

Mentale Modelle und Webnavigation

*Ein Usability-Experiment zur Informationssuche
im World-Wide-Web*

Als Diplomarbeit vorgelegt dem Vorsitzenden des
Prüfungsausschusses für die Diplomprüfung im Fach
Psychologie an der Universität zu Köln

von Stephan Noller aus Köln

angefertigt bei

Prof. Dr. N. Groeben, Universität
zu Köln

Prof. Dr. A. Kobsa, University
of California, Irvine

Köln, den 23. August 2000

Danken möchte ich meinen Betreuern, Prof. Dr. N. Groeben, Prof. Dr. A. Kobsa, Jürgen Koenemann, Johannes Naumann und Tobias Richter für ihre hervorragende Betreuung über alle Phasen der Arbeit.

Ein besonderer Dank gilt auch Ulrich Noller für die mühevollen Korrekturarbeit und Maria Brandt für ihre wundervolle Unterstützung und jede Minute in dieser Zeit. Nicht zuletzt möchte ich mich bei meiner geliebten Tochter Laura bedanken, die viel zu oft hören mußte, daß der Papa noch arbeiten muß...

Inhaltsverzeichnis

1.	Einleitung, Ausgangspunkt, Ziel der Arbeit	7
2.	Theoretischer Hintergrund	10
2. 1.	Die Theorie mentaler Modelle als Hilfskonstruktion in der Forschung zur Human-Computer-Interaction	10
2. 1. 1.	Vorstrukturierung	10
2. 1. 2.	Einführung - komplexe Probleme, Informationsverarbeitung und mentale Modelle	10
2. 1. 3.	Abgrenzung zu konkurrierenden Theoriekonzeptionen	21
2. 1. 3. 1.	<i>Vorbemerkung zur Auswahl</i>	21
2. 1. 3. 2.	<i>Information foraging (Pirolli & Card 1999)</i>	22
2. 1. 3. 3.	<i>Uses and Gratifications (Wirth & Schweiger, 1999)</i>	25
2. 1. 4.	Allgemeine Modelltheorie	25
2. 1. 5.	Eigenschaften und Funktionen mentaler Modelle	27
2. 1. 5. 1.	<i>Unvollständigkeit</i>	27
2. 1. 5. 2.	<i>Analogien, Metaphern</i>	28
2. 1. 5. 3.	<i>Schemata</i>	31
2. 1. 5. 4.	<i>Komplexes Problemlösen, Schlußfolgerndes Denken</i>	32
2. 1. 6.	Mentale Modelle und Navigation in Hypertexten	33
2. 1. 7.	Zusammenhang von spatial ability und Navigationskompetenz	34
2. 2.	Hypertext im Blickpunkt der Human-Computer-Interaction	36
2. 2. 1.	Vorstrukturierung	36
2. 2. 2.	Hypertext und hypermediale Lernsysteme	37
2. 2. 2. 1.	<i>Grundprinzip: Worin unterscheiden sich Hypertexte von linearen Texten? ...</i>	37
2. 2. 2. 2.	<i>Vor- und Nachteile des Lernens mit Hypertexten</i>	39
2. 2. 2. 3.	<i>Das Problem des Orientierungsverlusts und der Navigation in Hypertexten</i>	44
2. 2. 3.	Interaktionsstile im Design der Mensch-Maschine-Schnittstelle	47
2. 2. 3. 1.	<i>Vorbemerkung</i>	47
2. 2. 3. 2.	<i>Menu-Selection</i>	47
2. 2. 3. 3.	<i>Direct Manipulation</i>	49
2. 2. 4.	Probleme bei der Visualisierung komplexer Informationen	50

2. 2. 4. 1.	<i>Zusammenhang mit Menüsystemen</i>	50
2. 2. 4. 2.	<i>Fokussieren und Kontext</i>	51
2. 2. 5.	Verschiedene „information needs“ und aufgabenabhängige Suchstrategien ..	52
2. 2. 6.	Lösungsversuche	54
2. 2. 6. 1.	<i>Vorbemerkung zur Auswahl der Lösungsversuche</i>	54
2. 2. 6. 2.	<i>WebToc</i>	54
2. 2. 6. 3.	<i>Hyperbolic-Tree-Browser</i>	55
2. 2. 7.	Usability-Forschung	58
2. 2. 7. 1.	<i>Vorbemerkungen</i>	58
2. 2. 7. 2.	<i>usability-Ziele</i>	59
2. 2. 7. 3.	<i>Kriterien für „usable systems“</i>	61
2. 2. 7. 4.	<i>usability-testing</i>	61
2. 2. 7. 5.	<i>Zusammenfassung</i>	62
3.	Fragestellung und Hypothesen	64
3. 1.	Fragestellung der Untersuchung	64
3. 2.	Hypothesen	64
3. 2. 1.	Hypothese zum Einfluß der angebotenen Navigationshilfen	64
3. 2. 2.	Hypothesen zu Kovariaten	65
3. 2. 2. 1.	<i>Einfluß des Faktors spatial ability</i>	65
3. 2. 3.	Interaktionshypothesen	65
3. 2. 3. 1.	<i>Interaktion von Fragentyp und Navigationshilfe</i>	65
4.	Methode	66
4. 1.	Skizzierung der empirischen Vorgehensweise	66
4. 2.	Generelle methodische Probleme von Online-Untersuchungen	66
4. 3.	Untersuchungsmaterial	68
4. 4.	Operationalisierung der Hypothesen - verwendete Erhebungsverfahren	69
4. 4. 1.	UV1: Präsentation. Unterschiedliche Navigationshilfen	69
4. 4. 1. 1.	<i>Klassische Präsentation mit statischer Überblicksseite und einer Navigationsleiste</i>	69

4. 4. 1. 2.	<i>Lineares Inhaltsverzeichnis (realisiert mit WebToc)</i>	70
4. 4. 1. 3.	<i>Dynamischer graphischer Browser (realisiert mit Hyperbolic-Tree)</i>	71
4. 4. 2.	AV1: Fragentyp	72
4. 4. 2. 1.	<i>Vorbemerkung</i>	72
4. 4. 2. 2.	<i>AV1a: Retrieval-Fragen</i>	72
4. 4. 2. 3.	<i>AV1b: komplexe Struktur-Fragen</i>	73
4. 4. 3.	AV2: Eigenschaften des Navigationspfades (aus Logfile-Analysen)	74
4. 4. 3. 1.	<i>Vorbemerkung: Einschränkungen</i>	74
4. 4. 3. 2.	<i>AV2a: Anzahl besuchter Knoten</i>	74
4. 4. 3. 3.	<i>AV2b Anzahl vom Tool aus aufgerufener Knoten</i>	75
4. 4. 3. 4.	<i>AV2c: Stratum</i>	75
4. 4. 3. 5.	<i>AV2d: Compactness</i>	75
4. 4. 3. 6.	<i>AV2e Aufruf bestimmter Seitentypen</i>	75
4. 4. 4.	AV3: Lösungsgüte	76
4. 4. 4. 1.	<i>AV3a: Anzahl richtiger Antworten</i>	76
4. 4. 4. 2.	<i>AV3b: Bearbeitungsdauer</i>	76
4. 4. 5.	AV4: usability	76
4. 4. 6.	Kovariate: Spatial Ability	77
4. 4. 7.	Explorative Fragen zum Experiment	77
4. 5.	Der Versuchsplan im Überblick	78
4. 6.	Stichprobe	78
4. 6. 1.	Vorbemerkungen	78
4. 6. 2.	Zusammensetzung der Untersuchungsstichprobe	80
4. 7.	Ablauf der Untersuchung	81
4. 8.	Kontrolle von Störvariablen	82
5.	Durchführung der Untersuchung	84
6.	Darstellung der Ergebnisse	85
6. 1.	Vorstrukturierung	85
6. 2.	Technische Vorbemerkungen	86
6. 3.	Untersuchung auf Homogenität über die Zellen	86

6. 4.	Beurteilung der Voraussetzungen für eine varianzanalytische Untersuchung	86
6. 5.	Univariate Analysen der Retrieval bzw. Strukturfragen in Abhängigkeit der Navigationshilfe	89
6. 5. 1.	Psychometrische Kennwerte der Skalen	89
6. 5. 2.	Univariate Varianzanalysen: Anzahl richtiger Antworten	92
6. 5. 3.	Univariate Varianzanalysen: Bearbeitungsdauer	93
6. 6.	Analysen zur Kovariaten spatial ability	95
6. 6. 1.	Vorbemerkung	95
6. 6. 2.	Psychometrische Kennwerte der Skalen zur Erhebung von spatial ability	95
6. 6. 3.	Verteilung der Werte für spatial ability	95
6. 7.	Multivariate Analysen unter Einbeziehung der Kovariaten spatial ability	97
6. 7. 1.	Werte für Retrieval-Fragen	98
6. 7. 2.	Werte für Struktur-Fragen	101
6. 7. 3.	Untersuchung der zwei Aspekte der AV Fragentyp in Abhängigkeit der Navigationsbedingung	103
6. 8.	Analyse der Daten des usability-Fragebogens	105
6. 8. 1.	Reliabilität	105
6. 8. 2.	Faktorielle Struktur	106
6. 8. 3.	Ergebnisse für die unterschiedlichen Navigationsbedingungen	109
6. 9.	Analysen einzelner Kennwerte aus den Logfiles in Abhängigkeit von der Navigationsbedingung	113
6. 9. 1.	Unterschiede hinsichtlich Stratum und Compactness	113
6. 9. 2.	Anzahl besuchter Knoten	116
6. 9. 3.	Aufruf der Seitentypen „Index“ und „Suche“	117
6. 10.	Exploratorische Einzelanalysen der graphentheoretischen Logfilemaße für die Stufen des Faktors Navigationshilfe	118
6. 10. 1.	Faktorstufe ohne Hilfe	119
6. 10. 2.	Faktorstufe WebToc	120
6. 10. 2. 1.	Faktorstufe Hyperbolic-Tree	121

6. 11.	Analysen zu Bearbeitungsdauer und Anzahl richtiger Antworten in Abhängigkeit der Internet-Erfahrung	121
6. 11. 1.	Analyse in Abhängigkeit der Anzahl richtiger Antworten und der Bearbeitungsdauer	122
6. 11. 2.	Analyse des Zusammenhangs mit usability	126
6. 12.	Analysen der zusätzlichen Fragen zum Experiment	129
6. 12. 1.	Schätzung über die Anzahl der Knoten in der Untersuchungs-Site	129
6. 12. 2.	Weitere Zusatzfragen	129
6. 12. 3.	Korrelationen der Antwortausprägungen für die Zusatzfragen mit den abhängigen Variablen	131
6. 12. 4.	Aufruf der Hilfsfunktion	132
6. 13.	Analyse der Bearbeitungszeit des Einführungstutorials	133
6. 14.	Analyse von möglichen Lerneffekten	135
6. 14. 1.	Lerneffekte bei Retrieval-Fragen	135
6. 14. 2.	Lerneffekte bei Struktur-Fragen	136
6. 14. 3.	Lerneffekte über alle Meßzeitpunkte	137
6. 15.	Analyse des Dropouts	138
7.	Analyse der Ergebnisse und Diskussion	146
7. 1.	Zusammenfassung der Ergebnisse	146
7. 1. 1.	Ergebnisse der Suchaufgaben	146
7. 1. 2.	Usability-Beurteilung	149
7. 1. 3.	Logfile-Maße	150
7. 2.	Dropout-Analyse	152
7. 3.	Zusätzliche Fragen zum Experiment	153
7. 3. 1.	Weitere exploratorische Analysen	153
7. 4.	Diskussion	156
7. 5.	Ausblick	163
8.	Literaturverzeichnis	165
9.	Abbildungsverzeichnis	174

10.	Tabellenverzeichnis	176
11.	Anhang	178

1. Einleitung, Ausgangspunkt, Ziel der Arbeit

I feel ... an ardent desire to see knowledge so disseminated through the mass of mankind that it may, at length, reach even the extremes of society: beggars and kings.

Thomas Jefferson, Reply to American Philosophical Society, 1808

Obwohl bisher die häufig euphorisch gelobten Möglichkeiten des Arbeitens mit Hypertexten bisher nicht überzeugend gesichert werden konnten (Chen & Rada 1996, Tergan 1997), wird deren Benutzung in Hilfesystemen, Lernprogrammen, Informationssystemen usw. zunehmend Bestandteil unseres Alltages werden. Viele Informationen sind schon heute nur noch 'online' verfügbar und es ist absehbar, daß dieser Trend sich nicht zuletzt aus ökonomischen Gründen ungemindert fortsetzen wird. Insofern kann es nicht nur aus der Perspektive grundlagenwissenschaftlich-allgemeinpsychologischer Forschung, sondern auch im Sinne einer emanzipatorischen Perspektive der Psychologie weiterhin als hoch-interessante Frage gelten, wie das Lernen und Auffinden von Informationen mit Hypertexten vonstatten geht und wie evtl. vorhandene Hindernisse durch entsprechende Gestaltungsmittel und Hilfskonstruktionen überwindbar gemacht werden könnten.

Berücksichtigt man den ungeheuren Zuwachs der Nutzung des Internet (vgl. Musch, 1997; GfK-Online-Monitor, 2000), rückt desweiteren sowohl ein relevanter ökonomischer Faktor, als auch ein politischer in den Blickpunkt. Ökonomisch insofern, als daß in vielen großen Firmen bereits heute Kommunikation und Informationsaustausch auf der Basis der Hypertext-Technologie umgesetzt wird (auch in internen Firmennetzen, „Intranets“ bzw. „Knowledge-Management“-Umgebungen) und schon geringe konzeptionelle Mängel in den Navigationsmöglichkeiten bereits zu erheblichen Minderungen der Effizienz des Unternehmens führen würden (Cockburn & Jones, 1996). Der politische Faktor liegt darin begründet, daß vielerorts die Verbreitung des nicht hierarchisch geordneten Internet mit einem Potential zur Demokratisierung in Verbindung gebracht wird (so ist z.B. das Konzept der *universal usability* von B. Shneiderman zu verstehen, Shneiderman, 2000). Obwohl dabei in erster Linie die dezentrale Organisation hinsichtlich des Zuganges und der Möglichkeiten eigener „Publikationen“ gemeint ist (z.B. Press, 1997), kann auch generell die leichtere Zugänglichkeit von Informationen als ein Beitrag zur Nivellierung diesbezüglicher Hierarchien in der Gesellschaft interpretiert werden: („Research could pave the way

for broad citizen participation in quality online services and novel social, economic, and political programs“, Press 1997, S. 91 - oder im Hinblick auf totalitäre Systeme: „We see information flowing into and out of dictatorial nations. More important, we see it circulating within those nations, between individuals, political organizations, human-rights organizations, and so forth“, Press, 1997, S. 16). Hier würde sich z. B. eine hohe Einstiegsschwelle etwa durch Experten/innen-zentrierte Navigationsstrukturen ebenfalls hinderlich auswirken. Andererseits könnten durch verteilte Hypertextsysteme in Verbindung mit dem Internet tatsächlich Informationen mit vergleichsweise geringem ökonomischen Aufwand verfügbar gemacht werden, an die z. B. eine Schule in Afrika sonst nicht herangekommen wäre.

