

# „Fünfe gerade sein lassen“: Listenkontexteffekte beim Kategorisieren

*Ulf-Dietrich Reips<sup>1</sup>, Vinzenz Morger<sup>1</sup>, Beat Meier<sup>2</sup>*

## **Kurzbeschreibung**

In diesem Kapitel zeigen wir, dass einfachste Kategorisierungsentscheidungen wie „Ist 85 eine ungerade Zahl?“ in bestimmten Kontexten mehr Zeit benötigen als in anderen Kontexten und mit grossen Fehlerhäufigkeiten behaftet sein können. Es kann also selbst bei exakt definierten Begriffen zu vorübergehenden Modifikationen der Begriffsstruktur kommen. Unser Web-Experiment zu Listenkontexteffekten zeigt ausserdem, dass es möglich ist, mit Hilfe einer der unzuverlässigsten Techniken des Web-Experimentierens (clientseitiges Javascript) robuste Reaktionszeiteffekte zu messen. Die berichtete Replikation des Listenkontexteffektes im Rahmen einer web-technologiebasierten Untersuchung unterstreicht die externale Validität eines Effektes, der bisher nur unter streng kontrollierten Laborbedingungen aufgezeigt wurde. Das Experiment eignet sich als Ausgangspunkt für relativ leicht durchführbare Extensionen.

**Stichwörter** Listenkontext, Web-Experiment, Kategorisierung, Begriffsbildung

## **1. Einleitung**

Ein vielfach nachgewiesenes und immer wieder repliziertes Phänomen der psychologischen Begriffsforschung besteht darin, dass die Zugehörigkeit von verschiedenen Exemplaren zur übergeordneten Kategorie als abgestuft wahrgenommen wird. Eine Meise und ein Sperling werden beispielsweise als repräsentativere oder typischere Instanzen der Kategorie Vogel betrachtet als etwa ein Huhn oder ein Strauss. Diese begriffliche Abstufung zeigt sich in verschiedenen kognitiven Aufgaben wie etwa beim Typikalitäts-Rating (z.B. Eckes, 1985), bei der Exemplar-Produktion (z.B. Mannhaupt, 1983), oder der Begriffs-Verifikation (Flammer & Morger, 1985).

---

<sup>1</sup> Universität Zürich

<sup>2</sup> University of British Columbia, Vancouver; Beat Meier verbringt zur Zeit einen Forschungsaufenthalt an der University of British Columbia in Kanada, welcher vom Schweizerischen Nationalfonds (Stipendien-Nr. 8210-056614) unterstützt wird.

Wir danken Thomas Bamert und Andrea Frick für ihre Mithilfe beim Erstellen und Auswerten des Experiments und Christoph Neuhaus für das Bearbeiten des Javascript-Codes zur Messung der Antwortzeiten.

Bei semantischen Kategorien (Vögel, Früchte, Möbel, etc.) erscheint uns diese Abgestuftheit des begrifflichen Wissens intuitiv plausibel. Hingegen ist es einigermaßen verblüffend, dass sich dasselbe Phänomen ebenfalls bei exakt definierten Begriffen, etwa den geraden Zahlen oder den ungeraden Zahlen, zeigt (Armstrong, Gleitman & Gleitman, 1983; Bromme, 1990). So wird beispielsweise die Zahl 77 als typischeres Exemplar der Kategorie „ungerade Zahl“ eingeschätzt als die Zahl 65 und auch die Entscheidung, dass es sich hierbei um ungerade Zahlen handelt, kann bei 77 bedeutend schneller getroffen werden.

Um dieses Phänomen der abgestuften Typikalität zu erklären wird in der Regel von Begriffen als stabilen und statischen Wissensrepräsentationen ausgegangen. Besonders prominent ist dabei die Prototypentheorie von Rosch (1973, 1978). Ein Prototyp ist als summarische Repräsentation eines Begriffs zu verstehen, welche die Information der zentralen Merkmale desselben integriert. Die Abgestuftheit der Begriffszugehörigkeit kommt nun durch die unterschiedliche Ähnlichkeit konkreter Exemplare zu diesem abstrakten Prototyp (als der abstrakten Vergleichsnorm für den kritischen Begriff) zustande.

Wenn Begriffe als Prototypen repräsentiert sind, so bedeutet dies, dass die Kategorisierung von Exemplaren vom Kontext einer Kategorisierungssituation oder auch von kürzlich gemachten Erfahrungen mit der Kategorie nicht beeinflusst wird. Dieser Annahme widerspricht aber der Listenkontexteffekt, der darin besteht, dass die begriffliche Zuordnung eines Begriffsexemplars durch die Liste der zuvor bearbeiteten Exemplare beeinflusst wird. Nach einer Liste mit ähnlichen Exemplaren kommt es zu einer Erleichterung, nach einer Liste mit unähnlichen Exemplaren hingegen zu einer Interferenz (vgl. Morger, 2000). Diese Beeinflussung von Kategorisierungsleistungen durch die Liste der vorausgehend bearbeiteten Exemplare wurde nicht nur bei semantischen Kategorien festgestellt (d.h., *Habicht* wird nach einer Liste mit *Adler* und *Falke* schneller als Voglexemplar verifiziert als nach einer Liste mit *Ente* und *Gans*), sondern ebenfalls bei exakt definierten Begriffen wie „gerade Zahl“ und „ungerade Zahl“ (Morger, 1993, 2000). Ein besonders überraschender Effekt trat dabei bei den Experimenten mit der Zielkategorie „ungerade Zahl“ auf. Im Experiment von Morger (1993) wurde die Zahl 85 in einem unähnlichen Listenkontext (nach Zahlen wie 31, 53, 77) sehr häufig als ungerade Zahl zurückgewiesen (von 34 % der Versuchspersonen), während in der Bedingung mit dem ähnlichen Listenkontext (nach Zahlen wie 83, 65, 25) ein solcher Kategorisierungsfehler bei der Zahl 85 kaum auftrat (3 % der Versuchspersonen). Dieser Befund wurde von Morger (2000) in vier Experimenten mit annähernd übereinstimmenden Ergebnissen repliziert.

Der Listenkontexteffekt zeigt klar auf, dass das Kategorisieren nicht ausschliesslich als analytischer Prozess im Sinne eines abstrahierenden Verallgemeinerns verstanden werden kann. Vielmehr übt der Rückgriff auf kurzfristige episodische Wissensressourcen einen bedeutsamen Einfluss auf begriffliche Zuordnungen aus. Damit steht der Listenkontext zu jenen Theorien im Widerspruch, welche Begriffe als stabile, vom Kontext unbeeinflusste Repräsentationen auffassen (z.B. als Prototypen oder als Merkmalslisten). Hingegen ist er mit den Exemplartheorien (z.B. Brooks, 1987), aber auch mit dem Modell von Begriffen als momentane Konstruktionen von Barsalou (1989) vereinbar. Eine

ausführliche Darstellung der verschiedenen theoretischen Positionen der Begriffspsychologie findet sich bei Eckes (1996) sowie bei Morger (2000, Kapitel 2).

