App gepflegt werden können. In vielen Fällen stellen diese Apps Erinnerungsfunktionen zur Verfügung, so dass der Benutzer in gewissen Zeitintervallen nach Informationen zu seinem Kontext und der aktuellen Aktivität gefragt wird. In einigen Fällen wird dies auch mit einer Ortserkennung kombiniert, so dass bestimmte Fragen nur an bestimmten Orten gestellt werden. Diese Methode, der gezielten Befragung von Teilnehmern in bestimmten Situationen wird als **Experience Sampling Method (ESM)** bezeichnet [95]. ESM kann in vielfältigen Situationen zum Einsatz kommen, von der Therapie bei Angst- und Suchtpatienten bis zur Untersuchung von sinnvollen Privacy-Einstellungen bei der Verwendung von Smartphones.

Weitere technische Hilfsmittel können den Stakeholdern an die Hand gegeben werden, um Aspekte der Interaktion und der Umgebung zu erfassen und zu dokumentieren. Dies können z.B. Einmalkameras oder kleine digitale Aufnahmegeräte sein, die später ausgewertet werden. Ein Spezialfall solcher Hilfsmittel sind die sogenannten **Cultural Probes**, die ursprünglich eingesetzt wurden, um das kulturelle Leben und Umfeld von Stakeholdern zu erfassen. In einem konkreten Fall sollten beispielsweise die Lebensumstände älterer Personen untersucht werden. Zu diesem Zweck erhielten die Teilnehmer der Studie neben einer Wegwerfkamera weitere Hilfsmittel wie vorfrankierte Postkarten mit bestimmten Fragen und eine Reihe von Stadtkarten, auf denen relevante Orte gekennzeichnet werden konnten [37]. Das Konzept der *Probes* wurde seit seiner Einführung 1999 weiterentwickelt und wird in Bereichen eingesetzt, in denen traditionelle Beobachtungsmethoden an ihre Grenzen stoßen.

## 11.6 Personas und Szenarien

Das Verständnis über die zukünftigen Benutzer eines interaktiven Produktes und die Kontexte, in denen das Produkt verwendet wird, bzw. die Objekte, die neben dem Produkt eine Rolle spielen sollten im Fortlauf des UCD (im rechten Teil des Double-Diamond, Abbildung 10.1) den Entwicklern der Benutzerschnittstelle möglichst einfach zugänglich gemacht werden. Teilweise wird dies in Form von Berichten, Bild-, Video- und Textdokumenten geschehen, auf die der Entwickler zurückgreifen kann, wenn er eine spezielle Auskunft benötigt. Besser und einträglicher ist allerdings die Beschreibung typischer Benutzer des Systems und der häufigsten Situationen, in denen das Produkt eingesetzt werden soll. Eine Benutzerbeschreibung zu diesem Zweck heißt **Persona** und eine Schilderung einer typischen Benutzersituation **Szenario**. Beide Beschreibungen sind voneinander abhängig, da die Benutzer immer Teil des Szenarios sein werden und das Szenario durch die Aktivitäten der Benutzer geprägt wird.

*Personas* sind möglichst konkrete Beschreibungen fiktiver Personen, die aber als typischer Repräsentant einer Stakeholdergruppe gelten können. Die Persona ist ein Steckbrief, der neben einem Namen und dem Alter Hobbies, Vorlieben und auch familiäre Bindungen auflistet. Neben diesen eher allgemeinen Beschreibungen, die einen möglichst holistischen Eindruck geben sollen, enthält die Persona auch konkrete Statements zur Verwendung des zu entwickelnden Produkts, die konkrete Hinweise auf die Gestaltung der Benutzerschnittstelle geben. Personas helfen Entwicklern, sich zu jedem Zeitpunkt vorzustellen, wie der jeweilige Benutzer auf eine Veränderung des Designs der Benutzerschnittstelle reagieren würde. Mithilfe von Personas können sich Entwickler besser in ihre Benutzer hineinversetzen.

### Beispiel: Persona Beschreibung eines Lagerarbeiters

Matthias ist 35 Jahre alt. Er arbeitet seit 3 Jahren als Schichtleiter im Zentrallager der Firma Nil-Versand. Vorher hat er 10 Jahre lang als Lagerarbeiter in der Firma Rhein-Versand gearbeitet. Er hat einen Hauptschulabschluss und eine Lehre als Gleisschlosser abgeschlossen, ist aber auf Grund des besseren Verdiensts in die Logistik-Branche gewechselt. Er ist verheiratet und hat zwei Kinder, die zwölf und neun Jahre alt sind. Matthias liebt Computerspiele und verbringt einen Teil seiner Freizeit mit den neuesten Spielen. Zu diesem Zweck hält er seinen PC immer auf dem neuesten Stand. Für seine Kinder hat er die Familienkonsole Wuu gekauft, mit der er sich gerne auch selbst beschäftigt. Matthias ist stark kurzsichtig, hat aber eine ausgezeichnete Motorik, da er im örtlichen Badminton-Verein als Trainer und Spieler aktiv ist. Da er keine Fremdsprache spricht, macht Matthias mit seiner Frau und den Kindern lieber Urlaub in Deutschland und fährt selten ins Ausland. Matthias hat ein sehr gutes Verhältnis zu seinen Arbeitskollegen und ist dort sehr angesehen. Technologischen Neuerungen in der Firma steht er skeptisch gegenüber.

*Szenarien* ergänzen Personas indem sie Geschichten erzählen, in denen Benutzer Technologie in verschiedenen Kontexten einsetzen. Szenarien finden nicht nur im UCD Verwendung, sondern allgemein in der Softwareentwicklung und im Produktdesign. Sie sind eine wichtige Schlüsselkomponente, um erfolgreiches Design umzusetzen und werden in allen modernen Entwicklungsabteilungen eingesetzt [2]. Der Kernbestandteil eines Sze-

narios ist eine Geschichte, in der ein zukünftiger Benutzer, der durch eine Persona ausführlich beschrieben wird, die Hauptrolle spielt. Die Geschichte sollte detailliert sein und einen reichen Kontext zur Verfügung stellen. Die Interaktion mit einem neuen Produkt sollte zwar prominent vertreten sein, beansprucht aber nicht einen überwiegenden Anteil an der Geschichte. Mehrere Szenarien mit den gleichen Protagonisten geben dem Entwickler die Möglichkeit, das zukünftige Produkt aus verschiedenen Blickwinkeln zu bewerten und die Benutzerschnittstelle entsprechend anzupassen. Personas und Szenarien werden im Rahmen des UCD allen Entwicklern zur Verfügung gestellt. Sie sollten auch während des Entwicklungsprozesses angepasst und erweitert werden. Neue Szenarien und neue Personas sollten hinzugefügt werden, sobald sich neue Erkenntnisse, z.B. durch die Evaluierung von Prototypen im UCD ergeben.

#### Beispiel: Szenario für ein neuartiges Virtual-Reality Gerät

Matthias hat sich die neue Spiele-Brille *Okkultes Drift* bei Nil bestellt, weil seine Freunde aus dem Badminton-Verein beim letzten Training davon geschwärmt haben. Er öffnet das Paket sofort nachdem er es vom Postboten in Empfang genommen hat und überprüft den Inhalt. Die Brille ist leichter als er gedacht hatte und lässt sich problemlos an seine Kopfform anpassen. Er stellt fest, dass er das Gerät nur mit Kontaktlinsen verwenden kann, da die Optik der *Okkultes Drift* nur bedingt seine Sehschwäche kompensieren und er unter der *Okkultes Drift* seine normale Brille nicht verwenden kann. Der Anschluss an den Computer funktioniert erst, nachdem er die neusten Treiber von der Webseite geladen hat. Leider unterstützt sein Lieblingsspiel *Blind Craft* die neue Brille nicht, so dass er zunächst zum Testen mit einem anderen Spiel vorlieb nehmen muss. Die tolle Rundumsicht der Brille fasziniert Matthias sehr, und die leichte Übelkeit, die ihn nach 15 Minuten Benutzung überkommt, nimmt er dafür gerne in Kauf. Während des ersten Spiels spürt Matthias plötzlich ein Klopfen auf der Schulter. Es ist sein jüngster Sohn, der die Brille auch einmal ausprobieren möchte. Dazu muss Matthias die Kopfanpassung verändern. In der kleinsten Einstellung ist die Brille gerade klein genug, damit sie seinem Sohn nicht vom Kopf rutscht.

### Übungsaufgaben:



1. Beobachten Sie ihren Tagesablauf. Benutzen Sie dazu die Alarmfunktion ihres Smartphones oder ihrer Uhr, um sich alle 20 Minuten unterbrechen zu lassen. Notieren Sie ihren jeweiligen Aufenthaltsort, die Aktivität, die anwesenden Personen und die Objekte mit denen Sie gerade interagiert haben. Am Ende des Tages erstellen Sie eine Liste aller Aktivitäten, Orte, Objekte und Personen. Diskutieren Sie verblüffende Erkenntnisse.
2. Untersuchen Sie die Benutzbarkeit einer Webseite ihrer Wahl mithilfe eines Fragebogens. Erstellen Sie dazu eine Variante des Fragebogens ausschließlich mit Likert-Skalen und einer zweiten Variante, die ausschließlich das Semantische Differenzial als Antwortform verwendet (jeweils etwa 10 Fragen). Diskutieren Sie die Vor- und Nachteile beider Fragebögen.
3. Ihre Designfirma hat den Auftrag erhalten, ein neuartiges tragbares 3D-Fernsehgerät zu entwickeln. Überlegen Sie sich, wer die Stakeholder für ein solches Gerät sind und entwerfen Sie zwei unterschiedliche Personas dazu. Entwickeln Sie zusätzlich ein Szenario zu der Verwendung des neuen Geräts.

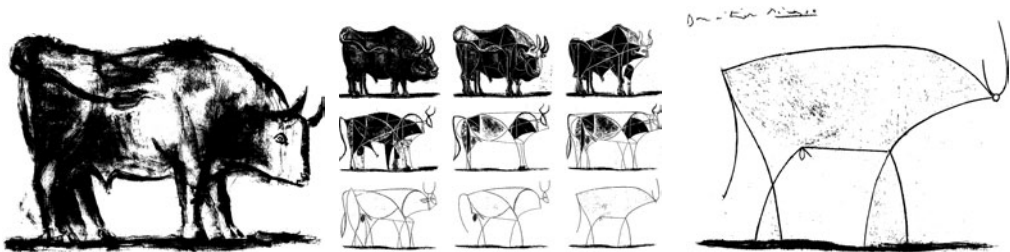


# 12 Skizzen und Prototypen

Im Zyklus aus Entwurf, Umsetzung und Evaluation (vgl. Kapitel 10) besteht eine ganze Reihe von Möglichkeiten zur Umsetzung eines interaktiven Systems. In frühen Konzeptionsphasen können grafische Darstellungen beispielsweise oft **skizziert** werden, wobei nur die wesentlichen Inhalte dargestellt und unwesentliche Teile bewusst weggelassen werden. In späteren Phasen kann man mehr oder weniger funktionale **Prototypen** evaluieren und auch dabei bewusst die jeweils unwichtigen Aspekte weglassen. Neben der schnelleren Umsetzung haben Skizzen und Prototypen also auch die Funktion der gezielten Vereinfachung und Fokussierung auf das Wesentliche.

## 12.1 Eigenschaften von Skizzen

In der bildenden Kunst werden einfache grafische Darstellungen als **Skizzen** bezeichnet. Diese Darstellungen zeigen oft die Idee oder das Konzept für ein späteres detailreicheres Werk. Sie können jedoch auch das Ergebnis einer bewussten Vereinfachung sein und haben einen ästhetischen Wert an sich. Pablo Picasso fertigte beispielsweise eine ganze Reihe von immer abstrakteren Lithografien eines Stiers an, um dieses Tier schließlich in der Darstellung ganz auf sein Wesen zu reduzieren (Abbildung 12.1).



**Abbildung 12.1:** Abstraktionsserie eines Stiers in Lithographien von Pablo Picasso. Diese Serie zeigt den Übergang vom *Abbild* zum *Urbild*.



mmibuch.de/s/12.1

Das Konzept der Skizze findet sich auch in anderen kreativen Fächern wie Architektur oder Musik, und auch in der Mathematik, beispielsweise als Beweisskizze. Eine solche Beweisskizze gibt die wesentlichen Schritte eines Beweises an ohne alle Schritte im Detail auszuführen. Auch Patentschriften enthalten bewusst einfach gehaltene Skizzen zur

Darstellung technischer Konstruktionen. Dem Fachmann genügt eine solche Skizze, um die Durchführbarkeit und Wirkungsweise des Beweises oder der Konstruktion zu beurteilen. Allen Skizzen gemeinsam ist eine Reihe von Eigenschaften, wie in **Bill Buxtons** Standardwerk zum Thema Skizzen in der MMI [17] dargelegt:

- Skizzen sind *schnell und billig* anzufertigen oder erwecken zumindest diesen Eindruck. Dadurch sind sie auch immer *zum richtigen Zeitpunkt* verfügbar, also genau dann, wenn sie gebraucht werden.
- Skizzen sind zum *Wegwerfen* gemacht, nicht für die Dauer. Das bedeutet natürlich nicht, dass sie wertlos sind. Der Wert einer Skizze besteht in der dargestellten Funktion, nicht in der Qualität der Darstellung. Sobald man so viel Energie in eine Skizze gesteckt hat, dass man sie nicht mehr leichten Herzens wegwerfen könnte, ist es keine Skizze mehr.
- Skizzen sind *selten Einzelstücke*. Sie sind meist Bestandteil einer Sammlung oder Reihe von Varianten.
- Skizzen *deuten an* und *untersuchen*, statt etwas *festzulegen*. Darin unterscheiden sie sich beispielsweise maßgeblich von Prototypen.
- Skizzen sind absichtlich *mehrdeutig*. Im Zusammenspiel von Skizze und Zeichner bzw. Betrachter entstehen so neue Ideen: Der Zeichner lässt bestimmte Interpretationsspielräume bewusst offen und dem Betrachter fällt daraufhin womöglich eine Interpretation ein, an die der Zeichner gar nicht gedacht hatte.
- Skizzen verwenden ein *eigenes grafisches Vokabular* und einen *charakteristischen Stil*, wie z.B. nicht ganz gerade Linien oder über den Endpunkt hinaus gezogene Striche. Diese Art der Darstellung vermittelt die bewusste Ungenauigkeit der Darstellung.
- Skizzen zeigen einen *minimalen Detailgrad* und gerade *so viele Feinheiten wie nötig*. Zusätzliche Details, die für die jeweilige Fragestellung nicht relevant sind, vermindern sogar den Wert einer Skizze, da das Risiko besteht, dass sie von den wichtigen Aspekten ablenken.