Mit dieser Motivation soll im Rahmen der vorliegenden Arbeit der Frage nachgegangen werden, ob (mehr oder weniger adäquate) mentale Modelle (Dutke, 1994; Johnson-Laird, 1983; Gentner & Stevens, 1983) bei der Navigation in komplexen Hypertexten des World-Wide-Web eine Rolle spielen und ob mit gezielten gestalterischen Maßnahmen die Generierung derartiger Modelle beim User gefördert werden kann oder nicht. Ein Schwerpunkt bei der Beurteilung der Auswirkungen der angebotenen Navigationsmittel soll dabei der Frage zukommen, inwieweit es möglich ist, das häufig berichtete Phänomen des Orientierungsverlusts ('Disorientation', 'Lost in hyperspace' usw.) in komplexen Hypertext-Umgebungen (Jih & Reeves, 1992; McDonald & Stevenson, 1996; Cockburn & Jones, 1996; Schnotz, Picard & Henninger, 1994; Gerdes, 1997a, b) zu vermeiden. Neben naheliegenden Analysen der Bearbeitungszeit und der Problemlösequalität wird dabei der Interpretation der Logfile-Daten besondere Aufmerksamkeit gewidmet werden. Hier soll im Zusammenhang mit der Operationalisierung der abhängigen Variablen versucht werden, Möglichkeiten einer psychologischen Interpretation dieser anfallenden Daten aufzuzeigen (vgl. Rauterberg, 1992; Sullivan, 1997; Weber & Specht, 1997; Rauterberg & Fjeld, 1998; Catledge & Pitkow, 1998).

Im folgenden wird zunächst die Theorie der mentalen Modelle angeführt und diskutiert, inwieweit diese einen geeigneten Hintergrund darstellen kann, um Suchoperationen in Hypertexten zu modellieren. Dabei soll auch auf konkurrierenden Theoriekonzeptionen kurz eingegangen werden. Es folgt eine Darstellung des Anwendungsbereichs der Human-Computer-Interaction. Hierbei sollen sowohl die Überschneidungspunkte zur Theorie der mentalen Modelle aufgezeigt als auch klassische Probleme dieses Forschungsfeldes beleuchtet

werden, die im Zusammenhang mit der vorliegenden Arbeit von Bedeutung sind. Um den Zusammenhang mit der später vorgestellten Untersuchung deutlich zu machen, werden hier bereits paradigmatische Lösungsansätze für zentrale Probleme vorgestellt. Für Untersuchungen zur HCI ist auch immer das Thema *usability* von großer Bedeutung, da HCI als angewandte Forschung nicht nur die theoretische Modellierung des Interaktionsprozesses zwischen Mensch und Computer zum Ziel haben kann, sondern sich immer auch darum bemüht, begründbare Kriterien für brauchbare Systeme aufzustellen. Einige Grundgedanken und davon abgeleitete Kriterien werden abschließend für den Theorieteil dargestellt.

Aufbauend auf diesen theoretischen Erörterungen wird dann die in diesem Zusammenhang durchgeführte empirische Untersuchung vorgestellt, wobei ein besonderer Schwerpunkt auf speziellen methodischen Problemen liegt, die sich durch das Design als Online-Untersuchung ergeben.

Es folgt eine ausführliche Analyse der gewonnenen Daten. Die Auswahl der analysierten Daten und der Analyseverfahren orientiert sich dabei an der Fragestellung der Untersuchung und soll die Grundlage zur einer Beurteilung der zuvor aufgestellten Hypothesen darstellen.

Im Diskussionsteil werden zunächst die Ergebnisse nochmals komprimiert dargestellt. Abschließend soll dann versucht werden, Erklärungsansätze für die vorgestellten Ergebnisse zu finden, insbesondere unter Rückbezug auf die zuvor formulierten theoretischen Erörterungen.

2. Theoretischer Hintergrund

2.1 Die Theorie mentaler Modelle als Hilfskonstruktion in der Forschung zur Human-Computer-Interaction

2.1.1. Vorstrukturierung

Im folgenden soll zunächst der Zusammenhang der Psychologie des komplexen Problemlösens, der Theorie, bzw. Theorien der mentalen Modelle und dem Lösen von Suchaufgaben in Hypertextumgebungen beleuchtet werden. In diesem Zusammenhang wird die Auswahl der Rahmentheorie auch in Abgrenzung anderer theoretischer Modelle erfolgen. Nach einer detaillierten Darstellung der Eigenschaften und Funktionen mentaler Modelle (unter Rückbezug auf die allgemeine Modelltheorie) werden dann Zusammenhänge mit Navigation und Informationssuche in Hypertexten hergestellt. Zusätzlich soll das Fähigkeitskonzept *spatial ability* in die Überlegungen mit einbezogen werden.

2.1.2. Einführung - komplexe Probleme, Informationsverarbeitung und mentale Modelle

Die Theorie der mentalen Modelle (Johnson-Laird, 1983; Gentner & Gentner, 1983) ist im Zusammenhang mit Überlegungen und Untersuchungsbefunden zum komplexen Problemlösen entstanden. Um den Entstehungszusammenhang und damit verbundene theoretische Implikationen zu beleuchten, kann diese Theorie nicht ohne einige Bemerkungen zum interdisziplinären „Informationsverarbeitungsansatz“ diskutiert werden, der in den letzten 30 Jahren im Zusammenhang mit Forschungen zur „cognitive science“ eine immer größere Bedeutung auch in der psychologischen Theorienbildung bekam.

Newell und Simon (1987) haben in ihrem Vortrag anlässlich der Verleihung des Turing-Awards 1975 im Rahmen eines „law of qualitative structure for symbol systems“ die Hypothese aufgestellt, daß Systeme, die in der Lage sind mit Symbolen zu operieren, prinzipiell auch in der Lage sind, intelligente Operationen auszuführen („*The Physical Symbol System Hypothesis*. A physical symbol system has the necessary and sufficient means for general intelligent action“, Newell & Simon 1987, S. 41). Die Autoren nehmen für diese Hypothese auch die Gültigkeit des Umkehrschlusses an: „the symbol system hypothesis implies that the symbolic behavior of man arises because he has the characteristics of a physical symbol system“ (Newell & Simon 1987, S. 49). Im Rahmen dieser Grundannahmen über die Arbeitsweise intelligenter Systeme wurden sowohl elaborierte psychologische Theorien über menschliche Prozesse der Informationsverarbeitung formuliert, als auch

Simulationsprogramme entworfen, bei deren Entwicklung anhand gleicher Grundannahmen versucht werden sollte, intelligentes Verhalten „herzustellen“. Der Indikator für eine Klassifikation von Verhalten als „intelligent“ ist dabei meist die Fähigkeit zur Lösung von (komplexen) Problemen. Abgesehen von der Frage, inwieweit es tatsächlich möglich ist, durch Computersimulationen die Richtigkeit der zugrundeliegenden Annahmen bezüglich menschlicher Informationsverarbeitung zu überprüfen (Newell und Simon weisen auf dieses Problem hin und fordern „psychological experiments [are] required to test the veridicality of the simulation models as explanations of the human behavior“, Newell & Simon 1987, S. 49) scheint es doch so zu sein, daß sich die Grundannahmen der *Physical Symbol System Hypothesis* für die Analyse der Prozesse beim Lösen komplexer Probleme als brauchbares Theoriengebäude erwiesen hat. Für die Erklärungsversuche anderer Theorien-Komplexe in der Psychologie, wie z.B. Gestaltpsychologie oder Behaviorismus, kommen Newell und Simon in ihrem Vortrag hingegen zu dem Schluß: „As a matter of fact, the alternative theories are so vague that it is not terribly difficult to give them information-processing interpretations, and thereby assimilate them to the symbol-system-hypothesis.“ (Newell & Simon 1987, S. 50).

Bevor nun mentale Modelle im Zusammenhang mit Problemlöseprozessen diskutiert werden, sollten zwei wesentliche Konzepte kurz definiert werden, die in der folgenden Diskussion zentral gegeneinander abgehoben werden: mentale Modelle vs. propositionale Repräsentationen.

Zur Definition eines mentalen Modells soll Johnson Laird (1998) zitiert werden:

Mental models are psychological representations of real, hypothetical, or imaginary situations. [...] Like pictures in Wittgenstein's (1922) 'picture' theory of the meaning of language, mental models have a structure that corresponds to the structure of what they represent. They are accordingly akin to architects' models of buildings, to molecular biologists' models of complex molecules, and to physicists' diagrams of particle interactions. (S. 1)

Im Gegensatz dazu sind Propositionen Strukturen, „die aus einer Relation und einer geordneten Menge von Argumenten besteht“ (Anderson 1996, S. 142). Aus diesen gegensätzlichen Annahmen bezüglich der Repräsentation des Wissens lassen sich weitere Schlußfolgerungen bezüglich der Operationen ziehen, die Menschen mit ihrem Wissen vollziehen, und vor allem hinsichtlich charakteristischer Probleme und Fehler, die dabei beobachtet werden können.

Im folgenden wird dieser Grundkonflikt insbesondere aus der Sicht Johnson-Lairds beleuchtet werden. Es sei aber bereits auf die Möglichkeit einer Abschwächung, bzw. einer integrativen Sichtweise verwiesen, wie sie Newell (1990) vorgeschlagen hat (s.u.).

Johnson-Laird beschreibt seinen Versuch, menschliche Kognitionen beim Lösen von Problemen mit der Annahme der Konstruktion mentaler Modelle zu beschreiben, explizit als eine Weiterentwicklung der genannten Informationsverarbeitungstheorien (Johnson-Laird, 1983). Die Notwendigkeit für Weiterentwicklungen prinzipieller Art sieht er in dem Umstand begründet, daß die bisherigen Theorien zwar brauchbare Analysen bzw. Modellierungen der Rationalität beim Lösen von Problemen bieten, aber zwei wichtige Aspekte nicht oder nur ungenügend abbilden können: typische Fehler und die Performanz/Schwierigkeit von Problemen.

The present theories are too fragile to bear the weight of human reason. [...] Newell holds fast to logical power, but fails to explain systematic error. The dilemma to be resolved is to allow for both rationality and human error. If my diagnosis is correct, what is needed is both a simpler and a more natural conception of a mental model. (Johnson-Laird 1983, S. 93)

Mit seiner Theorie der mentalen Modelle versuchte Johnson-Laird (1983) zunächst eine Alternative zur Vorstellung einer Speicherung von Informationen beim syllogistischen Schließen in der Form propositionaler Kodierungen zu finden. Wie mentale Modelle - auch in Abgrenzung zur allgemeinen Modelltheorie - genau zu verstehen sind, soll weiter unten ausgeführt werden. An dieser Stelle muß hingegen noch die prinzipielle Frage geklärt werden, in welcher Weise der Ansatz der „mentalen Modelle“ in besonderer Weise geeignet erscheint, um die psychologischen Prozesse bei der Informationsverarbeitung im Umgang mit Hypertextsystemen zu modellieren. Dies wird umso dringender, wenn man die spätere Nivellierung des oftmals stark betonten Unterschiedes propositionaler Kodierung und einer solchen mit mentalen Modellen von Newell (1990) in Betracht zieht, der feststellt „Models and propositions are both representations. [...] Furthermore, both propositions and models can be symbol structures, and both can be composed so as to satisfy a wide range of representational laws.“ (Newell 1990, S. 388f). Konsequenterweise führt Newell auch eine Zwischenform mentaler Modelle ein, und zwar „annotated models“, die dadurch gekennzeichnet sind, daß sie zwar nach den Regeln mentaler Modelle gebildet werden und entsprechend zur Problemlösung beitragen können (d.h. Lösungen können „abgelesen“

werden), aber zusätzlich gewissermaßen mit „Anmerkungen“ in propositionaler Struktur versehen sind.

Zur Charakterisierung der Unterschiede zwischen propositionaler Kodierung und der Informationsverarbeitung mit mentalen Modellen betrachtet Newell (1990) beide Konstrukte hinsichtlich zweier Dimensionen: Abbildungsmächtigkeit („representational scope“) und Verarbeitungsaufwand („processing cost“). Propositionen erweisen sich als stark hinsichtlich der Abbildungsleistung (es können z.B. auch Negationen und mehrdeutige Situationen problemlos abgebildet werden), insbesondere bei komplexen Problemen allerdings als schwach bzw. aufwendig hinsichtlich des anzunehmenden Verarbeitungsaufwandes. Im Gegensatz dazu zeigen mentale Modelle geringe Anforderungen an den Verarbeitungsaufwand und - sozusagen als Preis - eine verminderte Abbildungsmächtigkeit (dabei sollen Aussagen über die vermuteten kognitiven Prozesse gemacht werden, nicht jedoch über die etwaige Abbildungsleistung der Theorien im Sinne einer Beschreibung kognitiver Prozesse). Die Suche von Informationen in Hypertextsystemen gleicht, insbesondere bei komplexen Aufgaben, in vielen Punkten den vermuteten Prozessen bei der Lösung komplexer Probleme im engeren Sinne. So stellen Newell und Simon (1987) fest, daß „symbol systems solve problems by using processes of heuristic search“ (S. 60) - womit die Hypothese vergleichbarer Prozesse bei der Informationssuche in Hypertexten - auch wenn es dabei nicht primär um das Lösen komplexer Probleme gehen muß - weiter gestützt werden kann.

Wenn nun - unabhängig von empirischen Befunden zu dieser Frage - theoretisch entschieden werden sollte, welche der genannten Repräsentationen des Problemraumes im Falle der Informationssuche in Hypertexten eine stärkere Bedeutung haben könnte, kann m.E. die oben ausgeführte Klassifikation von Newell (1990) herangezogen werden, um die Frage zu beantworten. Aufgrund dieser Klassifikation kann angenommen werden, daß Repräsentationen in der Form mentaler Modelle aufgrund der geringeren Anforderungen an die Verarbeitungskapazität für die Informationssuche - vor allem in Internet-basierten Hypertexten - eine größere Rolle spielen könnten, als propositionale Kodierungen. Sicherlich ist dies nicht viel mehr als eine Spekulation und bedarf der empirischen Überprüfung, aber ein weiterer Anhaltspunkt für die Annahme dieser Präferenz kann darin gesehen werden, daß Hypertexte sowohl in der Theorie als auch im alltäglichen Sprachgebrauch und nicht zuletzt in ihren diversen Implementierungen überaus häufig unter Zuhilfenahme vielfältiger Metaphern dargestellt werden (z.B. „Netz“, „Baum“, „Landschaft“).