## 2. Fragestellung und Hypothesen

Es gibt mehrere Gründe, um das Experiment zum Nachweis von Listenkontexteffekten bei der Kategorie „ungerade Zahl“ als web-technologiebasiertes Experiment zu replizieren.

Erstens handelt es sich beim untersuchten Listenkontexteffekt um einen *überraschenden*, aber *robusten* Effekt aus einem zentralen Forschungsbereich. Kategorisierung und Begriffsbildung sind für das Verständnis vieler psychologischer Vorgänge von Bedeutung und werden deshalb nicht nur in der Kognitionspsychologie, sondern auch in der Sozial- und der Entwicklungspsychologie intensiv beforscht. Deshalb soll untersucht werden, ob sich der Listenkontexteffekt auch ausserhalb einer streng kontrollierten Laborsituation nachweisen lässt. Damit würde ein Beitrag geleistet zur situationsbezogenen Generalisierbarkeit und externen Validierung des Effektes (vgl. Reips, 2000). Da sich die Versuchssteuerung im Web von jener der ursprünglichen Experimente unterscheidet, wurde das Experiment ausserdem „duozentrisch“ lokal kontrolliert durchgeführt (vgl. Reips, 1997).

Zweitens trägt das Web-Experiment bei zur aktuell innerhalb der Online-Forschung stark diskutierten Frage der Praktikabilität von Reaktionszeitmessungen via Internet (Eichstaedt, 2000; Janetzko, 1999; Janetzko, Hildebrandt & Meyer, 2001). Der Listenkontexteffekt beeinflusst nicht nur die Fehlerhäufigkeit beim Kategorisieren, sondern ebenfalls die für die Verifikation benötigte Antwortzeit. Es soll deshalb geprüft werden, ob sich der Listenkontexteffekt im Web-Experiment – trotz technischer Einschränkungen – auch bei der abhängigen Variable Antwortzeit feststellen lässt.

Die zu prüfenden Hypothesen lauten wie folgt:

- (1) Die ungerade Zahl 85 wird in der Bedingung mit dem unähnlichen Listenkontext (d.h. nach der vorausgehenden Verifikation von Zahlen wie 31, 53, 77) häufiger als ungerade Zahl zurückgewiesen als in der Bedingung mit dem ähnlichen Listenkontext (nach Zahlen wie 83, 65, 25).
- (2) Für die Verifikation der Zahl 85 wird in der Bedingung mit dem unähnlichen Listenkontext mehr Zeit benötigt als in der Bedingung mit dem ähnlichen Listenkontext.
- (3) Die unter (1) und (2) postulierten Hypothesen lassen sich im Web-Experiment sowohl mit über das WWW rekrutierten Versuchspersonen wie mit einer lokal angeworbenen Stichprobe im traditionellen Versuchssetting bestätigen.

*Tabelle X.1:* Listenzusammensetzung und festgelegte Positionsanordnung der Items innerhalb der beiden Versionen.

Version	Position innerhalb der Liste												
	1	2	3	4	5	6	7	8	9	10	11	12	13
A	31	24	11	53	60	12	77	64	39	57	20	37	85
B	83	24	27	65	60	12	25	64	81	45	20	85	37