Die bewusste Ungenauigkeit von Skizzen ist also kein Mangel, sondern eine positive Eigenschaft. Die nicht vorhandenen Details können von jedem Betrachter nach Belieben gedanklich ergänzt werden, was auch neue Ideen hervorbringen kann, und in jedem Fall dazu führt, dass der Betrachter sich nicht an den falschen Details stört. Skizzen sind daher auch bewusst mehrdeutig: Eine grafische Benutzerschnittstelle, die mit Bleistift auf Papier skizziert ist, wird nicht nach ihrem Farbschema beurteilt, sondern allein nach ihrer grafischen Struktur. Ein Beispiel für eine solche UI-Skizze ist auch das Titelbild dieses Buches.

Obwohl es sich oft lohnt, Skizzen aufzubewahren, um später wieder zu früheren Ideen zurückkehren zu können, sind Skizzen ihrer Natur nach doch ein Wegwerfprodukt. Die Skizze ist zwar ein Schritt auf dem gedanklichen Weg von der Idee zur Umsetzung, sie findet jedoch in der endgültigen Umsetzung keine Verwendung mehr, sondern bleibt ein Werkzeug für unterwegs.



## 12.2 Eigenschaften von Prototypen

Im Gegensatz zu Skizzen treffen **Prototypen** bestimmte klare Aussagen zu einem Entwurf oder einer Konstruktion. Ein Prototyp setzt bestimmte Teile der Funktionalität eines geplanten Produktes oder Systems so um, dass diese Teile ausprobiert werden können. Der Prototyp eines Fahrrades kann beispielsweise aus ein paar Stangen und einem Sattel bestehen, wenn es nur darum geht, die Funktion zu überprüfen, dass man darauf sitzen kann. Geht es hingegen um die Funktion des Fahrens, dann sind sich drehende Räder wichtig, auf den Sattel kann jedoch verzichtet werden. Die im Prototypen umgesetzte Funktionalität kann also unterschiedlich verteilt sein.

### 12.2.1 Auflösung und Detailgenauigkeit

Die **Auflösung** eines Prototypen beschreibt den Umfang seiner Umsetzung. Sind nur wenige Teile des geplanten Systems umgesetzt (beispielsweise nur ein Startbildschirm), dann bedeutet dies eine niedrige Auflösung, sind schon viele Teile umgesetzt, dann ist die Auflösung hoch. Diese Terminologie wurde von Houden und Hill [51] eingeführt und ist eigentlich etwas irreführend, da man in der Umgangssprache den Umfang der Umsetzung eher als Funktionsumfang oder eben Umfang bezeichnen würde. Der Begriff hat sich jedoch im Interaktionsdesign eingebürgert. Die **Detailgenauigkeit** eines Prototypen beschreibt, wie nahe er bereits am geplanten System ist. Eine Papierskizze des groben Bildschirmlayouts hat also eine niedrigere Detailgenauigkeit als eine detaillierte Stiftzeichnung mit allen Bildelementen, und diese wiederum eine niedrigere als ein am Computer gezeichnetes Mockup des Bildschirms, in dem schon alle Farbschattierungen festgelegt sind.

Auflösung und Detailgenauigkeit eines Prototypen entscheiden darüber, für welche Art von Feedback er geeignet ist. Zeigt man dem Benutzer einen sehr detailliert grafisch umgesetzten Startbildschirm eines Programms, so kann man seine Meinung zu dieser grafischen Gestaltung und Aspekten wie Farben, Schriften und Wortwahl bekommen. Zeigt man ihm einen funktional sehr weit umgesetzten, dafür grafisch nicht ausgestalteten Prototypen, beispielsweise als funktionierendes Programm mit einfachen Bildschirmmasken und Blindtext, so kann der Benutzer viel eher Fragen zur logischen Struktur der Bedienabläufe beantworten. Wie bei Skizzen gilt es hier, den Detailgrad als Kombination von Auflösung und Detailgenauigkeit für die Fragestellung richtig zu wählen.

### 12.2.2 Horizontale und Vertikale Prototypen

Beim Prototyping von Software gibt es zwei häufig verwendete Arten, die Auflösung im Prototypen zu variieren. Betrachtet man die Funktionalität eines Programms beispielsweise als Baumstruktur mit Funktionen und Unterfunktionen, dann ist ein **horizontaler Prototyp** einer, der zu jeder vorhandenen Funktion den obersten Menüeintrag oder die erste Bildschirmmaske darstellt. Ein **vertikaler Prototyp** verzichtet auf diese Vollständigkeit an der Oberfläche, implementiert dafür aber eine bestimmte Funktion vollständig in die Tiefe. Auch Kombinationen dieser beiden Arten sind möglich.



mmibuch.de/a/12.2

**Abbildung 12.2:** Papierprototyp eines mobilen Gerätes. Hier ist nicht nur der Bildschirminhalt, sondern gleich das ganze Gerät als Prototyp umgesetzt. Knöpfe und Schalter können einfach gezeichnet werden.

### 12.2.3 Wizard of Oz Prototypen

In manchen Fällen ist es noch gar nicht möglich die tatsächliche Funktionalität eines geplanten Systems in einem Prototypen umzusetzen, oder es wäre viel zu aufwändig. In diesem Fall können Funktionen durch einen menschlichen Helfer so ausgeführt werden, dass es aussieht, als ob das Programm die Funktion ausgeführt hätte. In Anlehnung an **L. Frank Baums** Geschichte vom *Zauberer von Oz*, der sich letztendlich auch als Simulation durch einen Menschen entpuppte, heißen solche Prototypen auch **Wizard of Oz Prototypen**.

## 12.3 Papierprototypen

Eine häufig verwendete Art von Prototypen sind die **Papierprototypen**. Diese bestehen aus mehr oder weniger detaillierten Zeichnungen geplanter Bildschirminhalte, und zwar mit Stift auf einem Blatt Papier. Verschiedene Bildschirminhalte werden als verschiedene Zeichnungen umgesetzt. Der Versuchsperson wird nun von einem Helfer

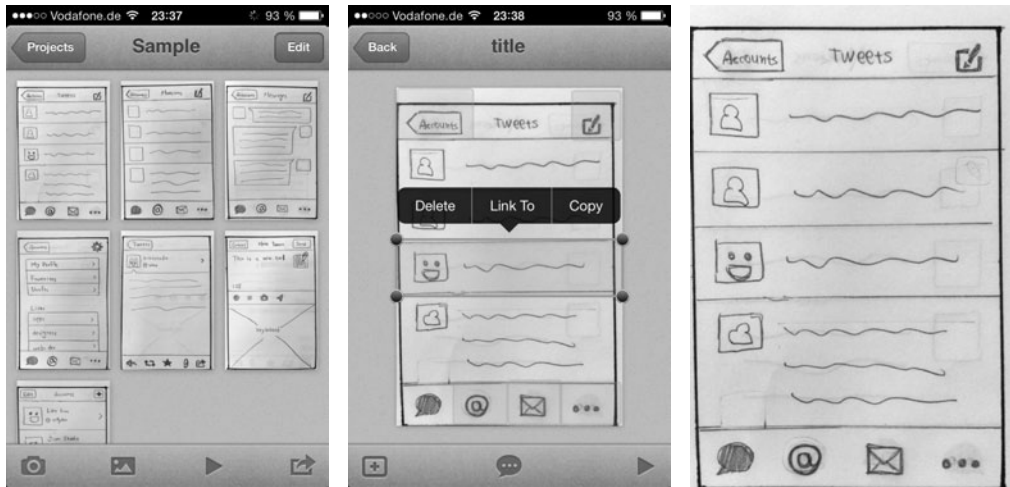


Abbildung 12.3: PopApp: ein Werkzeug zur Verwendung von Papierprototypen auf Smartphones



der Startbildschirm vorgelegt und abhängig von den getätigten Eingaben wird er gegen den jeweils passenden nächsten Bildschirm ausgetauscht. Statt kompletter Bildschirme können auch einzelne Masken, Menüs oder Dialogfelder ausgetauscht werden.

Diese Art der Simulation stellt sehr stark die eigentlichen Bedienabläufe in den Vordergrund. Die konkrete grafische Umsetzung wird durch die gezielt skizzenhafte grafische Umsetzung in den Hintergrund gestellt. Die Detailgenauigkeit kann also relativ niedrig sein, während die Auflösung bei entsprechendem Fleiß sehr hoch sein kann. Papierprototypen bieten außerdem den Vorzug, dass auf unvorhergesehene Eingaben oder Bedienschritte schnell reagiert werden kann, indem Elemente des Prototypen einfach mit einem Stift ergänzt werden.

Bisher wurden Papierprototypen vor allem bei der Entwicklung grafischer Benutzerschnittstellen für Bildschirme verwendet, da sich Bildschirm Inhalte gut zeichnen lassen. Auch für mobile Anwendungen und Geräte sind Papierprototypen ein sehr geeignetes Mittel (Siehe Abbildung 12.2). In letzter Zeit findet man zudem auch Ansätze, diese etablierte Form des Prototyping auf neue Interaktionsparadigmen, wie physikalische Interaktion zu übertragen [129]. Eine interessante Mischform stellt die iOS-Anwendung **PopApp** dar (Siehe Abbildung 12.3). Sie ermöglicht es, Papierprototypen von iOS-Anwendungen weiterhin mit Stift und Papier zu zeichnen, diese dann aber auf dem mobilen Gerät in funktionale **Klick-Prototypen** zu verwandeln. Hierzu werden die verschiedenen Bildschirm Inhalte wie oben beschrieben gezeichnet. Danach werden sie mit der Kamera des mobilen Gerätes abfotografiert. In den Aufnahmen werden dann Klick-Bereiche definiert, die jeweils zu einem anderen (gezeichneten) Bildschirm führen.

Ein derart digitalisierter Papierprototyp besitzt auf dem Bildschirm selbst immer noch die erwünschte skizzenhafte niedrige Detailgenauigkeit, während das gesamte Umfeld (Formfaktor des Geräts, Bedienung im Gehen) sehr realistisch ist und daher eine hohe Detailgenauigkeit besitzt.

#### Exkurs: Loslassen können

Eine sehr positive Eigenschaft von Papierprototypen kommt vor allem in kleinen Entwicklungsteams zum tragen, bei denen der Entwickler gleichzeitig die Tests der Prototypen durchführt. Hier wird es plötzlich wichtig, dass ein Papierprototyp ohne eine einzige Zeile Programmierung auskommt. Die gezeichneten Bildschirme werden vom Entwickler deshalb nicht als bereits entwickelte Bestandteile des geplanten Systems wahrgenommen. Eine Wiederverwendung von Programmcode ist ausgeschlossen, und statt einiger Tage Programmieraufwand hat der Entwickler nur einige Stunden Aufwand in das Zeichnen investiert. Liefert der Benutzertest nun ein kritisches oder negatives Ergebnis, so fällt es dem Entwickler emotional viel leichter, die schlechte Idee zu verwerfen und nach einer anderen Lösung zu suchen. Bei einem schon teilweise programmierten Prototypen wäre diese Hemmschwelle wesentlich höher.



## 12.4 Video Prototypen

Eine weitere zeitgemäße Form von Prototypen sind sogenannte **Video-Prototypen**. Viele interaktive Systeme werden heute mobil genutzt und Interaktion findet in den verschiedensten Alltagssituationen statt. Ein derart reichhaltiges Umfeld lässt sich in Papier- oder Software-Prototypen alleine oft nicht darstellen. Ein Video-Prototyp ist nichts anderes als eine Filmsequenz, die die Bedienung eines interaktiven Systems im geplanten Kontext zeigt und besitzt damit Ähnlichkeiten zu den in Abschnitt 11.6 vorgestellten **Szenarien**. In welcher Form das betreffende System dabei dargestellt wird, ist oft nicht so wichtig.

Beispielsweise kann ein Mobiltelefon einfach durch ein Stück Pappe dargestellt werden, wenn der eigentliche Bildschirminhalt unwichtig ist und stattdessen der gesamte Handlungsablauf in der Umgebung im Vordergrund steht. Video-Prototypen können ihrerseits auch wiederum Papier-Prototypen verwenden, wenn bis zu einem gewissen Detailgrad Bildschirminhalte dargestellt werden sollen.