Die im Zusammenhang mit Hypertextsystemen desweiteren überaus gebräuchliche Metapher der Navigation in einer „Informationslandschaft“ wurde bereits lange vor der Entwicklung des Internet von Wittgenstein (1999) 1945 im Vorwort zu den „Philosophischen Untersuchungen“ konzeptualisiert.

...daß meine Gedanken bald erlahmten, wenn ich versuchte, sie, gegen ihre natürliche Neigung, in *einer* Richtung weiterzuzwingen. - Und dies hing freilich mit der Natur der Untersuchung selbst zusammen. Sie nämlich zwingt uns, ein weites Gedankengebiet, kreuz und quer, nach allen Richtungen hin zu durchreisen. - Die philosophischen Betrachtungen dieses Buches sind gleichsam eine Menge von Landschaftsskizzen, die auf diesen langen und verwickelten Fahrten entstanden sind. (S. 231)

Läßt man derartige Metaphern als Hilfskonstruktionen zum Verständnis menschlicher Kognitionen bei der Rezeption von vernetzten Wissensstrukturen (wie z. B. Hypertexten) zu, so könnte sich wiederum die Konzeption der mentalen Modelle (Johnson-Laird, 1983; Gentner & Stevens, 1983) als ertragreiches Theoriengebäude erweisen, um die relevanten psychologischen Prozesse zu modellieren.

Spricht man von der „Theorie der mentalen Modelle“, ist prinzipiell von zwei Theoriesträngen auszugehen, die trotz der frappierenden zeitlichen Überlappung und der gleichen Bezeichnung des Kernkonstruktes klar voneinander getrennt werden können.

Obwohl die psychologische Kernannahme, daß menschliche Kognitionen vermutlich nur in den seltensten Fällen in der Form von Reiz-Reaktions-Mustern beschrieben werden können, sondern meistens eher von komplexen, aktiv konstruierten und individuell sehr unterschiedlichen „Gedankengebäuden“ auszugehen ist, in der Psychologie eine lange Tradition hat, wird der Beginn einer speziellen Ausrichtung der Forschung zu mentalen Modellen gemeinhin im Jahr 1983 angesetzt, als die beiden nahezu gleichlautenden Werke von Philip Johnson-Laird und Dedre Gentner und Albert Stevens erschienen.

Zwar kann es sicherlich nicht Ziel dieser Arbeit sein, in extenso Überlappungen und Differenzen zwischen diesen beiden Theoriesträngen herauszuarbeiten. Allerdings gibt es mindestens einen wesentlichen Aspekt in der Debatte (der auch aus weit älteren Kontroversen in der Psychologie bereits bekannt ist), der für die Konzeption und Diskussion in dieser Untersuchung von zentraler Bedeutung sein könnte. Gemeint ist die Frage, wie genau komplexe Informationen verarbeitet, abgespeichert und zum Problemlösen herangezogen werden können. Johnson-Laird u.a. vertreten hier die Auffassung, daß mentale Modelle immer das Ergebnis komplexer Verarbeitungsprozesse darstellen, die bezüglich der Verarbeitungstie-

fe, insbesondere aber in der logischen Abfolge eine Verarbeitung der Informationen in der Form propositionaler Kodierungen voraussetzt (ein entsprechender Zusammenhang konnte empirisch belegt werden, vgl. Mani & Johnson Laird, 1982). Aus diesen Voraussetzungen lassen sich konsequenterweise sehr klare Anforderungen an eine empirische Überprüfung formulieren. So sind die Untersuchungen von Johnson-Laird meist solche aus dem Bereich des Problemlösens, des logischen Schließens und der Forschung zur Psychologie der Textverarbeitung - klassischen Feldern der Allgemeinen Psychologie.

Nicht in direktem Gegensatz dazu, aber doch in wesentlichen Punkten weniger restringent sind die Modellierungen, die in der Folge und bereits im Sammelwerk von Gentner und Stevens (1983) ausgebreitet wurden.

Paradigmatisch für die unterschiedliche Konzeptualisierung des Begriffes „mental model“ erscheint in diesem Zusammenhang der Artikel „Understanding Micronesian Navigation“ (Hutchins, 1983) aus dem Sammelband - auch wenn inhaltlich der dort ausgeführte Ansatz der „social cognition“ wenig mit der hier diskutierten Thematik zu tun hat. Hier wird die Fähigkeit der Bevölkerung von Micronesien beleuchtet, ohne technische Hilfsmittel überhaupt komplexe Navigationsaufgaben zu lösen. Paradigmatisch ist dieser Beitrag in mehrfacher Weise, zunächst sticht der starke Praxisbezug direkt ins Auge, im Gegensatz zu den Laborexperimenten Johnson-Lairds wird hier direkt im Feld geforscht. Desweiteren findet man eine weniger ausgeprägte theoretische Elaboration zugunsten des dafür umso ausgeprägteren Praxisbezuges, und nicht zuletzt wird ein naheliegender Bezug auch für diese Arbeit deutlich - Navigationskompetenz. Offensichtlich werden diese Unterschiede in der Verwendung des Kernkonstruktes auch im Beitrag von Larkin (1983) zur Rolle von Problemrepräsentationen in der Physik („the role of problem representation in physics“). Die Autorin versucht, die unterschiedlichen Problemlösestrategien von Novizen und Experten bei der Bearbeitung komplexer physikalischer Aufgaben zu beschreiben, indem sie davon ausgeht, daß in beiden Gruppen unterschiedliche Arten von mentalen Repräsentationen des Problems generiert werden. Novizen generieren demnach „naive problem representation“, wohingegen die Experten mit „physical representations that contains fictitious, amagindes entities such as forces and momenta“ (Larkin 1983, S. 75) operieren. Auch hier ist es so, daß - im Vergleich zur Forschung im Rahmen des Johnson-Laird-Ansatzes - mit komplexeren, und eher im Feld verankerten Problemen gearbeitet wird. Dies hat offenbar zwei wichtige Konsequenzen: die psychologische Elaboration der zentralen Annahmen hinsichtlich der zugrundeliegenden Informationsverarbeitungsprozesse fallen sehr dürftig aus und

die experimentelle Prüfung weicht einer qualitativen Analyse von Protokollen lauten Denkens. Nichts daran ist grundsätzlich verwerflich, für die Verwendung des Konstruktes „mentale Modelle“ ist es aber durchaus angezeigt, diesen Unterschied zu berücksichtigen. Erschwerend kommt hinzu, daß in der zitierten Untersuchung von Larkin eine Konzeptualisierung der „physical representations“ verwendet wird, die aus der Sicht der Johnson-Laird-Klassifikation unter Umständen eher einer propositionalen Repräsentation des Problemraums gleichkommt, als einer in Form von mentalen Modellen. Jedenfalls wird dieser kognitionspsychologisch höchst relevante Aspekt in dieser Untersuchung nicht problematisiert und fließt demnach auch nicht in die empirische Überprüfung ein.

Bei näherer Betrachtung sind die Unterschiede in beiden Ansätzen bisweilen so gravierend, daß man den Eindruck gewinnt einen Ebenen-Fehler zu begehen, wenn man Untersuchungen aus der Johnson-Laird-Tradition mit solchen aus der Gentner-Stevens-Tradition vergleichen möchte. Für das Anliegen dieser Arbeit und die Interpretation der Ergebnisse jedenfalls ist bereits an dieser Stelle eine erhebliche Relevanz dieser Bruchstelle anzusetzen, zumal in den Aufsätzen des Gentner und Stevens-Bandes (1983) nahezu durchgängig eine Tendenz zu beobachten ist, die bereits weiter oben als problematisch in der Forschung und Theoriebildung zu Hypertext beschrieben wurde: Auch hier entsteht an vielen Stellen der Eindruck einer zu groben Generalisierung und ungenauer Korrespondenzannahmen (z. B. zur Vergleichbarkeit räumlicher Operationen im physikalischen Raum und entsprechender mentaler Operationen) hinsichtlich der angenommenen kognitionspsychologischen Konstrukte, was zusätzlich deshalb hier von besonderer Bedeutung ist, da eher die Gentner-Stevens-Forschungslinie in das Feld der Softwareergonomie und Human-Computer-Interaction Eingang gefunden hat.

Es sei noch angemerkt, daß das Verhältnis der Gentner und Stevens-Tradition zur der um Johnson-Laird in einem gänzlich anderen - aber sicherlich ebenso relevanten - wissenschaftstheoretischen Spannungsfeld diskutiert werden kann. So hat Groeben (1986) auf das prinzipielle Problem und die weitreichenden Implikationen bei der Festlegung der Einheiten psychologischer Forschung hingewiesen. Hier ist psychologiehistorisch als wichtige Anforderung an psychologische Theorien die Verabschiedung vom Molekularismus in der Einheitenfestlegung (wie er z.B. explizit von den Vertretern der Psychophysik betrieben wurde) zugunsten eher molarer, komplexer Einheiten weitgehend konsensuell. Groeben weist nun nach, daß diese wenig diskutierte Frage vermutlich keineswegs als gelöst betrach-

tet werden kann, und auch „moderne“ psychologische Theorien häufig unter einem latenten Molekularismus und einem ebensolchen Objektivismus in der Festlegung ihrer (Forschungs-)Einheiten leiden. Im Zusammenhang der vorliegenden Untersuchung und der Diskussion der verschiedenen Theorienstränge zum Konzept der mentalen Modelle kann sicherlich auch für die Gentner und Stevens-Tradition die Analyse eher komplexer Einheiten als Verständnis der kognitiven Prozesse unterstellt werden, als dies für die Forschungen um Johnson-Laird der Fall sein kann (obwohl für diese Annahme genauere Analysen notwendig wären, denn unter Umständen ist dieser Eindruck komplexerer Untersuchungseinheiten nur ein Artefakt unterschiedlicher Untersuchungsdesigns und nicht auf die theoretischen Grundannahmen zurückzuführen). Andersherum gesagt könnte man Johnson-Laird einen latenten Molekularismus in der Einheitenfestlegung bei seinem Konzept der mentalen Modelle vorwerfen (der sich auch in den Experimenten zum syllogistischen Schließen wieder spiegelt) - obwohl seine Konzeption mentaler Modelle auch bereits unter dem Eindruck unangemessener Komplexitätsreduktion bei manchen Modellierungen von Problemlöseprozessen mit propositionalen Kodierungen entstanden ist, wie das folgende Zitat zeigt:

From a psychological perspective, the treatment of propositional attitudes in the philosophical literature seems odd because it has concentrated on single isolated beliefs. People do not have such beliefs: they have systems of beliefs; they have ideologies. (Johnson-Laird, 1983, S. 436).

Leider gestattet es der Rahmen dieser Arbeit nicht, diese Fragen mehr als nur anzureißen, ein gewisses Unbehagen des Autors bezüglich dieser fundamentalen ungelösten Frage soll aber nicht verschwiegen werden - zumal die Einhaltung des Postulates der nicht von unnötigen Reduktionen belasteten Gegenstandskonzeption in der psychologischen Forschung durch die Gegenstands-Methodik-Interaktion unter Umständen auch gravierende Konsequenzen für die Planung und Durchführung entsprechender Untersuchungen haben könnte (z.B. die Ablehnung des Experiments als - primär zu wählender - Forschungsmethode, vgl. Groeben, 1986).

Trotz der offensichtlichen Unterschiede der beiden Forschungsstränge zur Theorie der mentalen Modelle sind, insbesondere im Hinblick auf die Rezeption und Anwendung der Konzepte im folgenden, aber auch in vielfältiger Hinsicht theoretisch inhaltliche Parallelen festzustellen. So geht auch Johnson-Laird (1983) von einem Zusammenhang von Analogien, Schemata und mentalen Modellen aus, und was er als physical model beschreibt ist bei-

spielsweise gut mit dem vereinbar, was Gentner und Genter 1983 als Modell der Elektrizität postuliert haben. Der wesentliche Unterschied besteht darin, in welcher Weise der Zusammenhang propositional kodierter Informationen und solcher in Form von mentalen Modellen modelliert wird.

Für die vorliegende Untersuchung soll in Anlehnung an die Argumentation von Dutke (1994) und auf der Grundlage der oben angeführten Entschärfung der Quasi-Dichotomie von Propositionen und mentalen Modellen von Newell (1990) von einer einheitlichen Theorie der mentalen Modelle ausgegangen werden, wenngleich in der Diskussion der Ergebnisse sicherlich der Faden dieser Auseinandersetzung noch einmal kurz aufgenommen werden sollte.

Aus der Sicht der Psychologie erscheint es nicht verwunderlich, daß Theoriekonzeptionen wie die der mentalen Modelle in angewandten Wissenschaften wie auch der Human-Computer-Interaction Eingang finden konnten, zumal der Computer in diesem Fall nicht nur Objekt der Forschung ist, sondern im Zusammenhang mit der „kognitiven Wende“ in der Psychologie auch häufig primär zum Versuch der Modellierung menschlicher Kognitionen verwandt wird. So ist es nur konsequent, wenn bei der Frage nach der Art der Interaktion von Personen mit Computersystemen (wie es Hypertextsysteme ja immer auch sind) solche (psychologischen) Theorien bevorzugt werden, die selbst dem Paradigma des Informationsverarbeitungsansatzes zuzuordnen sind. So finden sich z. B. in der elektronischen Bibliothek der Association for Computing Machinery (ACM) 303 Fundstellen über den Zeitraum von 1994-1999 zu Beiträgen die sich mit dem Konzept „mental models“ beschäftigen.

Norman (1991) beschreibt die Verbindung von Psychologie und Softwaredesign so:

Mental models have been represented in several ways in cognitive psychology. Scripts have been used to layout the expected series of events. Metaphors have been used to map the elements and relations from a familiar system to a less familiar one. Production rules have been used capture the knowledge that the user may have about the working of the system. In each case, the idea of a cognitive layout may be used to describe the way in which users may engage a particular model and cast a visual representation or layout of the model. Such a layout defines the way in which the user thinks about using the system and serves as a vehicle for formulating plans. (Norman, 1991)

Der Zusammenhang zwischen Skripten, Schemata, Metaphern und mentalen Modellen wird weiter unten (Abschnitt 2.1.5) noch beleuchtet werden. Zur Darstellung der Art und Weise, wie das Konzept der mentalen Modelle in die HCI-Forschung Eingang fand, ist die-

ses Zitat allerdings auch hier bereits sehr aufschlußreich, da es Norman (1991) und zahlreichen anderen Autoren/innen offensichtlich nicht um Grundlagenforschung im Bereich von Wahrnehmung und Problemlösen geht, vielmehr dient die Theorie der mentalen Modelle als „vehicle for formulating plans“.