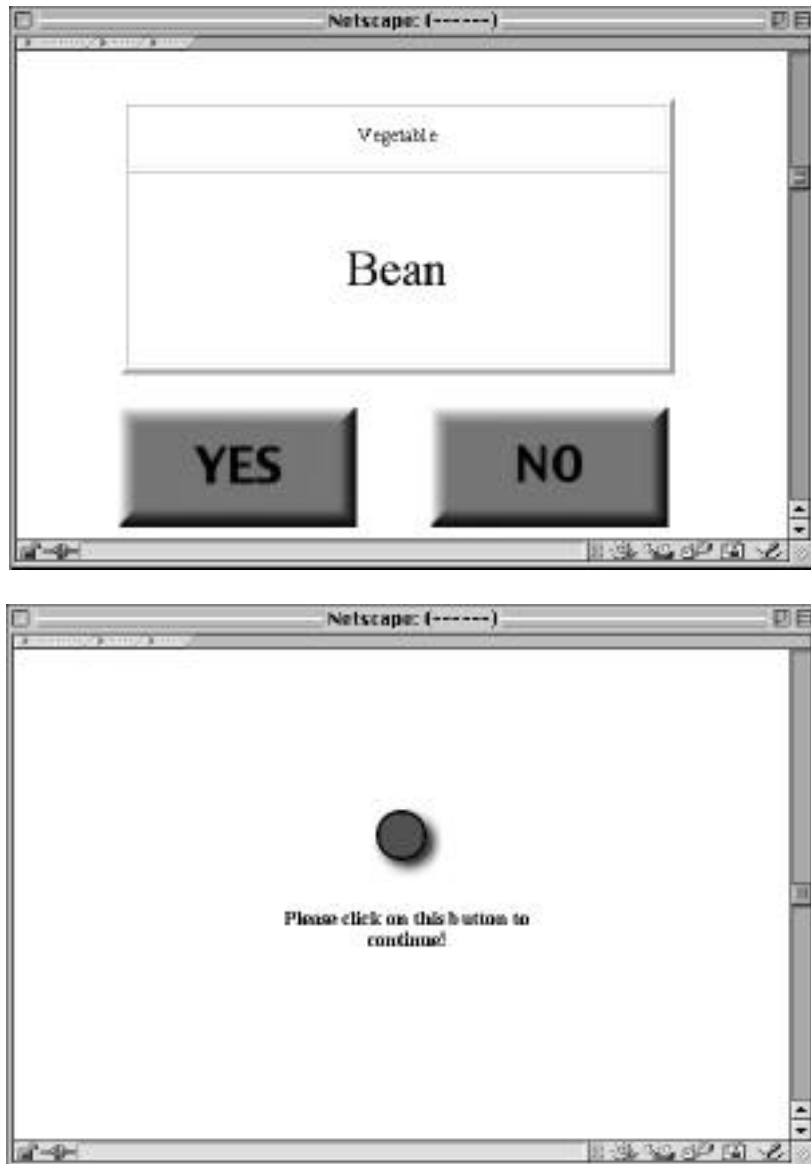
### 3. Methode

#### 3.1 Versuchsmaterial

Zwei Sets mit zweistelligen ungeraden Zahlen wurden zusammengestellt. Dabei wurde darauf geachtet, dass sich die Zahlen innerhalb jedes Sets ähnlich sind, den Zahlen des anderen Sets aber unähnlich. Die Zahlen des ersten Sets setzen sich durchwegs aus zwei ungeraden Ziffern zusammen (31, 11, 53, 77, 39, 57, 37), jene des zweiten Sets hingegen enthalten in der Zehnerstelle eine gerade Ziffer (83, 27, 65, 25, 81, 45, 85). Bei der Auswahl wurde zusätzlich darauf geachtet, Zahlen zu verwenden, welche in der Einer- oder in der Zehnerstelle übereinstimmende Ziffern mit der kritischen Testzahl ihres Sets (37, resp. 85) aufweisen. Mit diesen beiden Sets sowie fünf geraden zweistelligen Zahlen (Distraktoren) wurden zwei Versionslisten für den Verifikationstest zusammengestellt. Die genaue Zusammensetzung und die festgelegte Itemanordnung innerhalb der beiden Versionen ist aus Tabelle X.1 zu ersehen.

Zur Umsetzung der *Warm-up*-Technik zur Abbruchreduktion (Reips, 1999, 2000; s. Erläuterung in Kapitel 4 in diesem Band) und zum Einüben der Versuchsprozedur wurden drei Blöcke mit anderen Kategorien erstellt. Dabei handelt es sich um die Kategorien „Gemüse“ (5 *trials*), „Blume“ (15 *trials*) und „weiblich“ (11 *trials*).

Jeder *trial* bestand aus einer Webseite mit der Zielkategorie und einem zu verifizierenden Item, sowie einer zweiten Webseite zum Rejustieren des Mauspeils. Die Entscheidungs-Webseite zeigte jeweils zentriert oben die Kategoriebezeichnung und direkt darunter den zu kategorisierenden Begriff. Darunter wurden nebeneinander zwei Antwortfelder („Yes“ und „No“) präsentiert, mit denen die Versuchspersonen ihre Kategorisierungsentscheidungen fällten. Nach erfolgter Antwort wurde dann jeweils die Webseite zum Rejustieren der Maustaste aufgerufen. Auf dieser befindet sich ein zentrierter grüner Punkt, der angeklickt werden musste, bevor zum nächsten *trial* weitergeschaltet wurde. Abbildung X.1 zeigt die beiden Webseiten eines Versuchstrials.



*Abbildung X.1: Illustrationsbeispiel der beiden Webseiten eines Versuchstrials (aus der Übungsphase)*

### 3.2 Versuchsplan

Der Untersuchung liegt ein dreifaktorieller Versuchsplan (2 x 2 x 2) mit den drei Faktoren *Zahl* (37, 85), *Listenversion* (Version A, Version B) und *Durchführungsort* (Web-Labor für Experimentelle Psychologie, Experimental-Labor in Vancouver) zugrunde. Beim Faktor *Zahl* handelt es sich um einen wiederholten Messfaktor, während *Listenversion* und *Durchführungsort* zwischen den Versuchspersonen

variiert wurden. Es wurden zwei abhängige Variablen gemessen, nämlich die Korrektheit der Antwort, beziehungsweise die Anzahl der Fehl kategorisierungen, sowie die benötigte Antwortzeit. Die Anordnung der Antwortfelder auf dem Bildschirm („Yes“ und „No“ links oder rechts) wurde ausbalanciert.

### 3.3 Versuchsablauf

Da zur Teilnahme an diesem Experiment ein Javascript-fähiger Web-Browser benötigt wird, wurde für die Version der Web-Versuchspersonen auf der ersten Seite eine *Javascript-Weiche*<sup>3</sup> eingebaut. Bei Vorhandensein dieser technischen Voraussetzung wurde die Startseite des Web-Experiments angezeigt. Alle anderen Versuchspersonen wurden auf eine Seite mit Informationen zum Beschaffen eines Javascript-fähigen Web-Browsers beziehungsweise zum Einschalten von Javascript weitergeschaltet.

Für die Versuchspersonen in Vancouver, bei denen die Javascript-Kompatibilität des Web-Browsers durch den Versuchsleiter voreingestellt war, begann das Experiment gleich mit einer Startseite, welche persönliche Angaben über Computermaus- und Web-Erfahrungen sowie Vorerfahrungen mit psychologischen Untersuchungen erfragte.

Alle Versuchspersonen wurden dann über die Rahmenbedingungen der Studie informiert und um die Angabe von Teilnahmewiederholung, Alter, Geschlecht und E-Mail-Adresse oder Telefonnummer gebeten. Eine finanzielle Vergütung der Teilnahme fand nicht statt, weshalb bei den Web-Versuchspersonen zwar mit einer höheren Abbruchquote zu rechnen ist (Frick, Bächtiger & Reips, in press; vgl. auch O'Neil & Penrod, 2000), was sich jedoch nicht nachteilig auf die Replikation eines robusten Effektes auswirken sollte (Schwarz & Reips, in press). Danach folgte eine Webseite, welche die Teilnehmenden über die Dauer sowie den ausschliesslich wissenschaftlichen Zweck des Experimentes aufklärte. In diesem Zusammenhang wurde noch gefragt, ob sie bereits früher am Experiment teilgenommen hätten. Die Angabe bezüglich einer eventuellen früheren Teilnahme kann später zum Ausschluss von Mehrfachteilnahmen verwendet werden. Sie wird aber auch von Personen genutzt, die nicht ernsthaft teilnehmen möchten (z.B. nur aus Neugier oder aus Forschungsinteresse).

Anschliessend wurden die Versuchspersonen CGI-gesteuert zufällig einer von zwei Präsentationsbedingungen (Position der Antwortfelder: Ja-Nein, Nein-Ja) zugeordnet; die zufällige Zuordnung könnte zwecks Vereinfachung generell auch clientseitig von Javascripts übernommen werden (was aber Nachteile hat, vgl. Schwarz & Reips, in press; Janetzko, 1999).