Eine Stärke von Video-Prototypen ist die Darstellung der Dimension Zeit. Da die Bedienung im zeitlichen Verlauf gezeigt wird, wird dieser automatisch gut dargestellt. Durch Skalieren der Zeit (Zeitlupe, Zeitraffer) oder filmische Ausdrucksmittel wie Schnitte und Überblendungen können auch sehr lange oder sehr kurze Zeiträume vermittelt werden. Monatelange Interaktionen lassen sich so in wenigen Minuten vermitteln, was mit keiner anderen Prototyping-Technik möglich ist. Auch technisch (noch) unmögliche Dinge sind im Film möglich, da mit beliebigen Tricks gearbeitet werden kann. Einfache Video-Prototypen können mit einem Mobiltelefon oder einer Kompaktkamera ohne Schnitt



**Abbildung 12.4:** Szene aus dem SUN Starfire Video Prototyp von 1992. Darin wurde ein komplettes Szenario mit viel Sorgfalt umgesetzt, das zeigt, wie man aus damaliger Sicht in Zukunft (2004) leben und arbeiten könnte.



in wenigen Minuten gedreht werden, eine passende Vorbereitung der Szenen vorausgesetzt. Die kompetente Verwendung filmischer Ausdrucksmittel steigert natürlich die Ausdrucksstärke eines Video-Prototypen. Ein guter Schnitt, eventuell auch Zwischentitel oder ein Erzähler, zusammen mit einem Drehbuch, das eine gute Geschichte erzählt, sind Merkmale aufwändigerer Video-Prototypen.

Schließlich lassen sich im Video auch ganze Visionen zukünftiger Systeme darstellen. Ein Beispiel dafür, das eine gewisse Berühmtheit erlangt hat, ist das SUN Starfire Video<sup>1</sup> von 1992 (Abbildung 12.4). In diesem Video wird ein Szenario gezeigt, wie wir in Zukunft (aus der damaligen Sicht) arbeiten könnten. Viele der im Video gezeigten Technologien waren damals noch nicht anwendungsreif, in Forschungslabors jedoch schon ansatzweise vorhanden. Heute, mehr als zwei Jahrzehnte später, ist manches davon längst kommerzielle Realität (Videokonferenzen, Kollaboration am Bildschirm, schnelle Bildverarbeitung), während anderes noch immer Forschungsgegenstand ist (gekrümmte

<sup>1</sup><http://www.asktog.com/starfire/>

Displays) oder nie umgesetzt wurde. Das Starfire Video zeigt auf hervorragende Weise, wie interaktive Technologien in die Arbeitsabläufe und eine Erzählhandlung integriert werden und so eine insgesamt glaubwürdige Geschichte entsteht.



### Übungsaufgaben:

1. Schlüpfen Sie in die Rolle eines Studenten und konzipieren Sie eine mobile Anwendung, die dessen Partyleben organisiert (Verwaltung von Einladungen und Gegeneinladungen, Bestimmen eines Autofahrers, Loggen des Getränkekonsums, etc.). Erstellen Sie für diese Anwendung einen Papierprototypen.
2. Verwenden Sie den eben erstellten Papierprototypen für die Produktion eines Videoprototypen. Dabei sollen die wichtigsten Funktionen jeweils in einem realistischen Interaktionskontext gezeigt werden. Setzen Sie sich ein Zeitlimit von wenigen Stunden und arbeiten Sie mit einfachen und vertrauten Mitteln, wie z.B. der Kamera in ihrem Smartphone.
3. Entwerfen Sie auf Grundlage der Papier- und Videoprototypen einen Klick-Prototypen mit einem Tool Ihrer Wahl (z.B. Pop-App oder Axure<sup>a</sup>).

---

<sup>a</sup><http://www.axure.com/>

# 13 Evaluation

In der Informatik gibt es verschiedenste Kriterien, nach denen ein Computersystem oder ein Stück Software bewertet werden kann. Prozessoren werden meist nach ihrer Rechenleistung bewertet, gelegentlich auch nach ihrer Energieeffizienz. Ein Maß für die Qualität von Algorithmen ist deren Speicherplatz- oder Laufzeitverhalten, und Software wird beispielsweise danach bewertet, wie robust und fehlerfrei sie entwickelt wurde. Sobald wir aber interaktive Systeme betrachten, müssen wir immer das Zusammenspiel zwischen Mensch und Maschine betrachten (vgl. auch Kapitel 1). Neben Robustheit und Effizienz ist dabei die Bedienbarkeit des Systems durch den Menschen das zentrale Kriterium. Da Menschen aber keine technischen Vorrichtungen sind, ist es viel schwieriger, sie objektiv zu messen und zu bewerten, und es existiert ein ganzes Repertoire verschiedener Techniken hierzu, die jeweils ihre eigenen Stärken und Schwächen haben.

## 13.1 Arten der Evaluation

### 13.1.1 Formativ vs. Summativ

Zunächst kann man verschiedene Arten der Evaluation nach ihrem Verwendungszweck unterscheiden: **Formative Evaluationen** informieren den Entwicklungsprozess. Sie sind einer Design-Entscheidung zeitlich vorgelagert und ihr Ergebnis hat einen Einfluss auf die anstehende Entscheidung. Formative Evaluationen von Konzeptideen werden beispielsweise in der linken Hälfte des *Double-Diamond* (vgl. Kapitel 10) durchgeführt. **Summative Evaluationen** fassen das Ergebnis eines Entwicklungsprozesses zusammen. Sie liefern eine abschließende Bewertung für eine oder mehrere Design-Entscheidungen. Summative Evaluationen werden klassischerweise am Ende großer Projekte durchgeführt, um die erreichte Qualität zu überprüfen (rechte Hälfte des *Double-Diamond*). Da der Entwicklungsprozess im User-centered Design (**UCD**) iterativ verläuft, also aus ständigen Zyklen von Konzeption, Umsetzung und Evaluation besteht, haben viele Evaluationen in der Praxis sowohl formative als auch summative Aspekte: Die Bewertung eines Prototypen liefert in den allermeisten Fällen neue Erkenntnisse für dessen Weiterentwicklung.

### 13.1.2 Quantitativ vs. Qualitativ

Eine weitere Dimension, entlang derer man verschiedene Arten der Evaluation unterscheiden kann, ist die Art ihrer Ergebnisse: **Quantitative Evaluationen** liefern quantifizierbare, also in Zahlen ausdrückbare Ergebnisse. Dies können beispielsweise Messwerte wie Ausführungszeiten oder Fehlerraten sein, oder generell jede numerische



Bewertung, mit der sinnvoll gerechnet werden kann. Daten von Fragebögen, in denen Bewertungen beispielsweise auf **Likert-Skalen** erhoben werden (siehe Abschnitt 11.3.2) gehören dazu nicht. Trotzdem finden sich immer wieder Publikationen, in denen zu solchen Daten fälschlicherweise Durchschnittswerte angegeben werden.

Im Gegensatz dazu liefern **qualitative Evaluationen** nur Aussagen, die sich nicht numerisch fassen lassen. Beispiele hierfür sind frei formulierte Kommentare von Benutzern in Interviews oder Fragebögen, sowie Notizen im Rahmen einer Nutzer-Beobachtung. Qualitative und Quantitative Methoden ergänzen sich oft sinnvoll: während die quantitativen Verfahren statistisch auswertbare und damit wissenschaftlich allgemein als belastbar anerkannte Aussagen liefern, runden die qualitativen Methoden das Bild einer Studie ab. In Benutzerkommentaren und Beobachtungen stecken oft interessante Details und Aspekte, die sich aus rein numerischen Daten nicht entnehmen lassen.

### 13.1.3 Analytisch vs. Empirisch

Die dritte Dimension, entlang derer wir Evaluationstechniken unterscheiden können, ist schließlich deren grundlegende Herangehensweise: **Analytische Evaluationen** untersuchen ein System, indem sie es analysieren, also beispielsweise seine Arbeitsweise, Bestandteile oder Eigenschaften betrachten und erklären. **Empirische Evaluationen** befassen sich lediglich mit den Ergebnissen bei der Bedienung eines Systems und beachten dabei nicht, wie diese zustande gekommen sind. Prominentestes Beispiel für eine empirische Methode sind die später in diesem Kapitel beschriebenen kontrollierten Experimente. In Anlehnung an **Michael Scriven** [100] (Übersetzung in [131]) könnte man diesen Unterschied auch wie folgt beschreiben: Will man ein Werkzeug, wie z.B. eine Axt evaluieren, dann kann man dies entweder analytisch tun und untersuchen, welcher Stahl für die Haue und welches Holz für den Griff verwendet wurde, wie die Balance der Axt oder die Schärfe der Klinge ist, oder man wählt den empirischen Weg und misst, wie gut die Axt in der Hand eines guten Holzfällers ihren Zweck erfüllt, also wieviel Holz er in gegebener Zeit damit fällt. Analytische und empirische Evaluation sind (wie qualitative und quantitative Methoden) ebenfalls komplementär. Während empirische Untersuchungen wissenschaftlich belastbare Aussagen über die unterschiedliche Leistung zweier Systeme machen, liefern analytische Verfahren die möglichen Erklärungen hierfür: Messergebnisse ohne vernünftige Erklärung sind nur von begrenztem Wert, und umgekehrt nutzt die beste Analyse nichts, wenn ihre Ergebnisse sich nicht mit dem in der Realität beobachteten Verhalten decken. In den folgenden Abschnitten sollen einige wichtige Vertreter der verschiedenen Arten von Evaluation vorgestellt werden.

## 13.2 Analytische Methoden

Analytische Methoden der Evaluation besitzen den Charme, dass sie prinzipiell ganz ohne echte Benutzer auskommen. Damit sind sie prädestiniert für eine schnelle und kostengünstige Evaluation von Konzepten, sowie für eine Evaluation in Situationen, in denen ein Konzept noch geheim bleiben soll. Das offensichtliche Risiko dabei ist jedoch, dass die gefundenen Probleme nicht wirklich diejenigen der späteren Benutzer sind, sondern Artefakte der verwendeten Evaluationsmethode.



### 13.2.1 Cognitive Walkthrough

Bei einem **Cognitive Walkthrough** werden die Interaktionen eines fiktiven Benutzers mit dem System Schritt für Schritt durchgeführt und dabei auftretende Probleme dokumentiert. Hierfür muss einerseits der fiktive Benutzer genau charakterisiert werden (z.B. durch eine Persona, siehe Abschnitt 11.6) und andererseits die auszuführende Handlung bzw. das zu erreichende Ziel genau spezifiziert werden. Nun werden die gedanklichen Schritte des Benutzers nacheinander simuliert, wobei man davon ausgeht, dass der Benutzer immer den einfachsten oder offensichtlichsten Weg wählen wird. Die dabei in jedem Schritt gestellten Fragen orientieren sich am Modell der Ausführung zielgerichteter Handlungen (vgl. Abschnitt 5.3.1):

1. Ist die korrekte Aktion zur Ausführung einer Handlung ausreichend klar? Weiß der Benutzer überhaupt, was er tun soll?
2. Ist die korrekte Aktion als solche erkennbar? Findet sie der Benutzer?
3. Erhält der Benutzer eine ausreichende Rückmeldung nach Ausführung der Aktion, so dass er erkennen kann, dass die Handlung erfolgreich durchgeführt ist?

Während dieser simulierten Ausführung der Handlung werden in jedem Schritt potenzielle Probleme und deren Quellen notiert und auch sonstige Beobachtungen festgehalten. So steht am Ende eine konkrete Liste potenzieller Probleme, die ggf. noch nach ihrer Schwere kategorisiert werden und dann in einer neuen Version des Systems behoben werden können. Die Qualität des Ergebnisses dieser Technik steht und fällt damit, wie gut der fiktive Benutzer simuliert wird, und ob der Tester tatsächlich in jedem Schritt dessen Vorkenntnisse und Überlegungen richtig nachvollzieht. Die exakte Charakterisierung des Nutzers ist daher sehr wichtig.

### 13.2.2 Heuristische Evaluation

Die **Heuristische Evaluation** wurde durch **Jacob Nielsen** als sogenannte **Discount Usability** Methode eingeführt. Sie formalisiert die Vorgehensweise bei der analytischen Evaluation etwas stärker und basiert auf Jacob Niensens zehn **Heuristiken**:

1. *Visibility of system status*: Das System sollte den Benutzer immer darüber informieren, was gerade passiert, beispielsweise durch zeitiges und angemessenes Feedback.
2. *Match between system and the real world*: Das System sollte die Sprache des Benutzers sprechen und ihm vertraute Formulierungen verwenden statt technischer Fachbegriffe. Es sollte Konventionen aus der echten Welt einhalten und Informationen logisch und nachvollziehbar anordnen.
3. *User control and freedom*: Benutzer wählen Systemfunktionen oft versehentlich aus und sollten dann einen einfach erkennbaren Ausweg aus diesem ungewollten Systemzustand vorfinden. Wichtig hierfür sind beispielsweise **Undo** (rückgängig machen) und **Redo** (Funktion wiederholen).