Zwar erscheint dieser instrumentelle Umgang mit dem Konzept für die angewandte Forschung zunächst trivial, allerdings verbirgt sich hinter diesem trivialen Verlust an Elaborationstiefe auch möglicherweise ein entscheidender Punkt. Denn wenn die Ableitung von Erkenntnissen aus der grundlagenorientierten Psychologie für Anwendungen im Bereich der Softwareergonomie unter Verwertungsgesichtspunkten mit zu starken Vereinfachungen oder unhaltbaren Generalisierungen einhergeht (wenn also der unumgängliche Verlust theoretischer Elaborationstiefe nicht theoretisch kontrolliert erfolgt), kann der potentielle Erkenntnisfortschritt durch die Theorie schnell umschlagen und zu letztlich auch für die Umsetzung in der Technologie unbrauchbaren Ergebnissen führen. Ob derartige Ableitungen in Form von Untersuchungen und Programmen (z.B. zur Visualisierung) sich letztlich diesen Vorwurf machen lassen müssen, ist wiederum selbst keine triviale Frage und ein Teil des Forschungsprozesses, also auch der empirischen Bewährung.

Zur Veranschaulichung des Zusammenwirkens unterschiedlicher Forschungsfelder, aber auch um die angenommenen Wechselwirkungen zwischen Mensch und Maschine aus der Sicht eines/r HCI-Wissenschaftler/in darzustellen, hat Norman (1991) ein Prozeßmodell der Mensch-Computer-Interaktion formuliert (Abbildung 1).

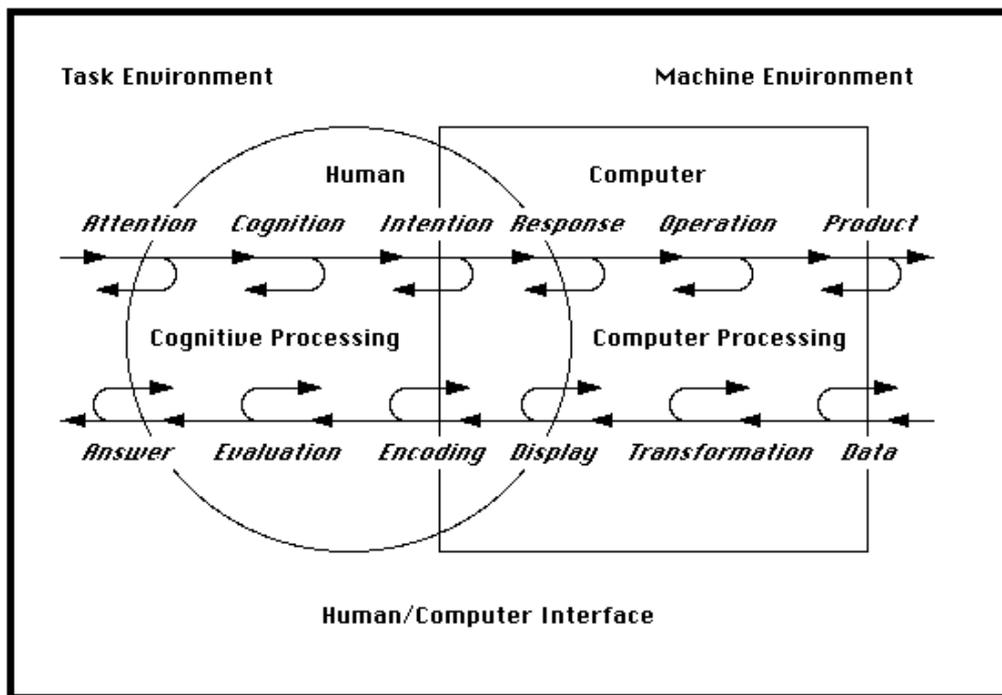


Abbildung 1: : Prozessmodell der Mensch-Computer-Interaktion nach Norman (1991)

An diesem Prozeßmodell wird deutlich, daß die HCI-Forschung ein interdisziplinäres Forschungsgebiet darstellt, welches gleichermaßen Prinzipien der Ablaufsteuerung, Visualisierung und Programmierung von Computersystemen zum Gegenstand hat, wie es eben auch originär psychologische Teilbereiche beinhaltet.

Neben der wissenschaftlich naheliegenden Verbindung von HCI und Forschungen zu mentalen Modellen weist Dutke (1994) auf einen weiteren Aspekt hin, der zur Anwendung dieser Theorie im Bereich der Softwareergonomie führte.

Die breite Verwendung von Metaphern in der Mensch-Computer-Interaktion begann mit der Ausweitung des Benutzerkreises: Personen, die mit den technischen Funktionsprinzipien von Rechnern nicht vertraut waren, sollten deren Bedienung erlernen. Metaphern halfen, bereits vertrautes Wissen zur Aneignung neuen Wissens zu nutzen. (Dutke, 1994, S. 82)

Aus diesem Zusammenhang läßt sich neben einem weiteren Grund für die Anwendung der Theorie mentaler Modelle in der HCI-Forschung ein weitaus bedeutsamerer Aspekt ableiten: Mentale Modelle eignen sich offenbar nicht nur, um komplexe Informationsverarbeitungsprozesse zu modellieren, sondern es lassen sich aus den Annahmen dieser Theorie grundlegende Prinzipien zur Konstruktion der Mensch-Maschine-Schnittstelle ableiten, mit

deren Hilfe z. B. auch klassische Probleme der *usability*-Forschung (Abschnitt 2.2.7) angegangen werden können.

Since our understanding of human perception does (or should) play a crucial role in the design of interfaces, research on mental models is a promising approach to analyzing human-computer interaction and improving interface designs. (Jih & Reeves, 1992, S. 44f)

Nach einer kurzen Abgrenzung zu konkurrierenden Theoriekonzeptionen soll im folgenden mit diesem Fokus die Theorie der mentalen Modelle unter dem Anwendungsgesichtspunkt HCI beleuchtet werden.

2.1.3. Abgrenzung zu konkurrierenden Theoriekonzeptionen

2. 1. 3. 1. Vorbemerkung zur Auswahl

Sicherlich gestattet es der hier zur Verfügung stehende Rahmen nicht, eine umfassende kritische Würdigung konkurrierender Theorien zum Konzept der mentalen Modelle vorzunehmen, allerdings erscheint es mir aus einer Reihe von Gründen dennoch angezeigt, eine kritische Abgrenzung von dem Konstrukt „information foraging“ (Pirulli & Card, 1999) zu formulieren. Strukturparallelen im kommunikations- bzw. medientheoretischen Ansatz der „uses and gratifications“ (ausführlich für den aktuellen Gegenstandsbereich angewandt in Wirth & Schweiger, 1999) sollen ebenfalls gestreift werden.

In beiden Fällen liegen im wesentlichen zwei Gründe für einen (abgrenzenden) Bezug vor. Zum einen stammen beide Theoriekomplexe aus dem großen Bereich des Informationsverarbeitungsansatzes in der Psychologie, der ähnlich wie die Theorie mentaler Modelle der Theorieentwicklung nach der vielzitierten kognitiven Wende in der Psychologie zugerechnet werden kann. Zum anderen werden beide Theorien vor allem in einem Kontext verwendet, der auch für das Thema dieser Arbeit zentral ist, nämlich der Informationssuche in computerisierten Informationssystemen, im besonderen in Hypertexten und im World-Wide-Web.

Natürlich ist diese kleine Auswahl an diskutierten Alternativkonzepten nicht repräsentativ und in keiner Weise erschöpfend für den möglichen Fundus an (psychologischen) Theorien, um Informationssuche und Navigation in Hypertexten theoretisch zu modellieren. Vor allem auf die Erörterung von (Lern-)Theorien aus dem Bereich des entdeckenden Lernens (Bruner, 1961, zit. n. Schulmeister, 1997) und des Lernens mit Mikrowelten (Papert, 1980,

zit. n. Schulmeister, 1997) soll hier ausdrücklich verzichtet werden, da nicht in gleicher Weise wie bei den zuvor genannten Theorien Abgrenzungsbedarf zu bestehen scheint und einer detaillierten Erläuterung der Zusammenhänge und Differenzen hier kein Raum gegeben werden kann.

Es sei noch vermerkt, daß eine weitere Theorienfamilie aus dem Bereich des Informationsverarbeitungsansatzes hier nicht ausführlich besprochen werden kann, und zwar handelt es sich um solche Theorien, die wie das GOMS-Modell (*Goals, Operators, Methods and selection rules*; Card, Moran & Newell; 1983), SOAR (Newell, 1990) oder ACT-R (Anderson, 1993) im engeren Sinne an das Informationsverarbeitungsmodell des Computers angelehnte Modellierungen formulieren. Diese Theorien wurden mit dem Anspruch formuliert, ein breites Spektrum der Phänomene menschlicher Kognitionen und Lernprozesse abbilden zu können und stellen somit gewissermaßen die Rahmentheorie zur Theorie der mentalen Modelle, wie sie hier angewendet wird dar. Insbesondere im Falle von SOAR wird von Newell (1990) detailliert dargestellt, inwiefern mentale Modelle als Teil von Problemlöseprozessen (nämlich Repräsentationen von Zuständen im Problemraum) der SOAR-Architektur verstanden werden können.

2. 1. 3. 2. *Information foraging (Pirolli & Card 1999)*

In einem umfassenden Beitrag haben Peter Pirolli und Stuart Card (1999) einen Theorieentwurf vorgelegt, der geeignet sein soll zu verstehen, „how strategies and technologies for information seeking, gathering, and consumption are adapted to the flux of information in the environment.“ (Pirolli & Card, 1999, S. 2).

Es handelt sich dabei um eine Ausarbeitung bzw. „an extensive revision“ des prominenten ACT-R Produktionssystems von John Anderson (1993), die von den Autoren als ACT-IF (*ACT-Information Foraging*) benannt wird. Die zentrale Hypothese der Theorie besagt, daß Individuen (bzw. „natural information systems“) nach Möglichkeit Zustände maximaler Informationsgewinnung bei Minimierung der Kosten anstreben.

Die Autoren entwickeln ein elaboriertes Modell dieses Prozesses der Suche nach Informationen, erläutern wie diese in verschiedenen Situationen an den Kontext angepasst wird und präsentieren ein Konstrukt, um die Auswahl und Bewertung von Informationen, die zu anderen räumlich benachbart sind (z.B. in einer graphischen Darstellung auf einem Computerbildschirm), zu modellieren (*information scent*). Auch die beobachtete Tendenz von

Versuchspersonen, relevante Dokumente zusätzlich einer Bewertung nach dem Aufwand für die Rezeption der gesuchten Information zu bewerten, wird mit dem Konstrukt „diet selection“ in die Theorie integriert, es werden sogar eine Reihe mathematischer Gleichungssysteme diskutiert, um diesen Prozeß optimal abbilden zu können.

Zur Integration dieser Komponenten einer Informationssuche und Rezeption präsentieren die Autoren schließlich ein Prozeßmodell der Informationssuche, das eine Abwandlung des ursprünglich von Anderson (1993) formulierten ACT-R-Modells darstellt. Die Erwartungen an diese Kombination sind nicht gering, es werden „mechanistic specifications that explain how the actions are effected“ (Pirolli & Card 1999, S. 10) erwartet.

In der Ablaufsteuerung des Modells bei der Informationssuche wird davon ausgegangen, daß mit einer Kombination von deklarativem Wissen (declarative memory), Produktionsregeln (production memory) und speziellen Vergleichs- und Bewertungsprozessen (evaluation functions), die die zuvor diskutierten Regeln des information foraging beinhalten, die Auswahl und Verarbeitung der gesuchten Informationen nach den Grundregeln der Theorie (Kostenminimierung und Informationsmaximierung) vorgenommen wird. Dabei wird das von Anderson (1993) bereits eingeführte Konzept der „spreading activation“ im semantischen Netz verwendet, um das Konzept des wahrgenommenen information-scent zu erläutern.

Prinzipiell problematisch an dieser aktuellen Theorieentwicklung scheinen m.E. jedoch nicht die Elaborationstiefe oder etwa Annahmen einzelner Operationalisierungen der Konstrukte zu sein, vielmehr ist eine prinzipielle wissenschaftstheoretische Einordnung derselben angezeigt. Diese Forderung läßt sich ausgehend von einer vermeintlichen Kuriosität der Terminologie von Pirolli und Card (1999) begründen. Diese beruht auf einer an vielen Stellen zur Erklärung von Grundannahmen herangezogenen Analogie zwischen der Nahrungssuche von Lebewesen (food-foraging z. B. von Vögeln) und der Informationssuche von Menschen (eben information-foraging). Vor dem Hintergrund der Entwicklung der Psychologie in den letzten Jahrzehnten, insbesondere der breitgefächerten Abkehr vom Modell der behavioristischen Psychologie (vgl. Groeben, Wahl, Schlee & Scheele, 1988) mit ihrer Grundannahme der Rattenähnlichkeit des Menschen ist man geneigt, diesen Kunstgriff der Autoren für ein didaktisches Experiment zu halten, das in der Substanz nicht trägt. Tatsächlich scheint es aber so zu sein, daß die Autoren in vollem Ernst von der Vergleichbarkeit dieser Situationen ausgehen, und die Analogie keineswegs nur didaktischen Charakter hat. Deutlich wird dies auch an anderen Stellen, wenn beispielsweise von „mechanistic specifi-

cations“ des Modells gesprochen wird, oder generell durch die Annahme der Berechenbarkeit des menschlichen Verhaltens in solch komplexen Situationen durch einfache mathematische Gleichungssysteme.

Letzendlich scheint diese Theorie eine Illustration verschiedentlich geäußelter Einwände gegen den ihr zugrundeliegenden Informationsverarbeitungsansatz in der Psychologie zu sein (vgl. Groeben, 1986; Groeben, Wahl, Schlee & Scheele, 1988), denn die Grundannahmen der Reduzierbarkeit selbst komplexer Verarbeitungsprozesse des Menschen auf mechanistische Regeln (bzw. solche die in Computerprogrammen zur Simulation herangezogen werden können) und Vergleiche mit Verhaltensweisen von Tieren sind Grundbestandteil des behavioristischen Ansatzes und werden somit den Annahmen des reflexiven, aktiv rezipierenden Subjekts nicht gerecht. Sicherlich kann diese (zugegebenerweise aufgrund des fehlenden Raumes etwas pauschal klingende) Kritik nicht zur Generalkritik am Informationsverarbeitungsansatz stilisiert werden. Wie die folgende Analyse zeigt, ist ACT-IF vielleicht einfach nur ein unglücklicher Vertreter dieser Theorienfamilie, und fällt in der Erklärungskraft bereits gegen im engeren Sinne konkurrierende Modelle zurück.