Nun folgte die eigentliche experimentelle Aufgabe des Kategorisierens. Zum Zweck der Umsetzung der Warm-up-Technik, der Hohe-Hürde-Technik (der Einsatz

---

<sup>3</sup> Solch eine Weiche kann beispielsweise darin bestehen, dass alle Versuchspersonen zunächst auf eine Webseite „X.html“ kommen. Im HTML Code des Kopfteils dieser Webseite steht das folgende Javascript: `<script language="javascript"><!-- window.location="mitjavascriptstart.html" // --></script>`. Wer Javascript hat, dem erscheint dadurch die Webseite „mitjavascriptstart.html“. Alle anderen verbleiben auf X.html

dieser Techniken resultiert in reduzierter Abbruchrate im eigentlichen Web-Experiment, Erläuterung s. Reips, Kapitel 4 in diesem Band) und zum Einüben der Versuchsprozedur wurden der eigentlichen Experimentalphase drei Übungsblöcke mit je einer anderen Kategorie vorgeschaltet. Bei den Kategorien dieser Blöcke handelte es sich in festgelegter Reihenfolge um Gemüse (5 *trials*), Blumen (15 *trials*) und „weiblich“ (11 *trials*). Jeder Übungsblöcke wurde durch eine Instruktionsseite eingeleitet, die im Falle der Kategorie Gemüse folgenden Wortlaut hatte:

### **Instruction 1**

In this experiment your task will be to categorize words. After having read these instructions, you will trigger a few demonstration trials.

First, a box will appear in the center of the webpage, containing a word pair with the words located one above the other. The top word represents the category name. The bottom word is the item you are required to categorize.

If the bottom word is a correct member of the category (top word), you should respond YES by pressing the **YES button** as quickly as possible with the mouse.

If the bottom word is not a member of the category (top word), you should respond NO by pressing the **NO button** as quickly as possible with the mouse.

After your response, the next webpage with a new word pair will appear.

Let's now do that with a few demonstration trials. The critical category in these trials will be **VEGETABLE**.

**To begin with the demonstration trials please use your mouse to click on the following button:**

**[LET'S GO]**

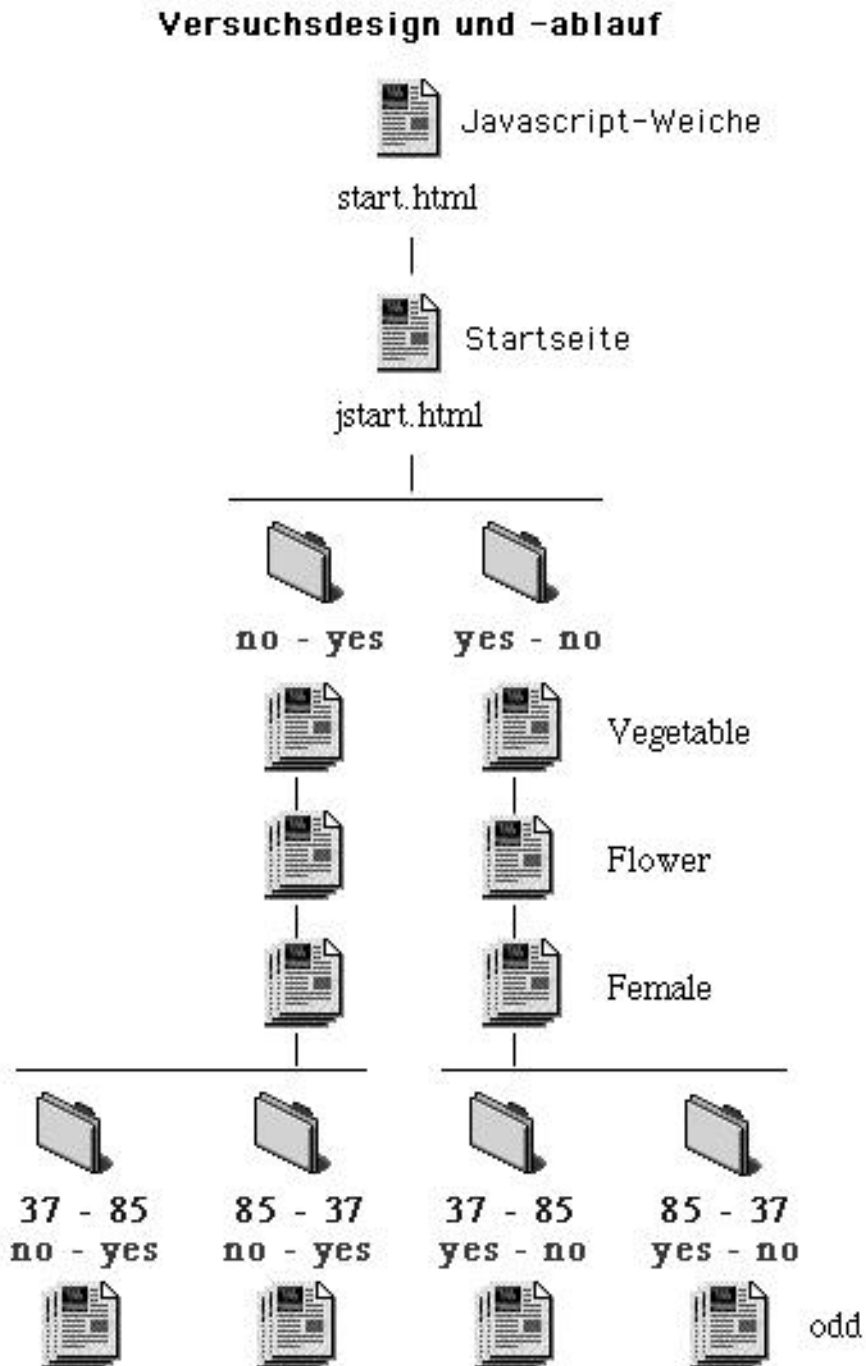
Nach Abschluss der Übungsblöcke fand eine CGI-gesteuerte zufällige Zuordnung zu einer der zwei Listenbedingungen (A oder B) der eigentlichen Experimentalphase statt. Im Anschluss an die Instruktionsseite wurde den Versuchspersonen eine Sequenz von 13 *trials* mit je einer zweistelligen Zahl präsentiert. Bei jeder Zahl sollten sie so schnell wie möglich per Mausklick auf das entsprechende Antwortfeld ihre Entscheidung darüber abgeben, ob die Zahl ungerade ist. Dabei wurde die Antwortausprägung und die benötigte Antwortzeit per Javascript und serverseitig registriert. Nach erfolgter Antwort erschien jeweils die Webseite zur Rejustierung des Mausepfeils, bevor der nächste *trial* präsentiert wurde.

Am Ende des Versuchs folgte eine Dank- und Abschiedsseite, welche Literaturverweise und einen Link zum Web-Labor für Experimentelle Psychologie enthielt.

Das Web-Experiment wurde am 25. Juli 2000 in die englischsprachige Version des Web-Labors für Experimentelle Psychologie (<http://www.genpsy.unizh.ch/Ulf/Lab/WebExpPsyLab.html>) aufgenommen und konnte dort bis zum 11. Oktober durchgeführt werden.