4. *Consistency and standards*: Benutzer sollten sich nicht fragen müssen, ob verschiedene Worte, Situationen oder Aktionen dasselbe bedeuten. Konventionen der jeweiligen Plattform sollten eingehalten werden.
5. *Error prevention*: Noch besser als Fehlermeldungen ist eine sorgfältige Gestaltung der Schnittstelle, die Fehler gar nicht erst entstehen lässt. Situationen, in denen Fehler wahrscheinlich sind, sollten gar nicht erst auftreten und vor potenziell fehlerhaften Eingaben ist ein Bestätigungsdialog sinnvoll.
6. *Recognition rather than recall*: Die Gedächtnisbelastung des Benutzers lässt sich verringern, indem Objekte, Aktionen und Optionen sichtbar und erkennbar gemacht werden (siehe auch Abschnitt 3.1.2). Der Benutzer sollte keine Informationen von einem Dialogschritt zum nächsten im Kopf behalten müssen. Bedienungshilfen sollten immer sichtbar oder zumindest leicht zu finden sein.
7. *Flexibility and efficiency of use*: Kurzbefehle und Tastaturkürzel – für den Neuling oft nicht erkennbar – können dem Experten eine sehr schnelle und effiziente Bedienung ermöglichen. So kann das System Neulinge und Experten gleichermaßen ansprechen (**Flexibilität**).
8. *Aesthetic and minimalist design*: Dialoge sollten keine irrelevanten oder selten benötigten Informationen enthalten. Jede unnötige Information konkurriert um die Aufmerksamkeit des Benutzers mit den relevanten Informationen und senkt deren relative Sichtbarkeit.
9. *Help users recognize, diagnose, and recover from errors*: Fehlermeldungen sollten eine einfache Sprache verwenden und keine Codes enthalten, das Problem präzise benennen, und konstruktiv eine Lösungsmöglichkeit vorschlagen.
10. *Help and documentation*: Obwohl ein System im Idealfall ohne Dokumentation bedienbar sein sollte, kann es manchmal nötig sein, Hilfe anzubieten. Hilfe und Dokumentation sollten einfach zu finden und zu durchsuchen sein. Sie sollten auf die Aufgaben des Benutzers fokussiert sein, konkrete Bedienschritte aufzeigen, und nicht zu umfangreich sein.

Bei der Durchführung einer heuristischen Evaluation werden diese sehr allgemeinen Kriterien in eine konkrete Checkliste überführt, die sich auf das zu evaluierende System bezieht. So lassen sich aus der Heuristik 1. *Visibility of system status* bei der Evaluation einer Website beispielsweise die folgenden konkreten Fragen ableiten:

- 1.1. Kann ich auf jeder Seite erkennen, wo ich mich gerade innerhalb der gesamten Website befinde?
- 1.2. Hat jede Seite einen aussagekräftigen Titel (z.B. zur Verwendung in Lesezeichen)?
- 1.3. Kann ich den Zweck und Inhalt jeder einzelnen Seite (z.B. Produkt-Information, Werbung, Reviews, Profil und Impressum) klar erkennen?

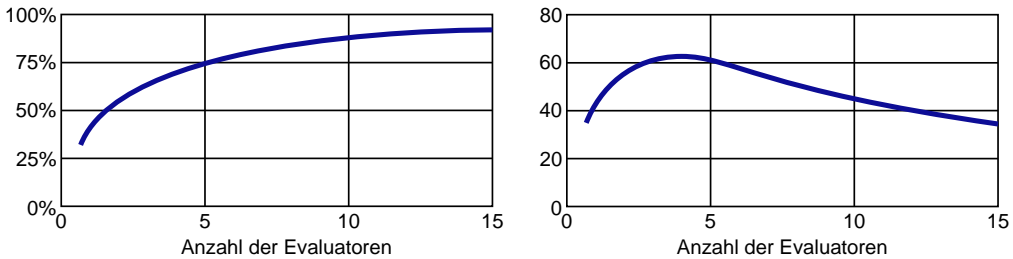


Abbildung 13.1: Links: Anzahl der gefundenen Probleme in Abhängigkeit von der Anzahl der Evaluatoren, Rechts: Verhältnis aus Nutzen und Kosten



Mit dieser abgeleiteten konkreten Checkliste kann dann das zu testende System Schritt für Schritt systematisch untersucht werden. Zur Durchführung dieser Untersuchung an allen Teilen eines Systems sind nicht unbedingt Usability Experten notwendig, da die Kriterien ja in der Checkliste klar festgelegt sind. Bei der Ableitung der Kriterien aus den allgemeinen Heuristiken wird hingegen die Expertise eines Usability Experten gebraucht. Ein gewisses Risiko im Zusammenhang mit heuristischen Evaluationen besteht darin, viele problematische Details zu identifizieren, die bei der späteren Benutzung womöglich vom Benutzer gar nicht als problematisch empfunden werden. Hier ist es wichtig, bei der Ableitung der konkreten Kriterien aus den allgemeinen Heuristiken mit Augenmaß vorzugehen.

Eine prinzipielle Einschränkung der Methode ist die Festlegung auf Nielsens zehn Heuristiken. Diese gelten bei Weitem nicht für alle denkbaren Kontexte. So ist beispielsweise die Heuristik *Aesthetic and minimalist design* ganz klar unserem westlichen Kulturkreis und auch hier nur unserem gegenwärtigen Stilempfinden zuzuordnen und müsste in anderen Kulturkreisen wie z.B. Indien oder China durch andere Kriterien ersetzt werden. Die Heuristik *Recognition rather than recall* bezieht sich auf Situationen, in denen Effizienz das oberste Ziel ist, wirkt aber beispielsweise negativ bei einem Spiel, das Rätsel-Elemente enthält. Auch für andere Nutzergruppen wie z.B. Kinder ist unklar, welches Gewicht die einzelnen Heuristiken noch behalten.

Nielsen stellt in seiner Darstellung der heuristischen Evaluation<sup>1</sup> auch eine interessante Überlegung bezüglich der Anzahl der benötigten Evaluatoren an: Unter der Annahme, dass kein Evaluator perfekt arbeitet, sondern jeweils nur einen bestimmten Anteil der auftretenden Fehler findet, lässt sich beschreiben, wie die gesamte Anzahl der gefundenen Fehler mit der Anzahl der Evaluatoren steigt und sich asymptotisch den 100% annähert. Andererseits steigt aber der Aufwand (also insbesondere die Kosten) linear mit der Anzahl der Evaluatoren an. Betrachtet man die Effizienz als Quotient aus Nutzen und Aufwand, so hat diese ein Maximum bei einer relativ kleinen Anzahl von Evaluatoren, beispielsweise vier oder fünf (siehe Abbildung 13.1).

<sup>1</sup><http://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/>

### 13.2.3 GOMS und KLM

Ein recht altes Modell zur analytischen Beschreibung von Bildschirm-Interaktionen mittels Tastatur und Maus ist der **GOMS** Formalismus [21] von 1983 in Verbindung mit dem Keystroke Level Model (**KLM**) von 1980 [20]. In GOMS werden Interaktionen des Menschen beschrieben durch **Ziele** (Goals), Operatoren, Methoden und Auswahlregeln (Selection Rules). Dabei lässt sich ein Ziel prinzipiell durch verschiedene Methoden erreichen. Jede dieser Methoden besteht aus einer Folge verschiedener Operatoren, und aus den möglichen Methoden wird mittels der Auswahlregeln eine gewählt. Damit können auch komplexe Interaktionsabläufe hierarchisch strukturiert und analysiert werden. GOMS gibt uns dabei eine *top-down* Perspektive auf den Ablauf.

Das Keystroke Level Model geht den umgekehrten Weg, also *bottom-up*: Es beschreibt für elementare Operationen, wie beispielsweise das Drücken einer Taste oder den Wechsel von der Tastatur zur Maus, die zu deren Ausführung benötigte Zeit. Card, Moran und Newell ermittelten experimentell eine Tabelle solcher durchschnittlich benötigten Zeiten, und wenn die genaue Folge von Operationen für eine komplexere Aktion bekannt ist, dann lässt sich damit durch bloßes Aufsummieren eine Vorhersage über die zur Ausführung benötigte Zeit machen. Dabei unterscheiden die Autoren fünf grundlegende Operationen (hier mit den zugehörigen durchschnittlichen Zeiten):

- *K* (Keystroke): Das Drücken einer Taste benötigt bei einem mittelmäßigen Maschinenschreiber etwa  $t_K = 0.28$  Sekunden.
- *P* (Pointing): Das Zeigen auf eine Bildschirmposition mit der Maus benötigt im experimentellen Durchschnitt etwa  $t_P = 1.1$  Sekunden.
- *H* (Homing): Der Wechsel zwischen Tastatur und Maus benötigt im experimentellen Durchschnitt etwa  $t_H = 0.4$  Sekunden.
- *M* (Mental preparation): Das geistige Vorbereiten einer nachfolgenden Operation benötigt im experimentellen Durchschnitt etwa  $t_M = 1.35$  Sekunden.
- *R(t)* (Response time t by the system): Antwortzeit des Systems von  $t$  Sekunden.

Nun betrachten wir einen Benutzer mit dem Ziel, eine Datei zu löschen, wobei die Hände vorher und hinterher auf der Tastatur sein sollen. Für dieses Ziel existieren in einer grafischen Benutzerschnittstelle unter anderem die beiden folgenden Methoden:

- *M1*: Wechsel zur Maus, Bewegen des Mauszeigers zur Datei, Anklicken der Datei, Ziehen zum Mülleimer und Loslassen, Wechsel zur Tastatur
- *M2*: Wechsel zur Maus, Selektieren der Datei, Wechsel zur Tastatur, Taste *Entf*.

Damit lassen sich die Ausführungszeiten wie folgt berechnen:

- $t_{M1} = t_H + t_P + t_K + t_P + t_H = 0.4 + 1.1 + 0.28 + 1.1 + 0.4 = 3.28s$
- $t_{M2} = t_H + t_P + t_H + t_K = 0.4 + 1.1 + 0.4 + 0.28 = 2.18s$

Daraus geht hervor, dass die Methode *M2* schneller ist, was sich mit dem Zweck von Tastaturkürzeln deckt, Zeit zu sparen. Auch komplexere Interaktionen lassen sich auf diese Weise ohne funktionierenden Prototypen analysieren. Das KLM ermöglicht es also, Interaktionen rein analytisch zu bewerten bevor mit der Implementation überhaupt begonnen wird.

Dabei gibt es die prinzipbedingten Einschränkungen, dass man bei dieser Art der Analyse immer von einer fehlerfreien Interaktion ausgeht, und dass ein durchschnittlicher Benutzer zugrunde gelegt wird. Bei den heutigen vielfältigen Formen der Interaktion rückt das für Bildschirm-Arbeitsplätze konzipierte und bewusst einfach gehaltene KLM auch deshalb immer weiter in den Hintergrund, weil sich viele Dinge, wie z.B. wechselnde Umgebungssituationen oder Ablenkung darin nicht modellieren lassen.

## 13.3 Empirische Methoden

Als **empirische** Methoden werden alle Formen der Evaluation bezeichnet, die durch Messung oder anderweitige Sammlung in **Experimenten**, **Beobachtungen** oder **Befragungen** Daten erheben, auf deren Basis sich wissenschaftliche Aussagen machen lassen. *Empirische* und *analytische* Evaluationen verhalten sich zueinander komplementär: Mittels empirischer Methoden lässt sich herausfinden *dass* ein bestimmter Sachverhalt auftritt, und mit analytische Methoden *warum* (vgl. Scriven [100]). Einen hervorragenden Überblick über den Entwurf, die Auswertung und Darstellung Wissenschaftlicher Experimente geben Field und Hole [33]. Für die wissenschaftliche Aussagekraft empirischer Untersuchungen gibt es bestimmte Qualitätskriterien:

- **Objektivität** bedeutet, dass die erhobenen Daten unabhängig von der spezifischen Messmethode oder den Erwartungen und Hypothesen des Experimentators sind. Zeiten oder Distanzen lassen sich einfach objektiv messen, Stimmungen oder Erlebnisse viel schwieriger.
- **Reproduzierbarkeit** bedeutet, dass die Untersuchung hinreichend genau beschrieben ist, damit andere Forscher sie wiederholen können und dabei zu den gleichen Ergebnissen gelangen.
- **Validität** bedeutet, dass die so ermittelten Ergebnisse nur genau das messen, was sie auch messen sollen (**interne Validität**) und repräsentativ für die Allgemeinheit sind (**externe Validität**). Wollte man beispielsweise Intelligenz durch wiegen der Gehirnmasse messen, dann wäre dies kein intern valides Maß. Wählt man für eine Studie zum Verhalten aller Autofahrer nur achtzehnjährige Probanden, dann wäre das Resultat nicht extern valide.
- **Relevanz** bedeutet, dass die Ergebnisse tatsächlich neue Erkenntnisse liefern und daher in irgendeiner Weise nützlich sind. Eine Studie, die objektiv, reproduzierbar und valide nachweist, dass Wasser bergab fließt, wäre beispielsweise nicht mehr wirklich relevant.

Bevor wir uns nun dem Entwurf und der Auswertung verschiedener empirischer Methoden zuwenden, müssen noch einige weitere Begriffe geklärt werden.

### 13.3.1 Variablen und Werte

Daten, oder Merkmale, die wir in einer empirischen Untersuchung messen oder einstellen, werden **Variablen** genannt. Dabei heißen die Daten, durch die Umgebung fest vorgegeben sind oder die wir als Experimentator kontrollieren und gezielt einstellen **unabhängige Variablen**, da sie nicht vom beobachteten Prozess abhängen. Diejenigen Daten, die als Ergebnis der Untersuchung gemessen werden und vom beobachteten Prozess abhängen, heißen **abhängige Variablen**. Ein Beispiel: Nehmen wir an, wir wollen messen, wie sich die Arbeitsumgebung auf die Geschwindigkeit beim Schreiben eines Lehrbuches auswirkt. Die verschiedenen Arbeitsumgebungen, die uns interessieren, sind das Büro und der Schreibtisch zuhause. Die unabhängige Variable *Umgebung* hat also die beiden möglichen Werte *Büro* und *zuhause*. Der beobachtete Prozess ist das Schreiben des Lehrbuches. Als abhängige Variable messen wir die Schreibgeschwindigkeit in *Seiten pro Tag*. Nun gibt es verschiedene Arten von Daten, die wir messen oder kontrollieren können:

- **Nominale Daten** benennen einfach verschiedene Kategorien, ohne dabei eine Ordnung vorzugeben. Beispiele dafür sind Ländernamen oder Fußballmannschaften oder die eben genannten Arbeitsumgebungen.
- **Ordinale Daten** lassen sich sortieren, also in eine Ordnung bringen, aber es ergibt keinen Sinn, mit ihnen zu rechnen. Beispiele hierfür sind Tabellenplätze in der Bundesliga oder Bewertungen beim online-shopping.
- **Kardinale Daten** geben Zahlenwerte an, mit denen sinnvoll gerechnet werden kann. Dabei können *diskrete* Daten nur bestimmte Werte annehmen und *stetige* Daten auch alle Zwischenwerte. Beispielsweise ist die Kinderzahl einer Familie immer eine nichtnegative ganze Zahl, während die Körpergröße in gewissen Grenzen beliebige Werte annehmen kann. Mit beiden kann man sinnvoll rechnen (doppelt so groß, doppelt so viele Kinder).