Die Information-Foraging-Theorie kann allerdings auch gemessen an den Anforderungen und der Erklärungskraft vergleichbarer Ansätze, wie z.B. SOAR (Newell, 1990) kritisiert werden - im besonderen im Hinblick auf Problemstellungen, wie sie bei der Informationssuche in Hypertextsystemen typisch sind. Eine wesentliche Schwäche des Ansatzes wird m.E. deutlich, wenn man versucht ein klassisches Problem der Psychologie des komplexen Problemlösens mit ACT-IF zu modellieren, und zwar komplexe Probleme in der Art des „Turms von Hanoi“ (vgl. z.B. Anderson, 1995). Analysen der Problemlösestrategien von Personen in diesem Fall haben deutlich gemacht, daß die Strategie der „Unterschiedsreduktion“ (Anderson, 1995), die darin besteht Zustände herbeizuführen, die in einer sukzessiven Annäherung des Zielzustandes bestehen, zu keiner Lösung des Problems führen kann. Tatsächlich kann das Problem nur dann gelöst werden, wenn in einer globalen Mittel-Ziel-Analyse die Strategie der Bildung von Zwischenzielen vor dem Hintergrund der Erreichung des Gesamtzieles bewertet wird. Im Rahmen dieser Strategie ist es dann durchaus möglich, kurzfristig Teilzustände zu akzeptieren, die nach der Methode der „Unterschiedsreduktion“ als Rückschritt zu bewerten wären. Wenn man nun die Definition von „information scent“ berücksichtigt, wie sie von Pirolli und Card (1999) formuliert wurde,

Information scent is the (imperfect) perception of the value, cost, or acces path of infor-

mation sources obtained from proximal cues, such as bibliographic citations, WWW links, or icons representing the sources (Pirulli & Card, 1999, S. 10)

dann wird deutlich, daß dem „Turm von Hanoi“-Problem analoge Suchaufgaben in Hypertextsystemen mit dieser Theorie nicht angemessen oder zumindest nicht ökonomisch modelliert werden können. Denn ähnlich wie mit der Strategie der „Unterschiedsreduktion“ werden nur die Bedeutung („value“ und „cost“) der aktuell im Fokus liegenden Knoten in die Entscheidung des weiteren Vorgehens einbezogen. Eine globale Problemlösestrategie hingegen, die eine Auswahl von Knoten mit niedrigem value und hohen Kosten zu Gunsten einer komplexeren Problemlösestrategie in Kauf nehmen würde, lässt sich mit der Theorie bestenfalls über komplexe Zusatzannahmen zum Konzept des „values“ eines Knotens modellieren. Für den von den ACT-IF-Autoren präferierten Vergleich der Informationssuche mit der Futtersuche im Tierreich ist diesbezüglich das Beispiel von Köhler (1927, zit. n. Anderson, 1995) illustrativ, der von einem Huhn berichtet, das nicht in der Lage ist, Futter hinter einem Zaun durch Inkaufnahme einer kurzfristigen Entfernung (um den Zaun zu umrunden) von diesem zu erreichen, und infolgedessen hilflos vor dem Hindernis stehen bleibt.

2. 1. 3. 3. *Uses and Gratifications* (Wirth & Schweiger, 1999)

Nur kurz streifen möchte ich den im Sammelwerk von Wirth und Schweiger (1999) in einem Beitrag für die Interpretation von Selektionsentscheidungen bei der Informationssuche im Internet angepriesenen Ansatz der „uses and gratifications“ (Tasche, 1999), der vor allem in der Kommunikations- und Medientheorie Anwendung findet, und sich bezüglich der wissenschaftstheoretischen Einordnung wenig von dem ebenfalls von diesem Autor als brauchbare Theorie vorgeschlagenen Ansatz der Stimmungsregulationstheorie (Zillmann & Bryant, 1985, zit. n. Tasche, 1999) zu unterscheiden scheint. Da es sich bei beiden Ansätzen um psychologisch wenig elaborierte Theorien zu handeln scheint, reicht an dieser Stelle der Hinweis, daß beide Theorien auf ähnlichen Kernannahmen bzgl. der menschlichen Kognition beruhen, und insofern in vollem Umfang die oben angeführte Kritik hinsichtlich der wissenschaftstheoretischen Einordnung greift.

2.1.4. Allgemeine Modelltheorie

Stachowiak (1973) unterscheidet in seiner „Allgemeinen Modelltheorie“ drei zentrale Merkmale von Modellen, und zwar das „Abbildungsmerkmal“, das „Verkürzungsmerk-

mal“ und das „pragmatische Merkmal“. Demnach ist es charakteristisch für Modelle, daß sie einen Gegenstandsbereich zwar abbilden, aber nicht vollständig, sondern unter pragmatischer Perspektive verkürzt und auf für einen Anwendungsfall relevante Informationen konzentriert. Der einschränkende Bezug kann im Hinblick auf bestimmte Individuen, Zeitpunkte und auf bestimmte Zwecke spezifiziert werden.

Obwohl diese allgemeine Modelltheorie zunächst nichts mit der Theorie der mentalen Modelle zu tun hat, ergeben sich bereits zwei wichtige Ansatzpunkte, um den Individuenbezug mentaler Modelle und deren charakteristische Unvollständigkeit abzubilden.

Die „Unvollständigkeit“ von Modellen im Vergleich mit dem abgebildeten Original ist ein wesentlicher Bestandteil derselben, zumal die Bildung von Modellen häufig kommunikativen Zwecken (z. B. Schulung von Ingenieuren) dient und insofern aus didaktischen und ökonomischen Zwecken nicht alle Informationen des Originals interessieren.

In diesem Fall können Abweichungen hinsichtlich zweier relevanter Dimensionen unterschieden werden: Es gibt Abweichungen in den Strukturrelationen und/oder im Hinblick auf die Elemente des Modells. Zur Abgrenzung beider Extremfälle spricht Stachowiak (1973) im Falle maximaler struktureller Ähnlichkeit von Isomorphie, und im Falle maximaler materialer Ähnlichkeit von Isohylie. Die für das Folgende wichtige Form der Analogie ist demnach der Fall einer hohen strukturellen und einer niedrigen materialen Ähnlichkeit von Original und Abbildung (Dutke, 1994).

Oberquelle (1984, zit. n. Dutke, 1994) erweitert die Modell-Definition um den relevanten Aspekt der aktiven Konstruktion des Modells sowohl auf Seite des Rezipienten als auch auf der des Produzenten eines Modells. Dieser Aspekt wird auch von Norman (1983) eingeführt und bereits explizit auf den Bereich Softwareergonomie angewendet. Er unterscheidet zwischen dem konzeptuellen Modell („conceptual model“), dem Ziel-System, das erlernt werden soll („target-system“), dem mentalen Modell des Benutzers/der Benutzerin und dem Konzept des Wissenschaftlers/der Wissenschaftlerin über das mentale Modell des Benutzers/der Benutzerin (Norman 1983, S. 7). Das konzeptuelle Modell ist dabei die explizite, vollständige und genaue Beschreibung des Zielsystems (also sowohl isomorph als auch isohyl). Das Konzept des Wissenschaftlers/der Wissenschaftlerin (oder Lehrers/Lehrerin o.ä.) des mentalen Modells des Users ist wiederum selbst ein mentales Modell.

2.1.5. Eigenschaften und Funktionen mentaler Modelle

Die Kernkonstrukte der Theorie(n) der mentalen Modelle können in Anlehnung an diese allgemeine Modelltheorie erläutert werden. Dieses kognitionspsychologische Konstrukt steht dabei im Rahmen der theoretischen Entwicklung, die versucht den Menschen als kognitiv aktiv und selbstreflexiv zu modellieren. Mentale Modelle sind also ähnlich wie z. B. Schemata, Skripte und Chunks theoretische Konstrukte, um bestimmte Aspekte der aktiven Informationsverarbeitung begreifbar zu machen. Sie setzen allerdings sehr komplexe und teilweise hoch individuelle Verarbeitungsprozesse voraus und stellen sicherlich ein empirisch vergleichsweise schwer zugängliches, aber für komplexere Verarbeitungsprozesse auch sehr fruchtbares Konstrukt dar.

2. 1. 5. 1. Unvollständigkeit

Unter 2.1.4 wurde bereits in der allgemeinen Modelltheorie die Unvollständigkeit als zentrales Merkmal eines Modells in Bezug auf das Original eingeführt. Bei mentalen Modellen muß diese Unvollständigkeit dahingehend erweitert werden, daß mentale Modelle per definitionem nicht explizit formuliert sind, sondern hypothetische Konstruktionen über kognitive Operationen darstellen. Es dürfte kein Zweifel daran bestehen, daß auch mentale Modelle kein 1:1-Abbild des Originals darstellen, allerdings ist die Bewertung und Klassifikation der jeweiligen Abbild-Original-Beziehung weitaus komplexerer Natur als es bei expliziten Modellen die Regel ist. Dies ist deshalb der Fall, da mentale Modelle „als kognitive Konstruktionen aufgefaßt [werden], die auf einer Interaktion von Wahrnehmung und Gedächtnis beruhen“. (Dutke, 1994, S. 12).

Wenn man nun von der in der heutigen Psychologie im Prinzip unbestrittenen Annahme der kognitiven Konstruktivität des Individuums ausgeht, wird klar, daß die beiden Determinanten der Bildung eines mentalen Modells, Wahrnehmung und Gedächtnis in höchstem Maße als individuelle Konstruktionen des jeweiligen Rezipienten/der jeweiligen Rezipientin zu verstehen sind und nur noch wenig mit den bewußt verkürzten Modellen expliziter Natur zu tun haben. Dennoch scheinen aber die Annahmen einer Strukturparallelität des Vorgehens beim Bilden von expliziten und impliziten (mental) Modellen hilfreich zu sein, um den Umgang von Personen mit komplexen technischen Systemen zu verstehen.

Norman (1983) beschreibt eine Untersuchung mit Probanden/Probandinnen, die Rechenaufgaben mit einem einfachen Taschenrechner lösen sollten. Dabei zitiert er die Äußerungen einer Probandin, die aus technischer Sicht sehr ineffektiv mit dem Gerät umging, da

sie z. B. die Möglichkeit, Zwischenergebnisse im Taschenrechner zu speichern, nicht wahrnahm, sondern diese auf einem Papier notierte. Ebenso betätigte sie unnötig (wie gesagt, aus technischer Sicht) oft die „Clear“-Taste des Rechners. Im Interview erklärte die Probandin ihr Vorgehen mit Unsicherheit gegenüber der Funktionsweise des Rechners. Interessanterweise hatte sie aber aufgrund des inadäquaten Modells von der Funktionsweise des Gerätes offensichtlich eine erfolgreiche Strategie entwickelt, um die Aufgaben zu lösen und ihre Unsicherheit zu verringern. Aus der Sicht der allgemeinen Modelltheorie wäre nur festzustellen, daß diese Person (wie übrigens die meisten anderen in der Untersuchung auch) ein inadäquates Modell vom Original gebildet hat, und didaktische Unterweisung nötig wäre, um diesen Mangel zu beheben.

Im Gegensatz dazu stellt Norman (1983) aber fest, „it seems a sensible simplification that eases and generalizes what would otherwise be a more complex, machine specific set of knowledge“ (Norman 1983, S. 10), denn die aus dem scheinbar inadäquaten mentalen Modell des Rechners abgeleitete Strategie, die Clear-Taste mehrfach zu drücken, war tatsächlich eine ökonomische (weil keine differenzierte Anzahl von benötigten Tastendrücken in verschiedenen Situationen gespeichert werden mußte) und auf andere Geräte generalisierbare Regel, die zudem mit großer Präzision funktionierte.

Dutke (1994) verweist auf einen weiteren Aspekt, der sich aus dem obigen Beispiel ebenfalls gut erklären lässt: mentale Modelle scheinen relativ resistent gegenüber Änderungen zu sein bzw. es kann sein, „daß eine Vermehrung einfacher und gut beherrschter Handlungen einer Elaboration des mentalen Modells vorgezogen werden“ (Dutke 1994, S. 13).

2. 1. 5. 2. Analogien, Metaphern

In Abschnitt 2.1.4 wurde bereits darauf hingewiesen, daß die Analogie einen speziellen Fall einer Abbild-Original-Beziehung darstellt, und zwar liegt hier eine tendenziell isomorphe Struktur vor.

Gentner und Gentner (1983; ebenso Gentner, 1983) führen in ihrem Beitrag die Analogie als zentrales Element in der Konstruktion von mentalen Modellen ein. Wesentlich für den didaktischen Effekt ist dabei der aus der allgemeinen Modelltheorie abgeleitete Umstand, daß in einer Analogie (anders als z. B. in einer Ähnlichkeitsbeziehung) nur strukturelle Überlappungen zwischen dem Zielbereich und der Analogie bestehen, nicht aber inhaltliche. Dadurch ist es möglich, bekannte Relationen (z. B. die Zusammenhänge der Planeten in einem Sonnensystem) auf zunächst unbekannte Bereiche (z. B. den Aufbau von Atomen)

zu übertragen. Die Autoren/innen bezeichnen diesen Vorgang als „structure mapping“ und präsentieren zwei Experimente, in denen verschiedene Analogien für die Zusammenhänge in der Elektrizitätslehre gelehrt wurden. Anschließend sollten die Versuchspersonen typische Problemkonstellationen aus dieser Domäne beantworten, wobei in Abhängigkeit vom gelehrt Modell (der Analogie) unterschiedliche Leistungen bei bestimmten Problemen erwartet wurden. Die erwarteten Effekte konnten empirisch gesichert werden. In einem weiteren Experiment der Autoren/Autorinnen war dies nicht der Fall, allerdings äußern sie die Vermutung, daß dies mit der unvollständigen Repräsentation der Domäne des Basisbereichs (aus dem die Analogie gebildet werden sollte) zusammenhing.

Eine weitere Variante der Abbild-Original-Beziehung, die nicht mehr ohne weiteres aus der allgemeinen Modelltheorie abzuleiten ist, stellt die Metapher dar. Nach Dutke (1994) ist eine Metapher „ein didaktisches Mittel, auf eine Analogiebeziehung hinzuweisen“ (S.19). Entscheidend ist dabei, ob die beiden Objekte der Metapher und des Zielbereichs über Attribute beschrieben werden können, die unterschiedliche Bedeutungsvarianten zulassen. So funktioniert die Metapher in der Formulierung „Er wurde zu Eis“ z. B. nur dann, wenn das Attribut „kalt“ sowohl zur Beschreibung der physikalischen Temperatur als auch zur Beschreibung von Persönlichkeitsmerkmalen zulässig ist. Ist dies aber der Fall, kann die Metapher dazu dienen, eine Strukturübertragung durch die Analogie von Wasser zu Zuständen einer Person herbeizuführen, die in diesem Fall durch die Analogie selbst nicht gut zu vermitteln gewesen wäre.