In Abbildung X.2 ist der Ablauf des Web-Experiments schematisch so dargestellt, dass die verwendeten HTML-Dateien (Webseiten), ihre Speicherung in

den den Experimentalbedingungen entsprechenden Ordnern und die Abfolge ihrer Darstellung deutlich werden.



*Abbildung X.2: Schematische Darstellung der Struktur und des Ablaufs des Web-Experiments.*



### 3.4 Versuchspersonen

Im gesamten Zeitraum der Untersuchung wurden insgesamt 1072 Versuchsteilnahmen<sup>4</sup> im Sinne von Aufrufen der ersten Webseite verzeichnet. Die erste Instruktionssseite wurde noch 726 mal aufgerufen, von 602 IP-Adressen aus. Alle Fälle, in denen die Teilnahme von einer zuvor bereits registrierten IP-Adresse aus stattfand, wurden ausgeschlossen. Insgesamt 372 Teilnehmende sahen die Instruktion zu Beginn des eigentlichen Experimentalblocks, alle anderen hatten die Untersuchung noch in der *Warm-up*-Phase abgebrochen. Während des Experimentalblockes gab es nur noch wenige Versuchsabbrüche (s. Abbildung 4.1 in Kapitel 4).

Für die Auswertung wurden jene Datensätze ausgeschlossen, in denen eine frühere Teilnahme angegeben worden war. Ebenfalls ausgeschlossen wurden die Datensätze von solchen Versuchsteilnahmen, bei denen die IP-Adresse schwankte. Bei dieser sogenannten dynamischen IP-Adressierung ist es zwar durch die Netzwerkzugehörigkeit, die zeitliche Nähe und die logische Abfolge der Logdatei-Einträge mit einer gewissen Wahrscheinlichkeit möglich festzustellen, dass es sich um ein und dieselbe Versuchsperson handelt, aber es ist sicherer, auf eine Auswertung dieser Einträge zu verzichten (vgl. Kapitel 4).

Nach Anwendung dieser Ausschlusskriterien blieben die Datensätze von 299 Versuchspersonen übrig. Nun wurden aber noch all jene Datensätze nicht in die Auswertung einbezogen, bei denen

1. fünf oder mehr Zahlen im Experimentalblock falsch kategorisiert worden waren (3 Versuchspersonen; davon 2 mit 12 resp. 13 Fehlern, was auf ein Missverständnis bei der Zuordnung der Antwortfelder schliessen lässt);
2. fünf oder mehr Antworten beim Experimentalblock im Datensatz fehlten (2 Versuchspersonen);
3. einer oder beide Einträge bei den kritischen Zahlen 37 und 85 fehlten (18 Versuchspersonen);
4. im Experimentalblock mehrere extrem lange Antwortzeiten gemessen wurden (2 Versuchspersonen mit 6, resp. 8 Antworten von über 5000 ms).

Nach diesem Vorgehen blieben die Datensätze von 274 Versuchspersonen für die Auswertung übrig. Dabei handelte es sich bei 233 um Versuchspersonen aus dem Web-Labor und bei 41 um Versuchspersonen aus dem Experimental-Labor der University of British Columbia in Vancouver (Kanada). Die Gruppe aus dem Experimental-Labor in Vancouver setzt sich aus Studierenden dieser Universität (vorwiegend im Fach Psychologie) zusammen, die durch einen Aushang angeworben wurden. Sie nahmen grösstenteils freiwillig im Anschluss an ein anderes Experiment teil, mit welchem sie eine Teilanforderung einer Lehrveranstaltung ableisteten. Auch diese Gruppe absolvierte das Experiment über das Internet, d.h. via

---

<sup>4</sup> Wir sprechen hier von Versuchsteilnahmen, da nicht mit Sicherheit festgestellt werden kann, ob es sich in jedem Fall um eine einzelne Person handelt, die nur einmal teilnimmt. Es gibt Hinweise darauf, dass wiederholte Teilnahmen sehr selten sind und ein Einfluss auf die Datenqualität durch geeignete Massnahmen praktisch ausgeschlossen werden kann (Reips, 1997, 2000).

Web-Labor für Experimentelle Psychologie an der Universität Zürich. Der einzige Unterschied bestand darin, dass ihre Teilnahme unter kontrollierten Bedingungen im Experimental-Labor, z.B. mit einem anwesenden Versuchsleiter und unter geringer Variation von Hardware und Software stattfand.

#### 4. Auswertung

Aus der Auswertung der Reaktionszeiten wurden in Analogie zu den Experimenten in den Originalarbeiten alle Werte über 3000 ms ausgeschlossen, um Verzerrungen durch nicht aufgabenbezogene Tätigkeiten der Versuchspersonen zu vermeiden (eine Einbeziehung dieser Werte würde sich laut unserer Analysen zugunsten des untersuchten Effekts auswirken). Für die Analyse der Antwortgenauigkeit wurden alle Messpunkte berücksichtigt.

In Tabelle X.2 sind die proportionalen Anteile korrekter Antworten bei den kritischen Zahlen 37 und 85 als Funktion der beiden Listenversionen und des Durchführungsortes aufgelistet. Es ist ersichtlich, dass die Zahl 85 in der Listenversion A eine höhere Fehlerrate aufweist als in der Listenversion B. Dieser Befund konnte mittels Chi-Quadrat-Test statistisch abgesichert werden und zwar sowohl bei der Gruppe aus dem Web-Labor [ $\chi^2(1, N = 233) = 7.23, p = .007$ ], als auch – trotz relativ geringer Stichprobengrösse – bei der Gruppe aus dem vancouverischen Labor [ $\chi^2(1, N = 41) = 4.14, p = .04$ ]. Dies liegt daran, dass die Effektstärke bei der Gruppe aus dem Labor in Vancouver mit  $h = .85$  gross war, während sie bei der Gruppe aus dem Web-Labor mit  $h = .335$  noch knapp einem mittelstarken Effekt entspricht (vgl. Cohen, 1977).

*Tabelle X.2:* Proportionaler Anteil korrekter Antworten und Anzahl Versuchspersonen als Funktion der Faktoren Zahl, Listenversion und Durchführungsort.

Listenversion	Zahl	Web-Labor		Vancouver-Labor	
		<i>Anteil korrekte Antworten</i>	<i>n</i>	<i>Anteil korrekte Antworten</i>	<i>n</i>
A	37	.98	101	1.00	18
	85	.87	101	.83	18
B	85	.96	132	1.00	23
	37	1.00	132	.96	23

Somit wird der in Hypothese 1 postulierte Listenkontexteffekt bei der Antwortgenauigkeit bestätigt. Nach einer Exemplarliste mit unähnlichen ungeraden Zahlen stellten sich bei der Verifikation der Zahl 85 bedeutend mehr

Kategorisierungsfehler ein als nach einer Exemplarliste mit ähnlichen Zahlen. Bei der Zahl 37 war der Fehleranteil insgesamt sehr gering und wurde nicht von der Listenversion beeinflusst.