Die Art der Daten, die wir erheben, hat eine maßgebliche Auswirkung auf die weitere Verarbeitung und Darstellung der Ergebnisse, wie wir weiter unten sehen werden.

### 13.3.2 Probanden

Die Teilnehmer einer empirischen Studie heißen **Probanden** und ihre Auswahl ist entscheidend für die **externe Validität** der Studie. Jeder einzelne Proband wird durch seine **demografischen Daten** wie Alter, Geschlecht oder Bildungsstand charakterisiert. Darüber hinaus kann es je nach Inhalt der Studie auch weitere wichtige Kriterien geben, wie beispielsweise die Erfahrung im Umgang mit dem untersuchten Gerät, Rechts- oder Linkshändigkeit, oder Sehkraft. Dabei sollte die Auswahl (**Stichprobe**) der Probanden möglichst repräsentativ sein für den gesamten Teil der Bevölkerung (**Grundgesamtheit**), über den man eine Aussage machen will. Ein Beispiel: Will man ein Seniorentelefon entwickeln und dessen Bedienbarkeit anhand eines Prototypen überprüfen, dann ist es sinnlos, dafür zwanzigjährige Studenten als Probanden zu nehmen. Die Ergebnisse, die eine solche Studie liefern würde, ließen sich nämlich höchstens auf

junge Erwachsene übertragen, nicht jedoch auf die eigentliche Zielgruppe der Senioren. Dass an Universitäten viele empirische Untersuchungen mit Studenten als Probanden durchgeführt werden, ist ein stetiger Kritikpunkt in der wissenschaftlichen Community, weil es die Aussagekraft der gewonnenen Ergebnisse oft auf diese Bevölkerungsgruppe einschränkt. Für die Anzahl der benötigten Probanden gibt es keine allgemeine Regel. Will man statistisch verwertbare Ergebnisse erzielen, dann wird dies um so einfacher, je mehr Probanden zur Verfügung stehen. Selbstverständlich steigt damit auch der Aufwand für die Studie, was Zeit und Kosten angeht. In der Praxis richtet sich die Zahl der Probanden oft nach dem experimentellen Design. Will man beispielsweise vier verschiedene Werte einer unabhängigen Variable vergleichen und jeder Proband kann nur einen Wert testen, dann sollte die Probandenzahl ein Vielfaches von vier sein, damit jeder Wert von gleich vielen Probanden getestet wird. In der Mensch-Maschine-Interaktion wird oft mit relativ kleinen Probandenzahlen zwischen zehn und 50 gearbeitet, die gerade eben statistisch verwertbare Ergebnisse liefern. In der Psychologie und Medizin werden oft erheblich größere Zahlen verwendet.

Bei der Rekrutierung von Probanden ist es wichtig, deren Einverständnis formal korrekt einzuholen. Man sollte ihnen erklären, worum es in der Studie geht, jedoch selbstverständlich ohne ein gewünschtes Ergebnis zu implizieren. Außerdem sollten sie darüber informiert werden, dass sie die Teilnahme jederzeit beenden können und dass ihre Daten komplett vertraulich bzw. anonymisiert und nur für den Zweck dieser Studie verwendet werden. Beispiele für solche Einverständniserklärungen finden sich vielfach im Internet und weitere Informationen sind in der Regel bei der Ethikkommission der jeweiligen Universität zu erfragen.

### 13.3.3 Beobachtungsstudien

Eine einfache Form empirischer Studien sind sogenannte **Beobachtungsstudien** (siehe auch Abschnitt 11.5). Dabei werden Prozesse beobachtet, ohne dass man in ihren Ablauf gezielt eingreift oder unabhängige Variablen kontrolliert. Stattdessen werden die Probanden verschiedenen Werten einer unabhängigen Variable zugeordnet, je nach dem, in welcher Kategorie sie sich von Natur aus befinden. Ein Beispiel: In einem Semester nehmen 108 Studierende an der Vorlesung Mensch-Maschine-Interaktion 1 teil. Davon besuchen 50 Prozent den freiwilligen Übungsbetrieb. Die unabhängige Variable *Übungsteilnahme* hat also die beiden möglichen Werte **ja** und **nein** und zu jedem Wert gehören jeweils 54 Probanden. Bei der Klausur am Ende des Semesters messen wir die abhängige Variable *Note* und es stellt sich heraus, dass die Noten der Übungsteilnehmer im Schnitt besser ausfallen als die der Nichtteilnehmer. Mit den passenden statistischen Auswertungen lässt sich daraus entnehmen, dass eine **Korrelation** zwischen *Übungsteilnahme* und *Note* besteht. Hierbei handelt es sich nicht um ein echtes kontrolliertes Experiment, sondern um ein **Quasi-Experiment**, da die unabhängige Variable *Übungsteilnahme* eben nicht vom Experimentator gesteuert wird, sondern sich aus dem Verhalten der Probanden ergibt.

Leider lässt sich aus der so ermittelten **Korrelation** keine **Kausalität** ableiten, denn es könnte ja sein, dass die *Note* nicht einfach nur von der *Übungsteilnahme* abhängt, sondern beides von einer dritten, **konfundierenden Variable** *Interesse*, die einen Charakterzug des Probanden beschreibt und die beiden Werte *interessiert* und *desinteressiert*



annehmen kann. In diesem Fall würden *interessierte* Studierende eher am Übungsbetrieb teilnehmen und durch ihr Interesse (nicht durch die Übungen!) auch den Stoff besser verstehen und damit eine bessere *Note* in der Klausur erzielen. Die bessere Note wäre in diesem Fall also nicht durch die Übungsteilnahme, sondern lediglich durch das Interesse des Studierenden zu begründen. Leider erlaubt es eine Beobachtungsstudie nicht, die beiden verschiedenen Fälle voneinander zu unterscheiden. Durch eine einfache Änderung ließe sich dieses *Quasi-Experiment* jedoch in ein echtes kontrolliertes Experiment überführen.

### 13.3.4 Kontrollierte Experimente

Im nächsten Semester nehmen wieder 108 Studierende an der Vorlesung MMI1 teil. Davon werden 50 Prozent zufällig ausgewählt, die zur Teilnahme am Übungsbetrieb gezwungen werden, während der Rest daran nicht teilnehmen darf. Die unabhängige Variable *Übungsteilnahme* hat also wieder die beiden möglichen Werte **ja** und **nein** und zu jedem Wert gehören jeweils 54 Probanden. In diesem Fall wurde die Zuordnung der Probanden zu einem der beiden Werte jedoch zufällig bestimmt und deshalb können wir davon ausgehen, dass sich die *interessierten* und die *desinteressierten* Studierenden gleichmäßig auf diese beiden Gruppen verteilen. Bei der Klausur am Ende des Semesters messen wir wieder die abhängige Variable *Note* und es stellt sich erneut heraus, dass die Noten der Übungsteilnehmer im Schnitt besser ausfallen als die der Nichtteilnehmer. Durch die zufällige Zuweisung können wir den Einfluss der dritten Variable *Interesse*, von der sowohl *Übungsteilnahme* als auch *Note* abhängen, in diesem Fall jedoch ausschließen. Das Beispiel zeigt, wie wichtig es ist, den genauen Aufbau des Experiments zu planen. Da alle relevanten unabhängigen Variablen kontrolliert werden, handelt es sich um ein **kontrolliertes Experiment**.

Zu Beginn eines jeden Experiments steht eine wissenschaftliche **Hypothese**, die man mit dem Experiment bestätigen möchte. Im obigen Beispiel ist die Hypothese:

$H$ : Übungsteilnehmer erzielen bessere Noten in der Klausur.

Um eine solche Hypothese zu beweisen, greift man jedoch in der Regel zu einem Trick: Da man die Größe des Unterschiedes im Durchschnittswert nicht genau vorhersagen kann, widerlegt man stattdessen das Gegenteil der Hypothese und hat damit gezeigt, dass die Hypothese selbst gilt, unabhängig davon, wie groß der Unterschied ist. Das Gegenteil der Hypothese ist die sogenannte **Nullhypothese**:

$H_0$ : Übungsteilnehmer und Nichtteilnehmer erzielen im Mittelwert die gleichen Noten in der Klausur.

Hat man gezeigt, dass dies nicht gilt, indem man im Experiment nachweist, dass es einen nicht zufällig entstandenen Unterschied gibt, dann kann man an den im Experiment ermittelten Werten gleich auch sehen, wie groß dieser Unterschied (**Effektgröße**) ist. Als Nächstes muss man dann eine im Experiment zu lösende **Aufgabe** (engl. **task**) festlegen. Will man beispielsweise zwei Methoden der Texteingabe vergleichen, dann



wäre eine sinnvolle Aufgabe die Eingabe eines vorgegebenen Textes. Will man die Effizienz einer Webseite für Online-Shopping messen, dann wäre ein sinnvoller Task das Einkaufen eines Gegenstandes. In beiden Fällen wären die benötigte Zeit sowie die Anzahl der gemachten Fehler sinnvolle Messgrößen als **abhängige Variablen**. Bleiben wir bei dem Beispiel der Vorlesung, dann besteht der experimentelle Task im Schreiben der Klausur und gemessen wird die erzielte Note.

Zur Widerlegung der Nullhypothese wird ein geeignetes **experimentelles Design** entworfen. Dazu werden die unabhängigen und die abhängigen Variablen festgelegt, sowie die verschiedenen **experimentellen Bedingungen** und ihre Aufteilung auf die Probanden. Im obigen Beispiel sind die beiden Bedingungen die beiden möglichen Werte der Variable *Übungsteilnahme*. Gibt es mehrere unabhängige Variablen, dann ergeben sich die möglichen Bedingungen aus allen möglichen Kombinationen der Variablenwerte. Will man beispielsweise den Erfolg der Übungsteilnahme in drei verschiedenen Studienfächern ermitteln, dann gibt es eine weitere unabhängige Variable *Fach* mit den (beispielhaften) Werten **MMI1**, **Analysis** und **Algebra**. Damit ergeben sich  $3 * 2 = 6$  mögliche Bedingungen:

<i>Fach:</i>		<b>MMI1</b>	<b>Analysis</b>	<b>Algebra</b>
<i>Übungsteilnahme:</i>	<b>Ja</b>	Bedingung 1	Bedingung 2	Bedingung 3
	<b>Nein</b>	Bedingung 4	Bedingung 5	Bedingung 6

Nun kann man entscheiden, ob jeder Proband alle Bedingungen ausführen soll (**within-subjects Design**) oder ob jeder Proband nur eine bestimmte Bedingung ausführt und man über diese verschiedenen Gruppen hinweg vergleicht. (**between-groups Design**). Im obigen Beispiel bietet sich ein between-groups Design an, da sonst jeder Student jede Vorlesung zweimal hören müsste, was im wirklichen Leben nicht viel Sinn ergibt. Außerdem lässt sich so die Studie in nur einem Semester durchführen, während man andernfalls 6 Semester braucht. Im Gegenzug werden insgesamt mehr Probanden gebraucht, aber mit der im Beispiel genannten Probandenzahl  $n = 108$  liegen wir für sechs experimentelle Bedingungen in einer vernünftigen Größenordnung, da so auf jede Bedingung immer noch 18 Probanden entfallen.