Die Wirksamkeit einer Metapher ist davon abhängig, ob sie als solche erkannt wird, was zunächst umso besser gelingt, je ähnlicher sich die Subjekte des Vergleichs- und des Zielbereichs sind. Allerdings besteht die motivationale und didaktische Wirkung von Metaphern gerade im Kern in der Ungewöhnlichkeit des Vergleichs, also einer möglichst großen Entfernung der Subjekte aus der Metapher und dem Zielbereich. Dutke (Dutke 1994) kommt zu dem Schluß, daß für den Gebrauch von Metaphern in diesem Zusammenhang einer mittlere Ähnlichkeit der Subjekte aus dem Basis- und dem Zielbereich anzustreben ist. Zusätzlich können Metaphern noch hinsichtlich der Anzahl an übertragbaren Relationen aus der Metapher (bzw. genauer: aus der Analogie, auf die die Metapher hinweist) auf den Zielbereich klassifiziert werden; eine große Anzahl spricht für die Nützlichkeit der Metapher. Daher ist natürlich eine möglichst hohe Ausprägung anzustreben. Allerdings ist es offensichtlich, daß es nicht immer einfach sein dürfte, eine derartige Metapher zu finden.

So schildert z. B. Johnson (1999) in seinem Buch „Interface Culture“ eine Reihe von Fällen, bei denen die Wahl einer geeigneten System-Metapher in einer Weise die Umgangsweise mit einer komplexen Maschine und ihre Verständlichkeit beförderten, wie es die Entwickler/Entwicklerinnen selbst nicht für möglich gehalten hätten. Die prominentesten Beispiele sind sicherlich die Entwicklung der Smalltalk-Umgebung in den Forschungslabors der Xerox-Company, die als Grundlage heutiger graphischer Betriebssystemoberflächen gelten kann, sowie die gleichzeitige Entwicklung der „Mouse“ als neuartiges Eingabegerät. Interessant ist unter diesen Beispielen auch die Analyse eines Versuchs der Entwickler/Entwicklerinnen des graphischen Betriebssystems „Windows“, welche versuchten, den PC mit einer noch vertrauteren Metapher („Bob“) breiteren Käufergruppen zugänglich zu machen und damit scheiterten - vermutlich weil die gewählte Metapher eines Wohnzimmers „nicht metaphorisch genug waren“ (Johnson 1999, S. 71).

Einen wesentlich elaborierteren Ansatz zur Analyse und Interpretation von Metaphern, Analogien und Ähnlichkeitsbeziehungen stellt Ortony (1979) in einem grundlegenden Beitrag mit dem Titel „Beyond Literal Similarity“ dar. In diesem Artikel wird der Versuch unternommen, die üblicherweise komplexe und häufig unscharf verwendete Klassifikation von Metaphern, Analogien und Ähnlichkeitsbeziehungen zu konkretisieren, indem alle drei Phänomene auf unterschiedliche Realisationen von jeweils zugrundeliegenden Vergleichen zurückgeführt werden. Um diese unterschiedlichen Realisationen zu analysieren, werden die Attribute des Objekts im Zielbereich (welches erklärt werden soll) und die des Objektes aus dem Quellbereich (welches an vertrautes Wissen anknüpfen soll und der Erklärung dient) aufgelistet und nach relativer Wichtigkeit (relativ, weil der Kontext einen erheblichen Einfluß auf diesen Faktor ausüben kann) sortiert. Ortony weist nun nach, daß das Ausmaß der wahrgenommenen Metaphorizität (vs. „wörtlicher“ Ähnlichkeit) davon abhängt, daß ein Ungleichgewicht der relativen Wichtigkeit der für den Vergleich herangezogenen Attribute besteht („salience imbalance“), und zwar genau derart, daß wichtige Attribute des Quellbereichs („vehicle“) auf unwichtige Attribute des Zielbereiches angewandt werden können. Besteht anstelle dieser Diagonalizität der Attributbeziehungen hingegen eine balancierte Struktur von Attributen hoher Wichtigkeit für beide Objekte, so wird der Vergleich als Ähnlichkeitsstatement und nicht als Metapher wahrgenommen. Für den Fall, daß zwar eine balancierte Struktur besteht, allerdings nur Attribute geringer (relativer) Wichtigkeit aus dem Quell- und Zielbereich in den Vergleich einbezogen werden, spricht Ortony

von einem „anomalous statement“. Diese Art von Statements bzw. Vergleichen erfüllen nur geringe kommunikative Funktionen und sind schwer verständlich. Eine Quelle für die Verstärkung dieses Effektes sind Fälle, in denen die verglichenen Attribute nicht als identisch, sondern nur als ähnlich wahrgenommen werden. Desweiteren Fälle, in denen eine Inkompatibilität der dem Vergleich zugrundeliegenden Wissens-Domänen vorliegt. In solchen Fällen sind zusätzliche (rekursive) Mechanismen notwendig, um Vergleiche auf zusätzlichen Ebenen mit höherem Abstraktionsniveau zu realisieren. So ist z.B. zum Verständnis des Statements „Blutgefäße sind wie Aquädukte“ eine solcher Prozeß notwendig, um feststellen zu können, daß in beiden Fällen das Attribut „fließt durch (Flüssigkeit)“ anwendbar ist. Dieser Prozeß findet, ausgehend von den Attributen „fließt durch (Wasser)“ und „fließt durch (Blut)“ anhand der zusätzlichen Abstraktion „ist Flüssigkeit (Wasser)“ und „ist Flüssigkeit (Blut)“ eine Ähnlichkeit zweiter Ordnung. Sind derartige Prozesse notwendig, um den zugrundeliegenden Vergleich zu verstehen, nimmt Ortony eine Verstärkung des Eindrucks der Metaphorizität des Statements (bzw. der „salience imbalance“) an.

Analogien werden in dieser Theorie als ein Spezialfall einer Ähnlichkeitsbeziehung angesehen, wobei der Unterschied darin besteht, daß „in standard analogy problems, part of the problem is to construct a schema that involves one of the pairs of concepts in some central way in such a manner that the relation between them can be applied to the other side of the 'equation'“ (Orntony 1979, S. 176). Abgesehen von diesem Unterschied, sind Analogien nach Ortony aber dadurch gekennzeichnet, daß sie - wie Ähnlichkeitsbeziehungen - durch eine balancierte Struktur in der relativen Wichtigkeit der Attribute des Vergleiches konstituiert werden.

2. 1. 5. 3. Schemata

Die Bedeutsamkeit schematischen Wissens für die Rezeption neuer Informationen ist in der Psychologie seit langem unbestritten (vgl. Anderson, 1995; Taylor, 1990).

Schemata sind „Strukturen allgemeinen Wissens, die typische Zusammenhänge eines Realitätsbereichs enthalten“ (Dutke, 1994, S. 24). Dutke (1994) und Ortony (1979) weisen nun darauf hin, daß Schemata über einen Wissensbereich notwendige Voraussetzungen für die Prozesse beim Erkennen von Analogien und Metaphern darstellen. Da diese nur dann erkannt werden können, wenn Phänomene des Basis- und des Zielbereiches auf schematische Bestandteile reduziert werden, geht Dutke davon aus, daß eine Strukturübertragung nur dann stattfindet, „wenn ein Gedächtnisschema aktiviert werden kann, aus dem sowohl der

Sachverhalt des Zielbereichs als auch der des Basisbereichs abgeleitet werden kann....Beide Sachverhalte müssen also gültige Instantiierungen des gleichen Schemas sein“. (Dutke 1994, S. 27).

Eine teilweise vergleichbare Annahme machen Jih und Reeves (1992), wenn sie unterstellen daß mentale Modelle entweder in einem „top-down“ oder „bottom-up approach“ ausgebildet werden. Nur teilweise vergleichbar ist diese Annahme deshalb, weil zusätzlich zum top-down auch ein induktiver bottom-up-Prozeß angenommen wird, und weil es bei Jih und Reeves (1992) im Prinzip keinen expliziten Verweis auf Schemata gibt. Die Autoren/innen sprechen nur von relevantem Vorwissen („prior knowledge“), das über die Art des Prozesses entscheiden kann.

2. 1. 5. 4. Komplexes Problemlösen, Schlußfolgerndes Denken

Auch für den Bereich des komplexen Problemlösens findet die Theorie der mentalen Modelle Anwendung, vor allem in der häufig gegen eine formallogische, strukturorientierte Modellierung menschlicher Kognitionsprozesse beim Lösen von Problemen argumentierenden Richtung wird häufig der Versuch unternommen, mit Schemata und mentalen Modellen den empirisch gesicherten Befund zu erklären, daß logische Probleme offensichtlich nicht unabhängig vom einkleidenden Kontext gelöst werden (Dutke, 1994; Green, 1996). Wenn auch mit anderer Terminologie, so konnte doch in der Sache auch für die Denkpsychologie ein Zusammenhang von vorhandenem Wissen (deklaratives und heuristisches prozedurales Wissen) und der Lösung von Problemen gezeigt werden (Hussy, 1993).

So spricht also einiges dafür, daß auch beim Lösen komplexer Probleme schematisches Wissen und dessen Instantiierungen in Form eines jeweiligen mentalen Modells eine Rolle spielen.

Ein weiterer Hinweis auf die Bedeutsamkeit von mentalen Modellen für diese Art kognitiver Operationen ergibt sich, wenn man deren wichtige Funktion einer Komplexitätsreduktion in Rechnung stellt (Dutke, 1994; Thüring, Hannemann & Haake, 1995). Schließlich bestehen viele der experimentell bestätigten Problemlösestrategien (Anderson, 1995; Newell, 1990; Hussy, 1993) im Umgang mit komplexen Problemen z.T. aus genau solchen Strategien der Reduktion und Organisation des vorhandenen Wissens und der zur Verfügung stehenden Informationen, wie sie auch von vielen Autoren/Autorinnen im Zusammenhang mit den Funktionen von mentalen Modellen bei der Verarbeitung von komplexen Hypertexten diskutiert werden (Jih & Reeves, 1992; Thüring, Hannemann & Haake, 1995).

2.1.6. Mentale Modelle und Navigation in Hypertexten

Auch wenn in frühen Untersuchungen zu mentalen Modellen bereits Studien zum Navigationsverhalten von Menschen zu finden sind (Hutchins, 1983) und auch Fragen der Human-Computer-Interaction bereits sehr früh beleuchtet wurden (Young, 1983), ist es dennoch nicht selbstverständlich, Navigation in Hypertextsystemen als einen Anwendungsfall für die Theorie zu postulieren. Denn anders als in den genannten Untersuchungen, in denen das Verständnis komplexer Systeme (wie z. B. ein Taschenrechner oder ein Textverarbeitungsprogramm) untersucht wurde, handelt es sich bei Hypertexten zunächst nicht um vergleichbar komplexe Systeme, und es ist auch in der Regel keine komplizierte Abfolge inhaltlich unterschiedlicher Schritte erforderlich, um einen Hypertext zu rezipieren. Formal betrachtet, handelt es sich bei Hypertexten tatsächlich um eine Sammlung gleichartiger autonomer Texteinheiten, die über ebenfalls (formal) gleichartige Links verbunden werden (Abschnitt 2.2.2.1). Sicherlich ist diese rein formale Klassifikation nicht ausreichend, doch bei einer inhaltlich erschöpfenderen Beschreibung von Hypertexten ist es zwar kein Problem mehr, auch hier komplexe Strukturen zu entdecken, die z.T. ein erhebliches Maß an kognitiver Aktivität beim Rezipienten erfordern, allerdings werden diese Strukturen weniger durch technische Merkmale des Hypertexts konstituiert, sondern vielmehr durch semantische, also die Bedeutung von Textbestandteilen (und ggf. anderen Medienelementen).

Zentral für die psychologische Beurteilung der Navigation in Hypertexten erscheint mir dabei eine Unterscheidung zu sein, die in vielen Studien nicht beleuchtet wird: Wenn die zu navigierende Struktur des Hypertextes in erster Linie durch semantische Informationen konstituiert wird, dann ist es sehr wahrscheinlich, daß die daraus resultierenden Hierarchien nicht mit denen des Filesystems eines Hypertexts, also der autonomen Textbestandteile untereinander, übereinstimmt. Denn es ist psychologisch unplausibel, Textverarbeitungsprozesse nur für die Ebene der Hypertext-Knoten anzunehmen, nicht jedoch innerhalb der Knoten, bis hinunter zur Satz- und Wortebene. Dieser Einwand steht im übrigen in direktem Zusammenhang mit der Zurückweisung der Hypothese der Korrespondenz, bzw. kognitiven Plausibilität von Hypertexten (vgl. 2.2.2), da auch hier die Verarbeitung innerhalb von Knoten außer Acht gelassen wurde.

Für die Beurteilung des Problems der Navigation in Hypertexten, aber auch im besonderen für die Konstruktion von Hilfsmitteln und Formulierung von Designprinzipien, die aus dieser Beurteilung resultieren können, wird es von großer Bedeutung sein, ob Navigation in Hypertexten auf Knotenebene den Kern des Problems darstellt oder nicht.

Unabhängig von derartigen prinzipiellen Einwänden gibt es bereits zahlreiche Untersuchungen, die versuchen, Navigationsprobleme mit „incorrect user models“ (Cockburn & Jones, 1996) in Verbindung zu bringen und aus dieser theoretischen Rahmenkonzeption Ansätze abzuleiten, wie die Navigation verbessert werden könnte (z. B. McDonald & Stevenson, 1996; Wright & Lickorish, 1990).

Jih und Reeves (1992) haben eine Reihe von Fragen formuliert, die eine Untersuchung des Zusammenhangs von mentalen Modellen und der Navigation in Hypertexten zu beantworten versuchen sollte:

1. What are the properties of the mental models that learners form of hypertext-bases ILS [Interactive Learning Systems] interfaces and how frequently do they occur?
2. What individual differences in selected learning styles and computer experience are related to differences in the mental models learners form?
3. How are differences in learning styles, previous computer experience, and mental models related to the navigational pathways of learners in ILS?
4. How are the individual differences, mental models, and/or navigational pathways of learners related to the prediction of learning? (Jih & Reeves 1992, S. 48)

Die vorliegende Untersuchung stellt einen weiteren Versuch dar, einer Beantwortung dieser Fragen etwas näher zu kommen, wenngleich aus verschiedenen untersuchungstechnischen und theoretisch-methodologischen Gründen (vgl. Abschnitt 4.2) ein möglicher Rückschluß auf Merkmale des je zugrundeliegenden mentalen Modells als problematisch gelten dürfte.