Tabelle X.3 zeigt die Mittelwerte der Antwortzeiten für die Zahlen 37 und 85, aufgeschlüsselt nach der Listenversion und dem Durchführungsort. Mit den Antwortzeiten wurde eine vierfaktorielle Varianzanalyse mit den Faktoren Durchführungsort, Listenversion, dem Messwiederholungsfaktor Zahl sowie dem Kontrollfaktor Anordnung der Antwortfelder durchgeführt. Für den Faktor *Zahl* ergab sich ein signifikanter Haupteffekt [ $F(1,233) = 28.96, MS = 2.62, p < .001$ ]; alle übrigen Haupteffekte waren statistisch nicht signifikant. Ausserdem ergab sich eine statistisch signifikante Interaktion *Zahl x Listenversion* [ $F(1,233) = 5.88, MS = 0.53, p < .05$ ], welche darauf zurückzuführen ist, dass für die Verifikation der Zahl 85 in der Listenversion A mehr Zeit benötigt wurde als in der Listenversion B [ $t(239) = 2.55, p = .01$ ]; für die Zahl 37 ergab sich kein analoger Unterschied zwischen den Listenversionen [ $t(239) = 0.52, n.s.$ ]. Ausserdem wurde die Effektstärke für den Unterschied der Antwortzeiten bei der kritischen Zahl 85 in den beiden Listenversionen berechnet. Nach Cohen's Konvention (Cohen, 1977) kann der Effekt in der Stichprobe aus dem Experimental-Labor mit  $d = .50$  als mittelgross, in der Stichprobe aus dem Web-Labor mit  $d = .31$  als gering bis mittelgross eingestuft werden.

Tabelle X.3: Mittlere Antwortzeiten in ms, Standardabweichungen und Anzahl der Versuchspersonen als Funktion der Faktoren Zahl, Listenversion und Durchführungsort.

Listenversion	Zahl	Web-Labor			Experimental-Labor		
		<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
A	37	1091	418	85	934	376	15
	85	1312	531	85	1328	469	15
B	85	1157	481	121	1126	343	20
	37	1048	397	121	991	245	20

Mit diesem Befund wird der in Hypothese 2 postulierte Listenkontexteffekt auf die Antwortzeiten beim Kategorisieren klar bestätigt. Die kurzfristigen Erfahrungen mit unterschiedlichen Exemplaren der Kategorie „ungerade Zahl“ in den beiden Listenversionen beeinflussen den für die Verifikation benötigten Zeitaufwand. Die Zahl 85 kann nach einer Liste mit ähnlichen Zahlen schneller verifiziert werden als nach einer Liste von unähnlichen Zahlen. Die Verifikation der Zahl 37 wird hingegen von der Listenversion nicht beeinflusst.

## 5. Diskussion

Mit diesem Web-Experiment ist es gelungen, einen Befund aus der Laborforschung zu replizieren, der sich bereits mehrmals unter verschiedenen, auch feldartigen Bedingungen gezeigt hatte (Morger, 1993, 2000): Entscheidungen über die Zugehörigkeit vorgegebener Exemplare zu einer Kategorie können vom aktuellen Kontext in Form vorangegangener Erfahrungen mit spezifischen Exemplaren beeinflusst werden. Bei der Kategorie „ungerade Zahl“ zeigt sich dieser sogenannte Listenkontexteffekt nicht nur bei der benötigten Antwortzeit, sondern er wirkt sich zusätzlich in erheblichem Masse auf die Fehlerquote aus. Da es sich bei der Kategorie ungerade Zahl – im Gegensatz zu den semantischen Kategorien – um einen exakt oder scharf definierten Begriff handelt, was bedeutet, dass für eine Zuordnung klare Merkmale oder Regeln existieren, kann dieser Befund als eher unerwartet eingestuft werden.

Wie lässt sich dieser Listenkontexteffekt theoretisch erklären? Vorerst scheiden jene Theorien als Erklärungsansätze aus, die den Kategorisierungsprozess als regelgeleitete Zuordnung auf der Grundlage stabiler und kontextunabhängiger Repräsentationen des begrifflichen Wissens konzipieren. Nachfolgend sollen zwei Modelle diskutiert werden, welche die Wirkung von kurzfristigen Exemplarerfahrungen erklären können.

Nach Barsalou (1989) werden Begriffe momentan oder ad hoc im Arbeitsgedächtnis konstruiert und sie enthalten sowohl kontextunabhängige (stabile), als auch kontextabhängige (variable) Information. Besondere Merkmale der vorausgehend bearbeiteten Exemplare einer Kategorie werden also in eine aktuelle Begriffs-Konstruktion integriert, wodurch diese Konstruktion (als der Vergleichsnorm für die Kategorie) sich dem ähnlichen Testexemplar annähert. Damit lässt sich zumindest erklären, dass die Zahl 85 nach ähnlichen Exemplaren schneller verarbeitet werden kann. Etwas problematisch ist hingegen die Erklärung des Fehlereffektes nach den unähnlichen Exemplaren. In eine Begriffs-Konstruktion wird ja auch kontextunabhängige Information integriert (und dazu gehört auch die definitorische Information), was eigentlich Fehl kategorisierungen verhindern sollte.

Einen zweiten Erklärungsansatz stellt die Exemplartheorie von Brooks (1987) dar. Anstelle eines Konzeptes können auch leicht abrufbare Exemplare erinnert und neue Stimuli in *Analogie* zu diesen Exemplaren kategorisiert werden. Damit würde an die Stelle eines analytischen Kategorisierungsprozesses eine alternative, nichtanalytische Vorgehensweise der kategoriellen Zuordnung treten. „Aufgrund einer globalen Ähnlichkeit zu spezifischen, leicht abrufbaren Exemplaren wirkt ein neuer Stimulus oberflächlich vertraut und wird deshalb ohne vollständige Analyse als korrektes Exemplar akzeptiert“ (Morger, 1993, 26). Es ist nun denkbar, dass eine Liste ähnlicher Exemplare eine solche Strategie fördert und damit die Bereitschaft sinkt, im Zweifelsfall einen analytischen Vergleich durchzuführen, was dann zu einer hohen Fehleranfälligkeit führt, sobald eine konzeptkontroverse Information in Form einer geraden Zehnerstelle enthalten ist, was ja bei der Zahl 85 der Fall ist. Diese Erklärung wird unter anderem durch die Tatsache unterstützt, dass die konzeptkonforme Zahl 37 auch nach beiden Listen gut kategorisiert wird.