Führt jeder Proband alle Bedingungen aus, dann spielt die Reihenfolge der Bedingungen eine wesentliche Rolle: Käme Bedingung 1 beispielsweise immer vor Bedingung 4 dran, dann könnte es sein, dass die Noten beim zweiten Hören der MMI Vorlesung konsistent besser sind, egal ob an Übungen teilgenommen wurde oder nicht. Dies wäre ein **Lerneffekt**, der im Studium ja tatsächlich so auch gewünscht ist, für das kontrollierte Experiment jedoch stört. Umgekehrt könnte aber auch ein **Ermüdungseffekt** eintreten, der dazu führt, dass man beim zweiten Hören der Vorlesung schon gelangweilt oder ermüdet ist und damit eine schlechtere Note erzielt. Beide Effekte lassen sich ausschalten, indem man die Reihenfolge der experimentellen Bedingungen variiert: Man kann diese Reihenfolge zufällig festlegen (**Randomisierung**) und bei einer sehr großen Zahl von Probanden sind die genannten Effekte damit ausgeschlossen. Für kleinere Probandenzahlen bietet es sich jedoch eher an, systematisch vorzugehen (**Gegenbalancierung**). Man könnte beispielsweise alle möglichen Reihenfolgen zuweisen (**Permutation**). Damit ergeben sich im obigen Beispiel  $6! = 6 * 5 * 4 * 3 * 2 * 1 = 720$  mögliche Reihenfolgen, und man braucht eine Probandenzahl, die ein Vielfaches von

720 ist. Da auch dies in den meisten Fällen unrealistisch ist, wird hier die Methode der **Lateinischen Quadrate** (engl. **latin square**) eingesetzt: Dabei wird nur gefordert, dass paarweise zwischen allen Bedingungen jede Reihenfolge gleich oft vorkommt und dass jede Bedingung einmal an jeder Position vorkommt. Ein lateinisches Quadrat für unsere Beispielstudie sieht wie folgt aus:

Bed. 6	Bed. 1	Bed. 5	Bed. 2	Bed. 4	Bed. 3
Bed. 5	Bed. 6	Bed. 4	Bed. 1	Bed. 3	Bed. 2
Bed. 2	Bed. 3	Bed. 1	Bed. 4	Bed. 6	Bed. 5
Bed. 1	Bed. 2	Bed. 6	Bed. 3	Bed. 5	Bed. 4
Bed. 4	Bed. 5	Bed. 3	Bed. 6	Bed. 2	Bed. 1
Bed. 3	Bed. 4	Bed. 2	Bed. 5	Bed. 1	Bed. 6

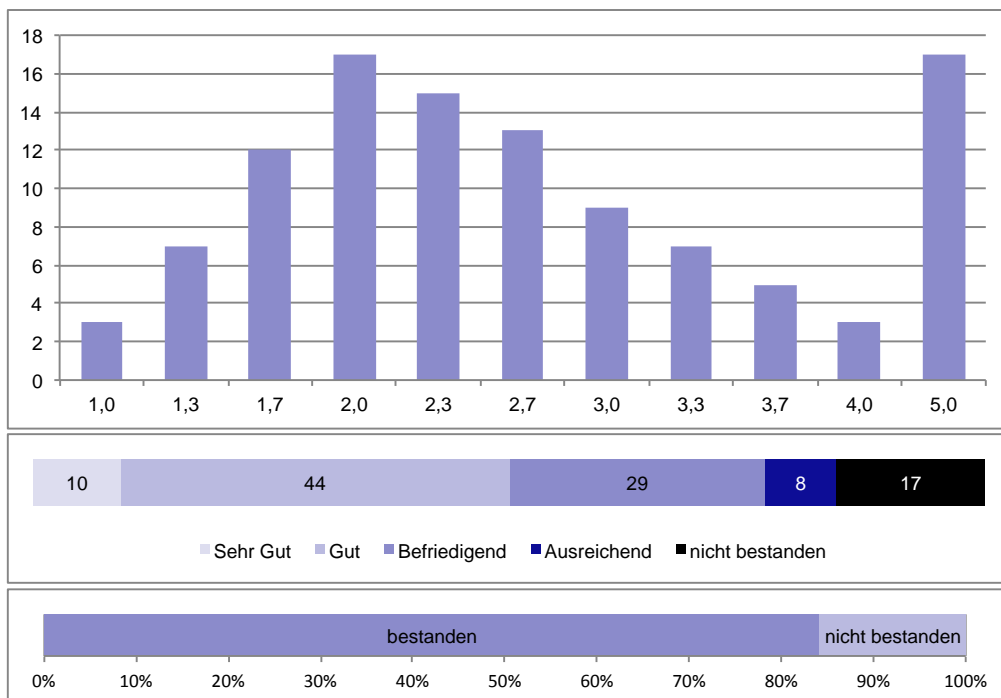
Somit können wir mit 6 verschiedenen Reihenfolgen eine Gegenbalancierung der 6 experimentellen Bedingungen erreichen und damit die Fehlerquellen durch die Reihenfolge der Ausführung ausschalten. Bei insgesamt  $n=108$  Probanden führen also jeweils 18 Probanden die Bedingungen in der gleichen Reihenfolge aus.

### 13.3.5 Darstellung der Ergebnisse

Bei der Durchführung eines Experimentes werden die abhängigen Variablen gemessen und zunächst einmal in einer großen Tabelle aufgezeichnet. Man erhält also eine Zuordnung von experimentellen Bedingungen und Probanden zu gemessenen Werten. Im obigen Beispiel ist dies die Notentabelle. Bei einem *between groups* Design wäre das eine Tabelle mit allen Klausurteilnehmern und der jeweiligen Note im Fach MMI, sowie der Information, ob sie am Übungsbetrieb teilgenommen haben oder nicht. Abhängig von der Art der Daten können diese nun unterschiedlich verarbeitet und dargestellt werden.

Bei nominalen Daten wie Noten bietet sich die Darstellung als Histogramm an. Dabei wird zu jedem nominalen Wert die Häufigkeit seines Auftretens dargestellt. Wird die MMI1 Klausur beispielsweise nach den üblichen Notenstufen von 1,0 bis 5,0 bewertet, dann wäre ein **Histogramm** (Abbildung 13.2 oben) eine angemessene Darstellung. Hierbei ist zu beachten, dass Notenstufen an sich erst einmal keine quantitativen Daten sind. Eine erreichte Punktzahl hingegen wäre ein quantitativer Wert, da man sinnvoll sagen kann, Student A habe doppelt so viele Punkte wie Student B. Auch quantitative Daten lassen sich so darstellen, indem man sie in verschiedene Äquivalenzklassen einteilt, beispielsweise auf ganze Zahlen rundet. Die Verteilung über so entstehende Kategorien ist wieder als Histogramm darstellbar und die Zusammenfassung hilft uns dabei, andere Zusammenhänge zu kommunizieren. Sollen nur wenige Werte dargestellt werden, dann braucht ein herkömmliches Histogramm unnötig viel Platz und man kann die Verteilung platzsparender als horizontale Balken (Abbildung 13.2 Mitte und unten) darstellen. Für den besonderen Anwendungsfall bei **Likert-Skalen** (siehe Abschnitt 11.3.2) lassen sich solche balkenförmigen Histogramme schnell, einfach und konsistent mit einem Online-Tool erzeugen (siehe Abbildung 13.3).

Erhobene Daten kann man mit den Mitteln der deskriptiven Statistik zusammenfassen und vergleichen. Nehmen wir als Beispiel die Zahlen (5, 5, 2, 5, 4, 5, 5). Um diese Reihe



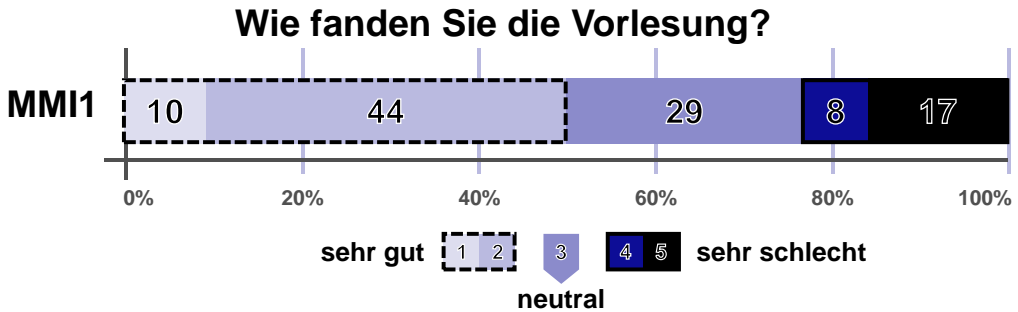
**Abbildung 13.2:** Histogrammdarstellung einer Verteilung über nominalen Daten, hier der Drittel-Notenstufen einer Klausur (oben), der ganzen Notenstufen einer Klausur (Mitte) sowie der Unterscheidung bestanden oder nicht bestanden (unten). Alle beruhen auf den gleichen Daten.



mmibuch.de/s/13.2

von Zahlen auf einen vergleichbaren Wert zu reduzieren, kann man folgende Größen abhängig von der Art der Daten angeben:

- Bei nominalen Daten den **Modus**: Der Modus einer Zahlenreihe ist die darin häufigste einzelne Zahl, im obigen Beispiel also die 5. Ein mögliches Problem mit dem Modus ist die Tatsache, dass mehrere Zahlen gleich häufig vorkommen können. Im Datensatz zu Abbildung 13.2 beispielsweise treten die Noten 2,0 und 5,0 gleich häufig auf. Da es zwei Modi gibt, nennt man eine solche Verteilung eine **bimodale Verteilung**.
- Bei ordinalen Daten *zusätzlich* den **Median**: Der Median wird berechnet, indem man die Zahlenfolge aufsteigend sortiert (2, 4, 5, 5, 5, 5, 5) und dann die Zahl an der mittleren Position herausgreift, hier also eine 5. Bei einer geraden Anzahl von Zahlen wird der Mittelwert der beiden mittleren Zahlen genommen. Im Vergleich zum Mittelwert verhält sich der Median robuster gegenüber **Ausreißern**, also einzelnen Werten, die sich von der großen Menge der Werte stark unterscheiden.



mmibuch.de/a/13.3

**Abbildung 13.3:** Balkendiagramm zur Darstellung der Ergebnisse von Likert-Skalen, generiert mit <http://www.likertplot.com/>

- Bei kardinalen Daten *zusätzlich* den **Mittelwert**: Der (arithmetische) Mittelwert einer Zahlenreihe  $x_1 \dots x_n$  wird berechnet als  $\frac{1}{n} \sum_{i=1}^n x_i$ . Im obigen Beispiel wäre dies  $31/7 = 4,42$ .

Bei einer solchen Zusammenfassung der Zahlenreihe geht natürlich Information verloren, denn sowohl die Reihe (1, 2, 3, 4, 5) als auch die Reihe (3, 3, 3, 3, 3) ergeben den gleichen Wert für Mittelwert, Median und Modus, nämlich 3. Dabei weichen die Zahlen in der ersten Reihe jedoch viel stärker voneinander und von ihrem Mittelwert ab. Als ein mögliches Maß für diese Abweichung vom Mittelwert kann man die **Standardabweichung**  $\sigma$  einer Zahlenreihe  $x_1 \dots x_n$  angeben, die nach folgender Formel berechnet wird wobei  $\bar{x}$  der arithmetische Mittelwert der Zahlenreihe ist:

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Die Standardabweichung für die erste Zahlenreihe (1, 2, 3, 4, 5) ist demnach  $\sqrt{10/4} = 1.58$ , die für die zweite Zahlenreihe (3, 3, 3, 3, 3) ist 0.0. Dies bringt zum Ausdruck, dass die Unterschiede zwischen den Werten in der ersten Reihe wesentlich größer sind. In statistischen Auswertungen wird die Standardabweichung oft explizit mit angegeben, in grafischen Darstellungen wird sie in Form von sogenannten Fehlerbalken dargestellt (siehe Abbildung 13.4).

### 13.3.6 Statistische Auswertung

Um nun eine wissenschaftlich haltbare Aussage treffen zu können, benötigen wir ein Konzept der induktiven Statistik, die statistische **Signifikanz**. Ein Unterschied zwischen zwei ermittelten Werten gilt als *statistisch signifikant*, wenn die Wahrscheinlichkeit, dass dieser Unterschied durch reinen Zufall zustande gekommen ist, kleiner als ein bestimmter Wert ist. Dieser bestimmte Wert heißt **Signifikanzniveau** und wird in den

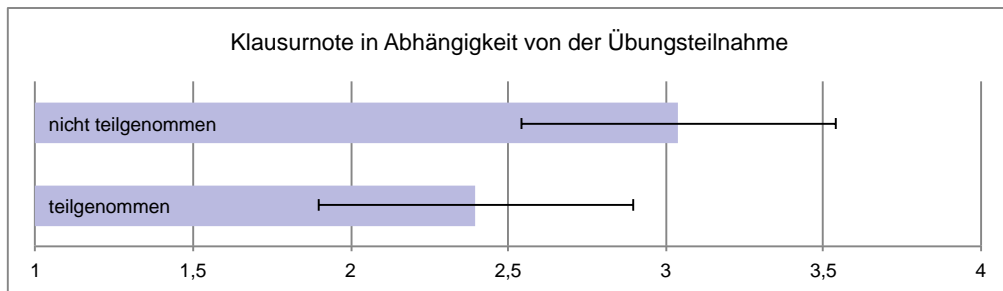


Abbildung 13.4: Fehlerbalken zur Darstellung der Standardabweichung bei der aggregierten Darstellung einer Zahlenreihe durch ihren Mittelwert



mmibuch.de/s/13.4

meisten Fällen mit 5 Prozent ( $p = 0.05$ ) angesetzt. In jedem Fall sollte das verwendete Signifikanzniveau bei einer statistischen Auswertung mit angegeben werden. Für die Berechnung dieser Wahrscheinlichkeit gibt es nun verschiedene etablierte Verfahren, die jeweils in verschiedenen Situationen zum Einsatz kommen. In der einfachsten Situation, in der lediglich zwei Zahlenreihen miteinander verglichen werden sollen, leistet dies der **t-Test**. Wurden beide Zahlenreihen von verschiedenen Probanden erzeugt (**between-groups Design**), dann wird insbesondere der t-Test für unabhängige Stichproben (auch **doppelter t-Test**) verwendet. Wurden die Reihen durch die gleichen Probanden erzeugt (**within-subjects Design**), so verwendet man den **abhängigen t-Test**. Liefert dieser eine Wahrscheinlichkeit  $p < 0.05$ , so können wir davon ausgehen, dass der Unterschied zwischen diesen beiden Zahlenreihen nicht zufällig entstanden ist, sondern ein Effekt der Veränderung der unabhängigen Variablen im Experiment ist (vorausgesetzt, wir haben keine anderen Fehler im experimentellen Design gemacht). Erinnern wir uns an die in Abschnitt 13.3.4 aufgestellte **Hypothese**  $H$  und ihr Gegenteil, die zugehörige **Nullhypothese**  $H_0$ :

$H$ : Übungsteilnehmer erzielen bessere Noten in der Klausur.

$H_0$ : Übungsteilnehmer und Nichtteilnehmer erzielen im Mittelwert die gleichen Noten in der Klausur.