2.1.7. Zusammenhang von *spatial ability* und Navigationskompetenz

Viele Untersuchungen im Bereich der Intelligenzdiagnostik, insbesondere solche, die sich auf das Intelligenzstrukturmodelle von Thurstone (1938) beziehen, werten den Faktor des räumlichen Vorstellungsvermögens als eigenständiges Fähigkeitskonzept, welches einer differenzierten Erfassung zugänglich zu sein scheint.

Da die Navigation durch komplexe Informationsmengen am Computer-Bildschirm in vergleichbarer Weise Operationen räumlicher Kognition erfordert, wie es z. B. viele Items in entsprechenden Test zu dieser Fähigkeit verlangen, ist es naheliegend, den Faktor *spatial ability* bzw. räumliches Vorstellungsvermögen im Zusammenhang mit interindividuellen

Unterschieden bei der Navigation in Hypertexten zu betrachten (Bringelson & Eberts, 1990; Chen & Czerwinski, 1997; Chen & Rada, 1996).

Offen bleibt dabei zunächst die Frage der Repräsentation des erlernten Wissens, denn selbst bei der Annahme einer ausschließlich propositionalen Repräsentation (wie z. B. bei Anderson, 1993; Pylyshyn, 1973) wären dennoch räumliche Operationen notwendig, um Informationselemente auf dem Bildschirm zu lokalisieren und je nach Visualisierungstechnik auch zu verändern, und um auf diese Weise Zusammenhänge erkennen zu können. Entschieden man sich allerdings zusätzlich für die Annahme anderer Repräsentationsformen (vgl. Kosslyn & Pomerantz, 1977; Paivio, 1971; Johnson-Laird, 1983), wie dies im Zusammenhang mit mentalen Modellen in der Regel der Fall ist, bekommt der Faktor der räumlichen Kognition natürlich noch weitaus größere Bedeutung.

Entsprechende Untersuchungen können an eine lange Tradition zahlreicher Versuche mit mentaler Rotation und inneren „Landkarten“ in der Psychologie anknüpfen, wie sie im Zusammenhang mit der oben angedeuteten Kontroverse zur Frage der Repräsentation räumlichen Wissens entstanden ist (vgl. Anderson, 1995).

In einer Studie zur mentalen Repräsentation räumlicher Beschreibungen (Mani & Johnson-Laird, 1982) konnten die Autoren/Autorinnen z. B. zeigen, daß räumliche Anordnungen von Gegenständen bei unklaren verbalen Beschreibungen schlechter erinnert werden konnten, als solche, die die Generierung einer eindeutigen analogen Vorstellung zuließen. Diese Unterschiede lassen sich mit einer propositionalen Kodierung nicht erklären. Zugleich zeigte sich ein zweiter Effekt, der durch eine weitere Untersuchung gesichert werden konnte: die Erinnerung an unklare räumliche Szenarien war zwar schlechter, allerdings konnten sich die Versuchspersonen in diesem Experiment besser an den Wortlaut der verbalen Beschreibungen dieser Szenarien erinnern als an den der eindeutigen Szenarien. Die Autoren/Autorinnen erklären diesen Effekt mit der Annahme zweier Repräsentationsformen bzw. zweier Formen der Abspeicherung von Informationen. Demnach werden verbale Informationen zunächst in einen propositionalen Code übertragen, um dann nach Möglichkeit eine elaboriertere Version in der Art eines mentalen Modells abzulegen. Die Ergebnisse legen zudem nahe, daß bei erfolgreicher Bildung einer solchen Repräsentation in der Form eines mentalen Modells die zuvor gebildete propositionale Kodierung vergessen wird.

Für die Visualisierung komplexer Informationen am Bildschirm dürften diese Erkenntnisse weitreichende Konsequenzen haben, obgleich bei der anhaltend kontroversen Diskussion dieser Kardinalfrage der Psychologie nicht vorschnell aufgrund einzelner Untersuchungen

theoriegeleitet Entscheidungen getroffen werden sollten. Insofern ist die vorliegende Untersuchung eher als ein weiterer Versuch zu verstehen, weitere empirische Belege für die Vorteile der Bildung eines mentalen Modells zu finden.

Jedenfalls ist nicht nur aufgrund zahlreicher empirischer Belege (Campagnoni & Ehrlich, 1989; Chen & Czerwinski, 1997, Vicente & Williges, 1988), sondern auch aufgrund dieser Überlegungen zum konzeptuellen Zusammenhang der Konstrukte *spatial ability* und „mentales Modell“ von einem wesentlichen Einfluß des genannten Fähigkeitskonstrukts im Zusammenhang dieser Untersuchung auszugehen (Chun & Plass, 1997; Walker, Meyer & Smelcer, 1993; Veer & Wijk, 1990), da bei der Visualisierung komplexer Informationen und der Rezipierung graphischer Navigationshilfen mit großer Wahrscheinlichkeit kognitive Prozesse relevant sind, die vom Gültigkeitsbereich des Konstruktes *spatial ability* tangiert werden.

2.2 Hypertext im Blickpunkt der Human-Computer-Interaction

2.2.1. Vorstrukturierung

Im folgenden Kapitel wird der Texttyp Hypertext und hypermediale Lernsysteme definiert und eine Reihe klassischer Probleme im Zusammenhang mit diesen Medien sollen ausgeführt werden. Ein besonderer Schwerpunkt liegt dabei auf der Einschätzung häufig euphorischer Annahmen bezüglich der Chancen vernetzter Hypertexte für das Lernen und die Aufnahme von Informationen. Des weiteren werden Navigationsprobleme diskutiert und kritisch gegen empirische Befunde abgewägt.

Hypertext werden dabei als Anwendungsfall einer Mensch-Maschine-Schnittstelle angesehen, wodurch die Verbindung zur HCI-Forschung deutlich gemacht werden soll. In einem weiteren Schritt werden dann klassische Probleme der HCI-Forschung, insbesondere im Zusammenhang mit dem Design von Programmsteuerungselementen (z.B. Menüsysteme) diskutiert und deren strukturelle Verwandtheit mit Problemen der Hypertextforschung analysiert. Dabei sollen auch bereits diverse Lösungsversuche in Form von Programmen zur Visualisierung von Daten vorgestellt und anhand der diskutierten Probleme eingeschätzt werden. Abschließend erfolgt eine Darstellung der *usability*-Forschung und deren zentraler Zielideen und Forschungsmethoden.

2.2.2. Hypertext und hypermediale Lernsysteme

2. 2. 2. 1. Grundprinzip: Worin unterscheiden sich Hypertexte von linearen Texten?

Auch wenn in der Literatur noch kein Konsens für eine allgemeingültige Definition von Hypertexten aufzufinden ist (Gerdes, 1997a; Schulmeister, 1997, Shneiderman, 1998), lassen sich doch relativ leicht anhand prototypischer Merkmale von Hypertexten die relevanten Abgrenzungen zu Lineartexten finden.

Hypertexte bestehen aus autonomen Informationseinheiten („Knoten“) die anhand von Verweisen („Links“) miteinander beliebig verknüpft werden können. Die zentrale Eigenschaft des Hypertext ist demzufolge seine Nicht-Linearität. Im Gegensatz zu linearen Texten besteht keine streng-hierarchische Gliederung der Informationen, und es wird der Leser/die Leserin keine bestimmte Leserichtung vorgegeben. Vielmehr kann der Leser/die Leserin nach eigenem Bedarf 'Pfade' durch den Hypertext legen und die eigenen Wissensbestände so mit einem flexiblen System sukzessive ergänzen. Damit ist der (prototypische) Hypertext auch als 'interaktiv' zu klassifizieren.

Die Knoten des Hypertextes werden in einem speziellen Betrachtungsprogramm („Browser“) aufgerufen und am Bildschirm dargestellt. Unabhängig von den Möglichkeiten zur Navigation, die der Autor/die Autorin des Hypertextes vorgesehen hat, bieten diese „Browser“ bestimmte Standardnavigationselemente an, und zwar das Vor- und Zurückblättern, das Aufrufen der Startseite und in je nach Hersteller unterschiedlicher Ausführung einen Zugriff auf bereits besuchte Seiten („history“).

Nach Gerdes (1997a) lassen sich Hypertexte zusätzlich nach der Art ihrer Vernetzung klassifizieren. So gibt es linear, hierarchisch und vernetzt organisierte Hypertexte.

Aus der Sicht der Softwareergonomie und im Hinblick auf die Verarbeitung von Hypertexten sind zwei Probleme der Gestaltung von Hypertexten besonders relevant:

Die Knotenlänge: Hier stellt sich die Frage, in welcher Weise Anforderungen an das Arbeitsgedächtnis der LeserIn berücksichtigt werden, und ob es gelingt, den Text so zu gestalten, daß jeder Knoten maximal eine Seite füllt. Andernfalls muß der Bildschirm nach unten „gescrollt“ werden, was sich als problematisch erweisen könnte.

Darstellung und Bezeichnung der Links: In der Regel werden Links am Bildschirm optisch durch blaue Unterstreichung ausgezeichnet. Zusätzlich verändert sich die Form des Mauszeigers, wenn dieser über einen Link im Text bewegt wird. Dies kann bei einer großen Anzahl von Verweisen zu einem sehr uneinheitlichen typographischen Bild führen. Andererseits ist es wichtig, „anklickbare“ Stellen (Links) zu kennzeichnen, um dem/der UserIn deren Verfügbarkeit zu signalisieren.

Links werden noch in verschiedener Hinsicht klassifiziert (vgl. Gerdes, 1997a), besonders relevant erscheint die Unterscheidung von assoziativen und typisierten Links (Gerdes 1997a). Typisierte Links verweisen im anklickbaren Text des Links explizit auf den Zusammenhang der vorliegenden Textstelle mit dem Zieltext, der beim Anwählen des Links angezeigt wird. Assoziative Links hingegen explizieren den Zusammenhang des Verweises nicht. Der User muß sich diesen selbst erschließen.

Ein weiteres psychologisch relevantes Problem stellt potentielle Verlust an lokaler und globaler Kohärenz in der Textrezeption dar (Ohler & Nieding, 1997). Dieses psycholinguistisch naheliegende Problem hängt aus der Sicht der Hypertextgestaltung zusätzlich direkt mit dem wichtigen Merkmal der Granularität (Schulmeister, 1997) des Hypertextes zusammen, mithin stehen beide Prinzipien in negativer wechselseitiger Abhängigkeit. Da wie oben erläutert ein wesentliches Merkmal von Hypertexten die autonomen Informationseinheiten darstellen, ergibt sich als logisches Problem die Frage der Abgrenzung derartiger Einheiten respektive deren Größe. Der Hypertext ist umso stärker vernetzt, und damit z. B. adaptiv zu nutzen, je kleiner diese Einheiten gewählt werden. Zugleich ist es offensichtlich, daß der Verkleinerung der Informationseinheiten natürliche Grenzen gesetzt sind. Jedenfalls geht mit der zunächst erwünschten Aufteilung des Textbestandes in einzeln ansteuerbare Informationselemente immer das Problem des Verlusts der globalen Kohärenz einher. Als positives Element ist dieser Effekt auch intendiert, der Autor/die Autorin des Hypertexts gibt sozusagen die Kontrolle über dieses Gestaltungselement an den Leser/die Leserin ab. Gelingt es diesem/dieser allerdings bei der Rezeption nicht, eine „eigene“ globale Kohärenz herzustellen, wird die Rezeption durch dieses Merkmal des Hypertextes behindert und es tritt das Phänomen des Orientierungsverlustes auf.

Zusätzlich zur gesteigerten Autonomie des Rezipienten/der Rezipientin beim Rezeptionsprozeß kann die Steigerung der kognitiven Komplexität und (im Falle des Gelingens) die Zunahme potentieller Inkongruenz im übrigen noch als weiteren positiven Effekt eine Steigerung der Neugier und damit ein erhöhtes Leseinteresse zur Folge haben (Groeben, 1982; Taylor, 1990).

Sollte die Herstellung „individueller“ globaler Kohärenz unter Einsatz entsprechender Hilfsmittel aber gelingen (Thüring, Hannemann & Haake, 1995), wäre mit der Präsentationsform des Hypertexts möglich, was lernpsychologisch (unter Berücksichtigung der ko-

gnitiven Konstruktivität des Individuums) schon immer sinnvoll erschien: Es gäbe dann so viele individuelle Text-Varianten wie Rezipienten/Rezipientinnen.

Für die Fragestellung dieser Arbeit kann allerdings als zentrales Problem die Nicht-Linearität von Hypertexten festgehalten werden, da diese wiederum das „Navigationsproblem“ nach sich zieht. Die UserIn wird also nicht nur mit der Strukturierung des dargebotenen Wissensbereichs, sondern (mehr oder weniger) auch mit der Navigation durch die Textbestandteile weitgehend alleine gelassen. Gelingt es allerdings, die Informationseinheiten souverän in der Form eines individuellen Lernpfades zu strukturieren, ohne daß die kognitiven Kapazitäten mit dieser Aufgabe zu sehr belastet werden, kann die Nicht-Linearität von Hypertext ein entscheidender Vorteil gegenüber linearen Lernmaterialien sein.

2. 2. 2. 2. *Vor- und Nachteile des Lernens mit Hypertexten*

Die anfängliche Euphorie hinsichtlich der Möglichkeiten von Hypertexten ist inzwischen deutlicher Ernüchterung gewichen, insbesondere nachdem viele Erwartungen durch die Ergebnisse empirischer Untersuchungen nicht bestätigt werden konnten. So konnte Landauer (1995) in seiner Analyse zur Wirksamkeit hypermedialer Systeme nur neun empirische Untersuchungen ausmachen, die seiner Einschätzung nach wissenschaftlichen Standards gerecht wurden. Zu diesem Zeitpunkt wurde aber bereits seit zehn Jahren mit Hypertextsystemen geforscht und deren Vorteile allorts beredt gelobt. Zusammenfassend kommt er zu dem Schluß, daß „in almost all cases, users were quicker and more successful using paper and print than the electronic form“ (zit. n. Dillon & Vaughan, 1997, S. 2).

Vielerorts zitiert ist auch die Metaanalyse von Chen und Rada (1996), in der die Autoren/innen 23 experimentelle Studien zum Vergleich von linearem Text und Hypertext analysierten. Im Ergebnis scheinen die Studien zu sehr unterschiedlichen Befunden zu kommen, allerdings mit einer klaren Tendenz zuungunsten der Probanden/Probandinnen in der Hypertext-Bedingung.

In der Analyse der möglichen Faktoren für diese Befundlage ergaben sich allerdings auch Hinweise auf theoretische Mängel und überzogene Konzeptionen. Schulmeister (1997) kommt in seinem Resümee der überzogenen Erwartungen dahin, die zugrundeliegenden Annahmen als „pädagogische Mythen“ zu bezeichnen.