Mit dem berichteten Experiment ist es gelungen, einen robusten Effekt aus früheren Laboruntersuchungen in einem web-technologiebasierten Experiment zu replizieren. Da Web-Experimente gegenüber Laborexperimenten eine Reihe von Nachteilen ausgleichen, die in der Vergangenheit zu Zweifeln an der externalen Validität von laborexperimentellen Befunden geführt haben, kann man durch zuverlässige Replikation von Effekten in multimethodischem Vorgehen (Web, Labor, Feld) eine höhere Generalisierbarkeit der Befunde erreichen (Reips, 1997, 2000). Konsequenterweise betrachten wir den Listenkontexteffekt als hoch valide.

Allerdings hat der Fehlereffekt mit 13 % für die Web-Labor-Gruppe und 17 % für die Gruppe aus dem Experimental-Labor in Vancouver nicht dasselbe hohe Ausmass angenommen, wie in den von Morger (1993, 2000) berichteten Experimenten, wo in verschiedenen Varianten des Experimentes Fehlerraten zwischen 18% und 34 % gefunden wurden. Dieser Unterschied dürfte aber darauf zurückzuführen sein, dass in den ursprünglichen Experimenten die Entscheidungen via Antworttasten der PC-Tastatur abgegeben wurden, während in der Javascript-basierten Version per Maus-Klick auf das entsprechende Antwortfeld geantwortet wurde. Da diese Art der Entscheidung aus zwei Phasen (Bewegen des Mauspeils auf das Antwortfeld und Klicken der Maustaste) besteht, gab es eine zweite Gelegenheit, um die Antwort zu korrigieren, was zu einer Reduktion des Fehlereffektes geführt haben dürfte. Für diese Einschätzung spricht, dass sich die Fehlerraten zwischen der Web-Stichprobe und der Labor-Stichprobe in Vancouver nur wenig unterschieden.

Erfreulicherweise ist es gelungen, den Listenkontexteffekt bei den Antwortzeiten ohne Einschränkung überhaupt festzustellen, obwohl Javascript neben der sehr groben serverseitigen Logfile-Zeitmessung zu den technisch schwächsten Zeitmessungsverfahren im Internet gehört. Mit Javascript gemessene Zeiten können clientseitig durch vielfältige Fehlerquellen wie Hardware-, Temperatur- und Betriebssystemschwankungen, Hintergrundprozesse, Browserimplementationen und lokale Netzwerkstruktur (von sich kumulierenden Problemen komplexer Interaktionen ganz abgesehen) verzerrt werden. Mit Java-basierten oder geeichten Verfahren (Eichstaedt, 2000; Janetzko, 1999; Janetzko, Hildebrandt & Meyer, 2001) sollten sich prinzipiell zuverlässigere Zeiten erreichen lassen.

## 6. Extensionen

Das dargestellte Experiment eignet sich für eine Reihe möglicher Variationen in fliessendem Kontinuum vom Verändern einer Zahl in einer Liste bis hin zur Verwendung anderer Begriffe und Kategorien, die relativ leicht umgesetzt werden können. Aus diesem Grund halten wir dieses Web-Experiment für einen geeigneten Ausgangspunkt zur schnellen und einfachen Entwicklung eines Experiments.

Hier einige Vorschläge für mögliche Experimente mit unserem Material. Dabei schildern wir jeweils die Art, wie das Material transformiert wird, um zu einem neuen Experiment zu kommen und geben ein konkretes Beispiel.

1. Modifikation durch *Ersetzung*:

Erster Vorschlag: Ersetzen Sie die zweistelligen Zahlen im Experiment durch *dreistellige Zahlen*, bei denen Sie die Unterschiede zwischen den Listenkontexten dadurch verstärken, dass die Zahlen der Liste A nun aus drei ungeraden Ziffern zusammengesetzt sind, jene der Liste B hingegen zwei gerade Ziffern enthalten (zum Beispiel: 731, 573, 177 versus 265, 683, 845). Wird dadurch der Listenkontexteffekt stärker?

2. Modifikation durch *Ergänzung*:

Verlängern Sie einzelne Listen. Verstärkt sich der Effekt? Erhöht sich die Abbruchquote?

3. Modifikation durch *Streichung*:

Verkürzen Sie die Warm-up-Phase zu einer kurzen Übungsphase. Welche Auswirkungen ergeben sich im Hinblick auf die Abbruchquote im eigentlichen Experiment? Hat dies vielleicht zur Folge, dass sich bei der Listenversion A keine nichtanalytische Strategie etabliert und der Listenkontexteffekt somit ausbleibt?

4. Modifikation durch *Komplementarität*:

Dies ist ein Spezialfall der Modifikation durch Ersetzung. Führen Sie das Experiment mit der Kategorienbezeichnung "*even*" im Zahlenblock durch. Kann der Listenkontexteffekt auf diese Art in gleicher Stärke repliziert werden?

*Weitere Ideen:*

Führen Sie das Web-Experiment lokal mit Leuten aus verschiedenen Personengruppen durch, zum Beispiel mit Kindern, regelmässigen Computerspielern, "Computer-Analphabeten", Personen mit Teilleistungsschwächen ... Könnte man die Aufgabe evtl. als Diagnoseinstrument einsetzen?

Benutzen Sie die *Multiple Site Entry* Technik (Reips, 1999, 2000, Kapitel 4 in diesem Band) - verlinken Sie Ihr Web-Experiment also auf verschiedenen Websites. Finden Sie den gleichen Effekt trotz der verschiedenen Zusammensetzung Ihrer Stichproben?

Versuchen Sie, die Messung der Antwortzeiten mit Hilfe eines anderen Verfahrens (Java, serverseitige Erfassung etc.) durchzuführen.

## 7. Literatur

Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. *Cognition*, 13, 263 - 308.

In dieser Arbeit wird erstmals gezeigt, dass auch exakt definierte Begriffe eine Typikalitätsstruktur aufweisen.

Barsalou, L. W. (1989). Intraconcept similarity and its implications for inter-concept similarity. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 76 - 121). Cambridge: Cambridge University Press.

Darstellung einer Theorie, in der Begriffe als adaptive, momentane Konstruktionen verstanden werden.

Bromme, R. (1990). Prototypikalität bei exakt definierten Begriffen: Das Beispiel der geraden und ungeraden Zahlen. *Sprache & Kognition*, 9, 155 - 167.