Ergibt der t-Test nun eine Wahrscheinlichkeit  $p < 0.05$ , dass der Unterschied zwischen beiden Versuchsgruppen zufällig entstanden ist, dann widerlegt dies die Nullhypothese  $H_0$  und bestätigt damit die Hypothese  $H$ . Aus dem Unterschied zwischen den beiden Mittelwerten können wir außerdem die Größe des Unterschiedes (die **Effektgröße**) ablesen, die im Datensatz zu Abbildung 13.4 bei etwa 0.7 liegt.

Will man mehrere Messreihen miteinander vergleichen, dann könnte man im ersten Ansatz immer paarweise überprüfen, ob der t-Test einen signifikanten Unterschied zwischen jeweils zwei der Reihen liefert. Dabei würden sich jedoch insgesamt größere Fehlerwahrscheinlichkeiten als das verwendete **Signifikanzniveau** ergeben, weshalb hier

üblicherweise ein anderes Verfahren zum Einsatz kommt. Die **Varianzanalyse** oder **ANOVA** (engl. *Analysis of Variance*) ermittelt, ob es unter  $n > 2$  verschiedenen Zahlenreihen einen signifikanten Unterschied gibt. Auch hier werden verschiedene Varianten, je nach experimentellem Design eingesetzt. Die statistische Auswertung komplexerer experimenteller Designs kann recht kompliziert werden und übersteigt den Umfang eines einführenden Lehrbuchs. Für den tieferen Einstieg sei hier auf eine gute Statistikvorlesung oder die Bücher von **Andy Field**, wie beispielsweise [33] verwiesen.

Die Methoden zur Durchführung kontrollierter Experimente stammen aus der Statistik und werden bereits seit langem in anderen Wissenschaften wie Psychologie, Medizin oder Soziologie verwendet. Die Experimente im Bereich der Mensch-Maschine-Interaktion sind im Vergleich zu den dort durchgeführten Experimenten meist recht klein, insbesondere was den Zeitumfang und die Probandenzahl angeht. Trotzdem ermöglichen sie uns, wissenschaftlich solide Aussagen über die Bedienung von Computern zu machen, und nicht einfach nach Geschmack und ästhetischem Gefühl zu urteilen, was von anderen Fachgebieten innerhalb der Informatik gelegentlich unterstellt wird. Lässt sich statistisch signifikant nachweisen, dass die gleiche Aufgabe vom Benutzer mit *System A* schneller oder mit weniger Fehlern erledigt wird als mit *System B*, dann ist damit die höhere Effizienz von System A wissenschaftlich erwiesen.

### 13.3.7 Feldstudien und Laborstudien

Viele Arten von Interaktion lassen sich gut im Labor studieren: Will man beispielsweise die Bediengeschwindigkeit einer bestimmten Benutzerschnittstelle am Personal Computer messen, so kann man einen solchen Computer in einem Laborraum aufstellen, alle äußeren Störeinflüsse wie Lärm oder andere Quellen der Ablenkung ausschalten, alle Umgebungsparameter wie Helligkeit, Raumtemperatur und Tageszeit konstant halten, und dann vergleichbar und reproduzierbar die Zeit zur Ausführung einer bestimmten Aufgabe messen. Solche **Laborstudien** sind dann aussagekräftig, wenn sich das spätere Bedienumfeld nicht wesentlich von der Laborumgebung unterscheidet, oder wenn der Unterschied sich auf alle getesteten Systeme gleich auswirkt.

Ein Gegenbeispiel: Wir vergleichen im Labor zwei Arten der Texteingabe und befinden, dass *Variante A* schneller ist als *Variante B*. Nun ist aber *Variante B* durch ihre Funktionsweise robuster gegen Unterbrechungen. In diesem Fall hängt es ganz vom späteren Einsatzzweck ab, welche Variante dafür besser geeignet ist: In ungestörten Umgebungen wird dies wie erwartet *Variante A* sein, in Umgebungen mit Ablenkungen, wie beispielsweise im Auto oder in der U-Bahn aber möglicherweise *Variante B*. Eine Möglichkeit besteht nun darin, im Labor die späteren Umgebungsbedingungen so gut wie möglich nachzubilden, indem man beispielsweise eine **Sekundäraufgabe** (engl. **secondary task**) gibt (siehe auch Abschnitt 3.5.3). In unserer Studie könnte das beispielsweise die Reaktion auf regelmäßig eintreffende Ablenkungen sein. In vielen Fällen kann man aber das spätere Umfeld gar nicht genau im Labor nachbilden oder es ist schlichtweg nicht vorhersehbar.

In solchen Fällen bleibt nur die Möglichkeit, die Studie in der tatsächlichen Anwendungsumgebung durchzuführen. Aus dem Sprachgebrauch anderer Wissenschaften wurde hierfür der Begriff **Feldstudie** übernommen. In solchen Feldstudien verzichtet man

auf die wohlkontrollierte Umgebung des Labors mit seinen garantierten Bedingungen, was möglicherweise einen Verlust an **interner Validität** bedeutet, da die Messung abhängiger Variablen oft schwieriger wird und die Reproduzierbarkeit durch die unvorhersehbare Umgebung sinkt. Dem gegenüber steht jedoch ein Gewinn an **externer Validität**, da eine Untersuchung des fraglichen Systems in seiner echten Anwendungsumgebung viel besser auf die spätere echte Anwendung übertragbar ist.

Feldstudien sind fast immer mit wesentlich höherem Aufwand an Zeit, Material und Organisation verbunden und stellen insbesondere auch hohe Anforderungen an die darin verwendeten **Prototypen**. Diese müssen auch außerhalb des Labors funktionieren und einen gewissen Funktionsumfang robust bereitstellen, um eine realistische Bedienung in der echten Nutzungsumgebung überhaupt zu ermöglichen. Weitere Details zu Feldstudien finden Sie in Abschnitt 11.5.

### 13.3.8 Langzeit- und Tagebuch-Studien

Eine weitere Komplexität kommt hinzu, wenn die experimentelle Aufgabe nicht in einem kurzen, eng umrissenen Zeitraum zu erledigen ist, sondern die Verwendung über lange Zeiträume hinweg beobachtet werden soll. Ein Beispiel: Wir haben eine neue Software zur Durchsuchung von Musiksammlungen auf dem Smartphone geschrieben und wollen in einem Experiment herausfinden, ob diese Software dazu führt, dass die Benutzer häufiger, länger oder andere Musik hören. Solche **Langzeitstudien** werden praktisch immer als Feldstudien durchgeführt. Im Gegensatz zu kürzeren Feldstudien kann aber hier der Experimentator nicht mehr während der gesamten Zeit der Studie anwesend sein. Auch dies stellt erweiterte Anforderungen an den verwendeten Prototypen, da sich dieser nun nicht mehr im direkten Einfluss des Experimentators befindet und eine Fehlfunktion damit zu erheblich größerem Schaden (Zeitverlust oder Abbruch der Studie) führt. Stürzt die Software regelmäßig ab, dann kann man von keinem Probanden erwarten, dass er sie langfristig verwendet.

Sofern das untersuchte System sowie sonstige ethische Überlegungen (Privatsphäre, Anonymität) es erlauben, können Daten in einer solchen Studie automatisch vom Prototypen erhoben werden. Unsere Musik-Software könnte beispielsweise in eine Datei schreiben, wann welches Musikstück gehört wird. Dabei beschränkt sich jedoch die erhobene Information auf Zeit und Identität der Stücke und andere wichtige Informationen, wie beispielsweise die Situation, in der die Software verwendet wurde, bleiben unbekannt. Ein umfassenderes Bild lässt sich mittels einer **Tagebuchstudie** gewinnen. Dabei werden die Probanden gebeten, zu bestimmten Zeiten (entweder feste Uhrzeiten oder bei bestimmten Anlässen) relevante Angaben in einem Tagebuch zu vermerken, und zwar in einer bestimmten, vorgegebenen Struktur, die eine spätere systematische Analyse ermöglicht. Die Tagebuchfunktion kann auch in den Prototypen selbst integriert sein, der dann zu bestimmten Zeitpunkten beispielsweise einen kleinen Fragebogen am Bildschirm präsentiert. Ein grundlegendes Problem bei dieser Art von Studien besteht in der niedrigen oder nachlassenden Begeisterung der Versuchspersonen. Jeder Eintrag in das Tagebuch bedeutet einen zusätzlichen Zeitaufwand, der daher minimiert werden sollte. Eine weitergehende Diskussion solcher Methoden findet sich im Abschnitt 11.5 und beispielsweise in dem Buchkapitel von Reis und Gable [95].



### Übungsaufgaben:

1. Bewerten Sie die Webseite Ihrer Universität mittels heuristischer Evaluation! Leiten Sie hierzu eine konkrete Checkliste aus Niensens Heuristiken ab, wählen Sie sich einen sinnvollen Ausschnitt (z.B. eine Fakultät, eine Verwaltung) und analysieren Sie, ggf. in Gruppen zu 4-5 Teilnehmern, diesen Ausschnitt. Falls Sie dies im Rahmen des Übungsbetriebes zur Vorlesung gemacht haben, vergleichen Sie Ihre Ergebnisse mit denen anderer Gruppen. Falls Ihnen erhebliche Probleme auf der Seite aufgefallen sind, erstellen Sie einen Report, in dem die Probleme genau dargestellt und bewertet, sowie Lösungsvorschläge gemacht werden, und senden Sie diesen (mit einer kurzen und höflichen E-Mail) an die Leitung der entsprechenden organisatorischen Einheit (Fakultät, Verwaltung). Dokumentieren Sie deren Reaktion!
2. Vergleichen Sie mithilfe der GOMS und KLM Modelle zwei verschiedene Methoden, eine Datei umzubenennen. Bei der Methode 1 wird die Datei im grafischen Interface angeklickt, ihr Name gelöscht und ein neuer Name eingegeben. Bei der Methode 2 wird in einer (bereits offenen und im richtigen Verzeichnis befindlichen) Kommandozeile das UNIX Kommando `mv name1 name2` verwendet. Unter welchen Bedingungen ist Methode 1 schneller, unter welchen Bedingungen Methode 2?
3. Entwerfen Sie ein kontrolliertes Experiment, das überprüft, mit welcher Methode man am schnellsten zu den medialen Zusatzangeboten dieses Buches gelangt. Vergleichen Sie die Verwendung der gedruckten QR-Codes mit dem Eingeben der ebenfalls abgedruckten URL. Definieren Sie genau die für Sie relevante Zielgruppe und Nutzungssituation und nehmen Sie zur internen und externen Validität Ihrer Studie Stellung.



# 14 Experience Design

Interaktion mit technischen Geräten erzeugt ein mehr oder weniger eindruckliches Erlebnis. Technik wird gelegentlich sogar gezielt zum Erzeugen von Erlebnissen eingesetzt, auch wenn sie dafür gar nicht gedacht war. So gibt es in Italien beispielsweise den Brauch, das Telefon eines vertrauten Menschen nur einmal kurz klingeln zu lassen, um ihm damit zu signalisieren, dass man gerade an ihn denkt. Da die Nummer des Anrufers ja übertragen wird, weiß der Angerufene auch, wer an ihn denkt. Dieser **Squillo** genannte Brauch ist ein kostenloses Nebenprodukt der Technik, und geht damit beispielsweise am Geschäftsmodell der Telefonanbieter komplett vorbei.

Seit einigen Jahren wird immer häufiger der Begriff **User Experience** (auch **UX**) verwendet, um das ganzheitliche Erleben bei der Interaktion zu beschreiben. Der Begriff ist zum regelrechten Buzzword geworden und seine genaue Definition leider oft unklar. In manchen Fällen wird er einfach synonym mit dem Begriff **Usability** verwendet. Es gibt jedoch auch eine wachsende Community, die mit *User Experience* ein recht konkretes Konzept bezeichnet, das über die reine *Usability* weit hinausgeht. In diesem Verständnis ist eine gute *Usability* zwar eine oft notwendige Voraussetzung für eine positive *User Experience*, nicht jedoch hinreichend dafür. Manchmal entsteht sogar trotz ausgeprägt schlechter *Usability* eine insgesamt positive *User Experience*. In seinem kompakten und grundlegenden Buch zu diesem Thema [43] erläutert **Marc Hassenzahl** sein Verständnis davon, wie eine gute *Experience* zu erzeugen ist.

## 14.1 Ziele und Bedürfnisse

Will man gezielt positive Erlebnisse erzeugen, dann kann man das beispielsweise, indem man bestimmte psychologische Grundbedürfnisse betrachtet und dafür sorgt, dass diese bei der Interaktion erfüllt werden. Hassenzahl beschreibt in seiner Arbeit eine Hierarchie von Zielen (siehe Abbildung 14.1). Demnach stehen an unterster Stelle die **motorischen Ziele**, die durch physikalische Aktionen wie das Drücken einer Taste erfüllt werden können. Sie beschreiben das *Wie* der einzelnen Bedienschritte. Die ausgeführten Aktionen entsprechen auch den elementaren Aktionen bei der Ausführung zielgerichteter Handlungen (siehe Abschnitt 5.3.1). Diesen motorischen Zielen übergeordnet sind die Ziele, die sich damit befassen, etwas zu tun (**Handlungsziele**), also die komplexeren Aktionen, die bestimmte Resultate erzielen wollen und durch komplexere Handlungen erreicht werden, wie beispielsweise das Verfassen eines Textes. Sie beschreiben das *Was* unserer Aktionen. Ihnen übergeordnet sind wiederum die Ziele, etwas bestimmtes zu *sein*, beispielsweise ein Schriftsteller (**Motive**). Sie beschreiben die inneren Antriebskräfte oder das *Warum* unserer Aktivitäten. Wir möchten beispielsweise reich, berühmt oder beliebt sein, und aus dieser Motivation begründen sich dann bestimmte Tätig-



**Abbildung 14.1:** Hierarchie von Zielen nach Hassenzahl [43]. Produkte sprechen oft nur die Ziele nahe der Welt an, das Gesamterlebnis wird jedoch durch alle drei Ebenen bestimmt.

keiten, die wir verfolgen, beispielsweise eine gute Ausbildung oder sozialverträgliches Verhalten. Interaktive Produkte werden in der Regel nach den Kriterien der beiden unteren Ebenen dieser Hierarchie entworfen: sie müssen auf einer pragmatischen Ebene bestimmte Aktionen oder Handlungen ermöglichen um bestimmte *Ergebnisse* zu erzielen. Das *Gesamterlebnis* wird jedoch durch die Ziele auf allen drei Ebenen bestimmt.