- Methodisch verbesserte Replikation der Arbeit von Armstrong, Gleitman und Gleitman (1983) zum Nachweis des Typikalitätseffektes bei geraden und ungeraden Zahlen.
- Brooks, L. R. (1987). Decentralized control of categorization: the role of prior processing episodes. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual Factors in categorization* (pp. 141 - 174). Cambridge: Cambridge University Press.
- Darstellung der Exemplartheorie als Alternative zur Prototypentheorie, welche aber ebenso die Typikalitätsstruktur von Begriffen zu erklären vermag. Postuliert, dass begriffliche Entscheidungen wesentlich von nichtanalytischen Prozessen gesteuert werden.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Der Klassiker zum Thema Effektstärke.
- Eckes, T. (1985). Zur internen Struktur semantischer Kategorien: Typikalitätsnormen auf der Basis von Ratings. *Sprache & Kognition*, 4, 192 - 202.
- Normierungsuntersuchung zur Begriffsstruktur bei semantischen Kategorien, operationalisiert durch Typikalitäts-Ratings.
- Eckes, T. (1996). Begriffsbildung. In J. Hoffmann (Hrsg.), *Enzyklopädie der Psychologie* (S. 273 - 318). Göttingen: Hogrefe.
- Hervorragende Überblicksarbeit zum Thema Begriffspsychologie. Darstellung der wichtigsten theoretischen Positionen.
- Eichstaedt, J. (2000). Reaction time measurement by JAVA-applets implementing Internet-based experiments. Manuscript submitted for publication.
- Empirische Überprüfung einer neuentwickelten Java-basierten Methode zur Reaktionszeitmessung in Web-Experimenten.
- Flammer, A., & Morger, V. (1985). Die Wirkungen von Voraktivierung und Typikalität auf die Verifikationszeiten von Begriffsexemplaren. *Schweizerische Zeitschrift für Psychologie*, 44, 1 - 16.
- Empirische Untersuchung zum Zusammenhang der Typikalität und der benötigten Antwortzeit bei der Begriffsverifikation.
- Frick, A., Bächtiger, M. T., & Reips, U.-D. (in press). Financial incentives, personal information and drop-out in online studies. In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet science*.
- Erster experimenteller Nachweis eines Einflusses von finanziellen Anreizen und dem Abfragen persönlicher Informationen auf die Abbruchrate in Online-Studien.
- Janetzko, D. (1999). *Statistische Anwendungen im Internet*. München: Addison-Wesley.
- Übersichtsarbeit zu den Möglichkeiten des Einsatzes statistischer Anwendungen im Internet.
- Janetzko, D., Hildebrandt, M., & Meyer, H. (2001). Zeitmessungen bei Fragebogenerhebungen. Manuskript in Vorbereitung.
- Empirische Studie zu einer selbstentwickelten Technik der Zeitmessungen bei Fragebogenerhebungen.
- Mannhaupt, H. R. (1983). Produktionsnormen für verbale Reaktionen zu vierzig geläufigen Kategorien. *Sprache & Kognition*, 2, 264 - 278.
- Normierungsuntersuchung zur Produktionsfrequenz von Begriffs-Exemplaren; bildet die Typikalitätsstruktur auf der Basis der Nennhäufigkeit von Exemplaren ab.
- Morger, V. (1993). Nichtanalytische Kategorisierungsprozesse bei der Exemplar-

- Verifikation exakt definierter Konzepte: Nochmals das Beispiel der geraden und ungeraden Zahlen. *Sprache & Kognition*, 12, 18 - 29.  
Erstmaliger Nachweis des Listenkontexteffektes bei der Begriffs-Verifikation mit den exakt definierten Begriffen „gerade Zahl“ und „ungerade Zahl“.
- Morger, V. (2000). *Erfahrungsnutzung bei begrifflichen Kognitionen: Episodisch verursachte Flexibilität des semantischen Wissens*. Lengerich: Pabst (Aktuelle Psychologische Forschung, Band 32).  
Enthält mehrere Replikationen des Listenkontexteffektes, sowohl mit exakt definierten, als auch mit semantischen Begriffen (Kapitel 6). In Kapitel 2 wird ein Überblick zur Theoriebildung im Bereich Begriffe und Kategorien gegeben.
- O'Neil, K. M. & Penrod, S. D. (2000). Effects of dropout rate and efforts to reduce it in one Web-based experiment. Paper presented at the 30<sup>th</sup> Society for Computers in Psychology conference, New Orleans, November 16.  
Replikationsversuch der Studie von Frick, Bächtiger & Reips (in press).
- Reips, U.-D. (1997). Das psychologische Experimentieren im Internet. In B. Batinic (Hrsg.), *Internet für Psychologen* (S. 245-265). Göttingen: Hogrefe.  
Erste deutschsprachige Grundlagenarbeit zum Web-Experimentieren.
- Reips, U.-D. (1999). Theorie und Techniken des Web-Experimentierens. In B. Batinic, A. Werner, L. Gräf, & W. Bandilla (Hrsg.), *Online Research: Methoden, Anwendungen und Ergebnisse* (S. 277-296). Göttingen: Hogrefe.  
Schildert die theoretischen Grundlagen und eine Reihe von Techniken des Web-Experimentierens.
- Reips, U.-D. (2000). The Web Experiment Method: Advantages, Disadvantages, and Solutions. In M. H. Birnbaum (Ed.), *Psychological Experiments on the Internet* (pp. 89-114). San Diego, CA: Academic Press.  
Umfassende Darstellung der web-experimentellen Methodologie, besonders auch im Vergleich zur laborexperimentellen Methode. Erläuterung von Vor- und Nachteilen, Techniken, Lösungen und Anwendungen.
- Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In T. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111 - 144). New York: Academic Press.  
Darstellung von Untersuchungen, welche die Prototypentheorie als Erklärungsansatz der begrifflichen Wissenstruktur unterstützen.
- Rosch, E. (1978). Principles of categorisation. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorisation* (pp. 27 - 48). Hillsdale, NJ: Erlbaum.  
Ausführliche Darstellung der Prototypentheorie, wie sie von Rosch im Rahmen der Begriffspsychologie propagiert wurde.
- Schwarz, S. & Reips, U.-D. (in press). CGI versus JavaScript: A Web Experiment on the Reversed Hindsight Bias. In U.-D. Reips, & M. Bosnjak (Eds.), *Dimensions of Internet Science*.  
Web-experimentelle Arbeit zum Auftreten des umgekehrten Rückschaufehlers in Abhängigkeit von Überraschung und Betroffenheit und zur Problematik der Verwendung von Javascript in Online-Studien.