Will man nun gezielt positive Erlebnisse erzeugen, so ist es wichtig, von unseren inneren Beweggründen und Motiven auszugehen, also vom *Warum* unseres Handelns. Dieses kann beispielsweise durch psychologische Bedürfnisse bestimmt sein. In einem grundlegenden Paper [103] und späteren, darauf aufbauenden Arbeiten sammelt **Kennon M. Sheldon** zehn psychologische Bedürfnisse des Menschen, deren Erfüllung zu einem glücklichen und zufriedenen Leben beiträgt, und ermittelt in drei Studien eine recht universell gültige Rangfolge dieser Bedürfnisse. Demnach sind die wichtigsten psychologischen Bedürfnisse:

- *Autonomie*: das Gefühl, dass man selbst die eigenen Handlungen kontrolliert und nicht durch andere Kräfte oder Druck von außen bestimmt wird,
- *Kompetenz*: das Gefühl, dass man sehr gut und effektiv ist in dem was man tut,
- *Verbundenheit*: das Gefühl, dass man regelmäßigen engen Kontakt hat zu Leuten, denen man etwas bedeutet, und die einem etwas bedeuten.

Weitere Bedürfnisse bestehen nach *Selbstachtung*, *Sinnerfüllung*, *Sicherheit*, *Beliebtheit*, *Einfluss*, *physischem Wohlergehen* und *Geld & Luxus*. Diese psychologischen Bedürfnisse sind nicht so lebenswichtig wie die rein physischen Bedürfnisse nach *Essen*, *Trinken* und *Schlafen*. Sie bauen vielmehr darauf auf: erst ein Mensch, der alles absolut Lebensnotwendige hat, kümmert sich um weitere, psychologische Bedürfnisse. Stillt der Umgang mit einem interaktiven System nun eines der genannten psychologischen Bedürfnisse, so wird das Gesamterlebnis positiv bewertet. Eine positive *Experience* kann also durch die gezielte Erfüllung psychologischer Bedürfnisse bei der Interaktion geschaffen werden.

## 14.2 Beschreibung von User Experience

Um ein Erlebnis, also eine bestimmte **Experience** mit einem interaktiven System gezielt hervorrufen zu können, muss dieses Erlebnis bereits im Entwurfsprozess möglichst genau definiert und durch alle Entwicklungsstadien hindurch konsistent beschrieben sein. Nun gibt es zur Beschreibung von Erlebnissen derzeit leider keinen Formalismus und keine definierte Beschreibungssprache, wie sich das der Informatiker wünscht. Stattdessen lassen sich Erlebnisse aber in Form von Geschichten erfassen, die beim Leser die beabsichtigten Gefühle und Bedürfnisse nachvollziehbar machen. Der Aufbau guter Geschichten und ihre Wirkungsweise wiederum ist nicht Gegenstand der Informatik, sondern wird von anderen Berufsgruppen wie Autoren, Journalisten oder Regisseuren studiert. Von diesen Berufsgruppen können wir lernen, wie sich eine gute Geschichte aufbaut, wie sie einen Spannungsbogen entwickelt, und welche Grundkonstrukte der Erzählung verwendet werden, um gezielt Spannung oder andere Gefühle zu erzeugen. Wenn wir dann ein bestimmtes Erlebnis gezielt mit einem Produkt erzeugen wollen, besteht eine Erfolg versprechende Möglichkeit darin, das Erlebnis in Form einer Geschichte zu erfassen, und in den verschiedenen Entwicklungsstadien des Produkts immer wieder abzugleichen, ob das in der Geschichte vermittelte Erlebnis immer noch durch das System in seinem derzeitigen Entwicklungsstadium erzeugt wird [59]. Die Idee, eine Geschichte zur Definition eines Erlebnisses zu verwenden, hat zudem den Vorteil, dass alle am Prozess Beteiligten (Designer, Psychologen, Ingenieure, Management) die Geschichte gleichermaßen gut verstehen können und kein Fachwissen über eine Beschreibungssprache brauchen.

Diese Vorgehensweise stellt uns vor zwei wesentliche Herausforderungen: Einerseits ist es wichtig, tatsächlich in allen Stadien die Geschichte als Basis zu betrachten, was eher technisch denkenden Ingenieuren in Entwicklungsabteilungen derzeit noch schwer vermittelbar ist. Andererseits müssen die in den Zwischenstufen entstehenden Prototypen so konkret sein, dass sie das Erlebnis erfahrbar machen und damit überhaupt erst ermöglichen, die erfolgreiche Vermittlung des Gesamterlebnisses in Form einer Evaluation zu überprüfen.

## 14.3 Evaluation von User Experience

Im Prozess des **UCD** folgt auf jeden Prototypen eine Evaluation, um den Erfolg des zugrunde liegenden Designs zu bewerten. Ist das Ziel des Designs nun, eine bestimmte **Experience** hervorzurufen, dann muss eine Evaluation diese auch messen können. Das ist schon prinzipiell schwieriger als das objektive Messen von Zeiten oder Fehlerraten, da es hierbei ja um ein Erlebnis im Kopf des Benutzers geht, in den man nun mal nicht hinein schauen kann.

Dieses Erlebnis im Kopf des Benutzers lässt sich nur sehr eingeschränkt von außen, also physiologisch messen. Die Bewertung von Gesichtsausdrücken oder die Messung von Herzfrequenz oder Hautleitwert als Indikatoren für den Erregungszustand sind vergleichsweise grobe Werkzeuge, die zu einer differenzierten Bewertung des Erlebnisses derzeit nicht taugen. Es bleibt also keine andere Möglichkeit, als den Benutzer nach seinem Erlebnis zu befragen. Dies sollte jedoch nicht direkt getan werden (*Haben Sie sich*

*verbunden gefühlt?*), da von den Benutzern kein detailliertes Verständnis psychologischer Bedürfnisse erwartet werden kann, sondern durch eher indirektere Fragen (*Hatten Sie das Gefühl, anderen Menschen näher zu sein?*).

Solche Fragen können beispielsweise in **semi-strukturierten Interviews** (siehe Abschnitt 11.2) gestellt werden. Um sich dem Kern des Erlebnisses immer weiter anzunähern, kann die sogenannte **laddering** Technik eingesetzt werden, bei der die Antworten des Gesprächspartners immer wieder nach dem *Warum* hinterfragt werden, und zwar so lange, bis man beim psychologischen Grundbedürfnis angelangt ist.

Zur Evaluation von *User Experience* gibt es auch einige standardisierte Fragebogen, deren Validität, teilweise auch über verschiedene Kulturkreise hinweg, durch die Autoren sicher gestellt wurde. Generelle Gefühlsreaktionen lassen sich beispielsweise mit der **PANAS** (Positive Affect Negative Affect Scale) Familie von Tests ermitteln. Die ausführliche Variante **PANAS-X** besteht dabei aus 60 Fragen, die kurze **I-PANAS-SF** [114] aus zehn Fragen. Eine Bewertung der hedonischen und pragmatischen Qualitäten eines Systems erlaubt der **AttrakDiff** Test [44], der damit eine Einschätzung liefert, ob ein System zwar nützlich aber langweilig, oder eher verspielt, aber nutzlos ist. Das Gebiet **User Experience** entwickelt sich derzeit noch sehr schnell und zum Druckzeitpunkt dieses Buches fand sich beispielsweise eine wachsende Sammlung von Evaluationsmethoden in einem Blog von Forschern des Gebiets.<sup>1</sup>

## 14.4 User Experience: ein Beispiel

Als Beispiel für den gesamten geschilderten Prozess wollen wir das System **Clique Trip** betrachten. Im zugehörigen Paper [60] werden die einzelnen Schritte des benutzerzentrierten Entwurfs dieses Systems genauer besprochen und auch Details wie die Interviewführung diskutiert. Das Ziel der Autoren von *Clique Trip* war, eine Technologie zu entwickeln, die das Grundbedürfnis der *Verbundenheit* im Automobil anspricht, da das Auto uns derzeit oft von der Umgebung und anderen Menschen abschirmt. Dazu wurden zunächst in Interviews viele verschiedene reale **Erlebnisse** von Interviewpartnern gesammelt, die Verbundenheit im Auto betrafen. Diese Erlebnisse wurden auch mittels des **PANAS-X** Fragebogens bezüglich ihrer Gefühlswirkung analysiert. Die so erhobenen Erlebnisse wurden sodann auf ihre Essenz oder Kerngeschichte reduziert. Vier der Geschichten folgten dabei einem ähnlichen Muster, und so wurde dieses Muster als Basis für eine eigene fiktive Geschichte genommen (Zitat aus Knobel et al. [60]):

Max, Sarah, Marianne, Martin, Monica, and Matthias have known each other for ages. Lately, they don't spend time together as often as before. But one event is always fixed: each year they visit their favorite city as a group – Paris. As usual, they go there in two cars. This year, however, something is different. Max invites all to the trip via Clique Trip, a new app he wants to try out. This app promises to make its users feeling close to each other, even when being in two different cars. All friends are excited to test Clique Trip because they hate the feeling of being separated during the

---

<sup>1</sup><http://www.allaboutux.org>

trip. It is time to depart. Max is driving one car, Sarah the other. Sarah is a very sporty driver (some say reckless) and Max drives very relaxed (some say painstakingly slow). Consequently, they tend to lose each other on the motorway, with Max getting more and more behind. But Clique Trip helps out. It changes the navigation system so that it guides Max (in the rear car) to Sarah (in the leading car). Ah, Sarah takes the scenic route. „Good choice“, Max thinks. He announces „I guess the others plan to visit the nice little cafe in the city centre of Reims. Let me try to catch up“. He does, and when the cars are close to each other, Clique Trip opens a communication channel. They can now talk to each other, as if sitting in one car. „Hey,“ Max yells, „I hope you are not planning to have a first glass of Champagne already? I am driving!“

Diese Geschichte wurde dann in Form eines Storyboards weiter konkretisiert. Dabei steigt der Detailgrad von der Essenz der Geschichte über die Geschichte selbst bis zum Storyboard kontinuierlich an. In den Abbildungen des Storyboards werden beispielsweise schon technische Details vorgeschlagen. Abbildung 14.2 zeigt dabei drei wichtige Szenen aus dem gesamten Ablauf, nämlich die Tourplanung (Links), die Kontaktaufnahme (Mitte) und die Kommunikation zwischen den Fahrzeugen (Rechts).



**Abbildung 14.2:** Bilder aus dem Storyboard zu *Clique Trip* [60], einem System, das *Verbundenheit* zwischen den Teilnehmern einer Gruppe in verschiedenen Autos vermittelt.



Im nächsten Schritt entstand dann ein erster Prototyp als Apps auf einem Smartphone und im Entertainment System eines Autos, der iterativ weiter verfeinert wurde. Der endgültige Prototyp wurde dann mit Probanden im Auto evaluiert und das bei der Benutzung entstehende Erlebnis mittels Fragebogen erhoben. Dabei konnte das Ergebnis mit dem Design-Ziel und die erzeugten Emotionen mit denen der anfangs erhobenen Geschichten verglichen werden. An diesem Beispiel wird klar, dass zur gezielten Gestaltung von Erlebnissen immer noch die grundlegende Methodik des **UCD** eingesetzt werden kann, dass jedoch die in den einzelnen Schritten eingesetzten Werkzeuge teilweise völlig andere sind. Dieses Buchkapitel kann das Thema auf wenigen Seiten natürlich nicht in seiner gesamten Breite erfassen. Da das Gebiet des *Experience Design* aber auch noch so stark in der Entwicklung ist und es noch keinen über Jahrzehnte bewährten Methodenkanon gibt, wäre das auch gar nicht sinnvoll. Aufbauend auf dem hier ver-

mittelten Grundverständnis wird dem Leser also wieder nichts anders übrig bleiben, als sich in die aktuellen Entwicklungen nach Erscheinen dieses Buches selbst einzuarbeiten. Die Gruppe um Marc Hassenzahl<sup>2</sup> wird dafür vermutlich auf absehbare Zeit ein guter Startpunkt bleiben.



### Übungsaufgaben:

1. Angenommen, Sie haben eine Stunde Freizeit und einen Computer mit Internetzugang. Wie würden Sie diese Stunde verbringen, und warum? Analysieren Sie, welche psychologischen Bedürfnisse durch den von ihnen dann genutzten Dienst (z.B. Web, E-Mail, soziale Netze, Online-Spiel) angesprochen werden, auf welchem Wege dies passiert, und was davon aktiv durch die Entwickler gestaltet ist, bzw. was vermutlich Zufall bei der Entstehung des Dienstes war.
2. Vergleichen Sie die hedonistischen und pragmatischen Qualitäten von zwei Webseiten Ihrer Wahl. Verwenden Sie dazu den Attrakdiff-Fragebogen<sup>a</sup>. Diskutieren Sie die Unterschiede der beiden Webseiten.

---

<sup>a</sup>Informieren Sie sich dazu im Internet, z.B. unter <http://attrakdiff.de/>

---

<sup>2</sup><http://www.marc-hassenzahl.de>