Zentrale Konzepte, um die sich die vermutlich überzogenen Erwartungen an Hypertextanwendungen gruppieren lassen, sind die angebliche kognitive Plausibilität von Hypertexten,

die Annahme des nicht-linearen Denkens bzw. der kognitiven Flexibilität, die Annahme des Vorteils durch multiple Repräsentationen, Annahmen über Vorteile durch die Art des Informationszugriffes und schließlich die häufig verwendete Metapher der Navigation.

Im folgenden soll kurz zu den genannten Annahmen ausgeführt werden, inwiefern es sich dabei um „misleading theoretical assumptions“ (Tergan, 1997) handelt, und inwiefern die zugrundeliegenden Annahmen evtl. modifiziert und für erfolgversprechendere empirische Überprüfung zugänglich gemacht werden könnten.

Die Annahme der kognitiven Plausibilität

Dieser Annahme liegt - vereinfacht gesagt - die Vermutung zugrunde, die Anordnung der Inhalte in Hypertexten in der Form autonomer, miteinander vernetzter Einheiten entspräche weitgehend der Anordnung von Informationen im menschlichen Gedächtnis in der Form von semantischen Netzen (vgl. z. B. Jonassen & Wang, 1992). Daraus ließe sich dann leicht eine Annahme zugunsten des Hypertextes im Vergleich mit linearen Texten ableiten. Es bedarf im Prinzip kaum näherer Betrachtung, um die Unhaltbarkeit einer solch starken Korrespondenzhypothese zu erkennen, allerdings lässt die weite Verbreitung dieser Annahme bzw. zumindest deren implizite Unterstellung es dennoch sinnvoll erscheinen, darauf einzugehen.

Eine wie auch immer geartete Parallelität der Strukturen, Inhalte und Vernetzung von semantischen Netzen im Gedächtnis und derjenigen von Informationseinheiten in Hypertexten anzunehmen, ist theoretisch aus zwei Gründen abzulehnen. Einerseits ist die der Hypothese zugrundeliegende Annahme einer Struktur- und Inhaltsparallelität der Elemente, also von Textteilen eines Hypertextes (oder evtl. sogar Graphiken und Multimediaelementen) und Bausteinen eines semantischen Netzes als solche bereits aufgrund der Forschung zu semantischen Netzen als unzutreffend abzulehnen (vgl. hierzu Kintsch, 1974; Whalley, 1990).

Sollte die Annahme der Strukturparallelität der beiden Informationsspeicher desweiteren Sinn machen, wäre zudem eine primitive „Übernahme“ von Knoten aus dem Netz des Hypertext ins semantische Netz des Rezipienten/der Rezipientin anzunehmen. Derartige Annahmen über die Akquisition von Wissen werden schon seit langer Zeit bereits auf der Ebene einfacher Wahrnehmungsprozesse als unzutreffend und zu stark vereinfachend zurückgewiesen. Für komplexe Verarbeitungsvorgänge, wie sie bei der Rezeption eines Hypertextes angenommen werden müssen, wäre eine solche Theorie vor dem Hintergrund

zahlloser empirischer Untersuchungen im Bereich der Psycholinguistik und Gedächtnisforschung zurückzuweisen.

Ohler und Nieding (1997) verweisen zusätzlich auf die mangelhafte Berücksichtigung „eines hypothetisch ressourcenbegrenzten Arbeitsgedächtnisses bei der Informationsenkodierung und bei der Organisation von Informationsaufnahme Prozessen“ hin.

Vor diesem Hintergrund erstaunt es wenig, daß sich für die Annahme der kognitiven Plausibilität bisher keine empirischen Belege finden ließen - wie bereits erwähnt deuten bisherige Untersuchungen eher in eine entgegengesetzte Richtung.

Ob sich eine schwächere Variante der Korrespondenzhypothese im Sinne eines metaphorischen Gebrauchs des Topos „Netz“ zur Beschreibung der beiden Informationsspeicher, wie sie z. B. von Kuhlen (1991) vertreten wird als brauchbar für die Ableitung empirischer Untersuchungen erweisen wird, kann in Anbetracht der Unschärfe des Vergleichs und der komplexen Lage der Empirie ebenfalls bezweifelt werden.

Die Annahme des nicht-linearen Denkens bzw. der kognitiven Flexibilität

Eines der wichtigsten Ziele bei der Rezeption von Texten ist neben dem reinen Textverständnis sicherlich der Zugang des kritischen Lesens, insbesondere im Rahmen wissenschaftlicher Literatur (vgl. Groeben, 1982). Die bereits mehrfach angeführte und bei Hypertextsystemen regelrecht erzwungene aktive Akquisition des Textsinnes durch Bildung individueller Pfade im System entspricht zunächst wichtigen Prinzipien, die die Psychologie der Textverarbeitung und des Textverstehens als relevant für kritisches Lesen herausgearbeitet hat.

Auf der anderen Seite trifft hier in ähnlicher Weise wie bei den anderen Punkten der Einwand einer zu stark vereinfachenden Abbildung des Rezeptionsprozesses in einem derartigen Vergleich. Problematisch für die empirische Erfassung eines entsprechenden Effektes erweist sich im übrigen ein Störfaktor, der in der Natur eines derartig anspruchsvollen didaktischen Zieles zu liegen scheint: Bei einigen Untersuchungen liegt nach Beeman et al. (zit. n. Schulmeister, 1997) offensichtlich der Schluß nahe, daß die erzielten Effekte eher auf die besondere Motivation der Dozierenden bzw. den interessanten Stoff zurückzuführen sind, als auf Merkmale der Präsentation des Materials. Aufwendigere Kontrollgruppendesigns wären hier für weitere empirische Überprüfungen angezeigt (wenngleich bei einigen Variationsmöglichkeiten ethische Bedenken zum Tragen kommen dürften).

Die Annahme des Vorteils durch multiple Repräsentationen (Konstruktivismus-Annahme)

Es wird angenommen, daß vermittelt durch die multiple Repräsentation des Inhaltes in der Form unterschiedlicher Darstellungsweisen (Text, Bild, Video, evtl. sogar Ton) und durch die dem Hypertext eigene Möglichkeit, kleine Informationseinheiten individuell zu inhaltlich kohärenten Strukturen zu verknüpfen besser als bei linearen (Print-)Texten die Möglichkeit zur Verfügung gestellt wird, individuelle Sinnkonstruktionen herzustellen (Jacobson & Levin, 1995; Tergan, 1997).

Problematisch an dieser vom Konstruktivismus geprägten Vorstellung des Rezeptionsprozesses ist auch hier (wie bei den meisten überzogenen Anforderungen an Hypertext-Systeme) eine relativ naive Vorstellung von der Art, wie Menschen Informationen aufnehmen. So ist es mitnichten so, daß ein linear vorgegebener Text z. B. in der Form eines Buches auch zur Bildung linearer Kognitionsstrukturen führen würde, ebensowenig wie die Annahme ausschließlich nicht-linearer Rezeption bei einem vernetzten Hypertext zwingend erscheint. Vielmehr ist hier wie dort von einem komplexen Verarbeitungsprozeß auszugehen, indem häufig Rezipienten-/Rezipientinnen-seitige Faktoren eine mindestens ebenso große Rolle spielen dürften wie die Anlage des Informationsmaterials.

Auch hier kann wiederum auf die oben bereits angeführte mangelhafte empirische Bestätigung eines entsprechenden Lernvorteils durch Hypertexte verwiesen werden.

Annahmen zum Vorteil durch die Art des Informationszugriffes

Im Umfeld dieser Annahme wird ein Vorteil von Hypertexten durch die computerisierten Möglichkeiten des Informationszugriffes, ähnlich wie bei Datenbanksystemen angenommen.

Rouet (1992) konnte allerdings zeigen, daß traditionelle, lineare Texte bei der Suche nach Informationen trotz der zusätzlichen Möglichkeiten des Hypertexts diesem dennoch überlegen zu sein scheinen. Die bei allen Annahmen bzgl. der Nutzung von Hypertext relevante Frage der Übung/Vertrautheit mit den jeweiligen Systemen (lineare vs. nicht-lineare) stellt sich bei diesem Aspekt natürlich im besonderen. So ist es für diesen Aspekt gut denkbar, daß in neueren Untersuchungen mit großen Informationsmengen ein entsprechender Vorteil des Hypertexts dennoch gesichert werden könnte.

Annahmen zur Metapher der „Navigation“ in Hypertexten

Keine Untersuchung zur Hypertextrezeption der letzten zehn Jahre konnte ohne ein Kapitel zur Navigation, und im Zusammenhang damit zum vielzitierten Problem des „lost in hyperspace“ auskommen. Daß es sich bei der „Navigation“ in Hypertexten immer noch um eine Metapher handelt, scheint mitunter aus dem Blickfeld zu geraten, zumal die aus dem Gebrauch der Metapher abgeleiteten Probleme, Designkonsequenzen und Studien zahlreich sind.

Dennoch gibt es gute Gründe, die Angemessenheit der Metapher und deren Abbildungsleistung in Bezug auf den Rezeptionsprozeß genauer zu beleuchten, insbesondere auch deshalb, weil der bei allen zuvor genannten Einwänden gegen übertriebene Erwartungen angeführte Punkt einer zu stark vereinfachenden Konzeption auch bei diesem zentralen Konzept der Hypertextforschung zu greifen scheint.

Ausgehend von der dürftigen Befundlage empirischer Untersuchungen zur Hypertextrezeption zeigen Dillon und Vaughan (1997) in einem grundlegenden Artikel die theoretischen Annahmen und die daraus abzuleitenden Probleme bzw. Fehlschlüsse des Konzeptes der „Navigation“ auf.

Tragend für die Verwendung des Begriffes „Navigation“ im Zusammenhang der Hypertext-Forschung ist eine Analogie zwischen der Navigation in physischen Räumen und der „Navigation“ in textuellen „Landschaften“ eines Hypertexts (z. B. Parunak, 1989; Botafogo & Shneiderman, 1992). Konsequenterweise wird dann auch von Landmarken und Routen gesprochen, die User bei der Verwendung eines Hypertexts registrieren und wiederverwenden (Vinson, 1999).

Die Autoren/Autorinnen (Dillon & Vaughan, 1997) weisen darauf hin, daß es weder tragfähige empirische Absicherungen dieser grundlegenden Analogie gibt, noch daß die aus der Analogie abzuleitenden Folgerungen (die z.T. in der Literatur selbstverständlich verwendet werden), wie z.B. die angeblich unveränderliche Abfolge von Landmark- über Routen zu Überblickswissen durch entsprechende Untersuchungen gesichert werden konnten. Im Falle der genannten Teilfähigkeiten finden sich eher Studien, die eine aufgabenabhängige Verwendung der jeweiligen Strategie belegen, nicht aber eine starre hierarchische Abfolge, wie sie vielleicht beim Erwerb von Orientierung in einer physischen Umwelt zu beobachten wäre. Hinzu kommt die mißliche „Neutralität“ der Navigationsmetapher, wenn es um die Einbeziehung psychologischer relevanter Faktoren, wie z.B. von Werturteilen oder emotionalen Komponenten bei der Konfrontation mit bestimmten Inhalten, gehen soll.

Berücksichtigt man die kognitive Konstruktivität des Individuums bei der Rezeption, so wird deutlich, daß eine mechanische Konzeption in einen Hypertext implementierter Landmarken und vorgegebener Routen weder adäquat den individuellen Lernprozeß des Users unterstützen kann, noch daß diese (bzw. deren Verwendung) per se als Indikator für eine bestimmte Art der Rezeption angesehen werden kann.

Ein für vergleichende empirische Forschung und nicht zuletzt für aus dieser abzuleitende Designprinzipien erfolgreicher Hypertextanwendungen äußerst problematisches Fazit aus dieser Befundlage könnte sein, daß „the form of knowledge manifest by users was driven more by individual differences in cognition than any characteristics of the electronic space“ (Dillon & Vaughan, 1997, S. 6). Die Autoren/Autorinnen schlagen vor, ein Modell des Rezeptionsprozesses zu verwenden, das stärker die rezeptionsseitigen Prozesse in Rechnung stellt und führen die Gestalt („shape“) als Alternative ein. Als zentralen Vorteil dieses Begriffes sehen die Autoren/innen die Möglichkeit, neben strukturellen Aspekten des Hypertextes auch für die Orientierung und das Auffinden von Informationen ebenso wichtige semantische Informationen einbeziehen zu können („Shape is a property of information that is conveyed both by physical form and information content“ o.a., S. 9). Sie weisen in diesem Zusammenhang auf einen zentralen Mangel der Analogie mit der Navigation in physischen Umwelten hin: Hier besteht der Sinn der Navigation in der Erreichung eines bestimmten Zieles auf der „Karte“ - im Falle der „Navigation“ in Hypertexten ist allerdings das Erreichen eines Zieles alleine noch nichts wert, zugleich geht es hier auch immer um die Enkodierung der gegebenen sprachlichen Information, ja die Navigation selbst ist in hohem Maß von diesem Verständnisprozeß beeinflusst.

Als Folge dieser wichtigen theoretisch fundierten Korrektur der Terminologie in der Erforschung von Hypertexten schlagen Dillon und Vaughan (1997) schließlich vor, das Konzept der Navigation nicht zu verwerfen, sondern in der Evaluation von Nutzungsprofilen nicht allein die erfolgreiche Navigation von Seite A zu Seite B im Blick zu haben, sondern anhand des Konzeptes der „Gestalt“ ein Mix aus verschiedenen Faktoren zu untersuchen, die schließlich das *Verständnis* eines Hypertextes fokussieren und nicht nur dessen Navigabilität.

2. 2. 2. 3. *Das Problem des Orientierungsverlusts und der Navigation in Hypertexten*

Eine Suche nach dem von Conklin (1987) geprägten Begriff „lost in hyperspace“ ergibt in der Suchmaschine Altavista 3676 Fundstellen, zur Kombination von „hypertext“ und „dis-

orientation“ werden 1393 Dokumente gefunden. In einer Sammlung von Artikeln zur Human-Computer-Interaction der Association of Computing Machines (ACM) werden immer noch 67 Artikel gefunden, die sich diesem Thema widmen. Dennoch gibt es auch hier Anlaß zur kritischen Prüfung des Konstrukts, zumal viele Untersuchungen und Befunde im Zusammenhang mit demselben erklärt werden.

Der Grundgedanke ist einfach und naheliegend: Durch die Nicht-Linearität des Hypertextes muß der Benutzer/die Benutzerin selbst eine Struktur und Abfolge der Sinneinheiten (d.h. Knoten des Hypertextes) herstellen. Dazu ist es notwendig, sinnvolle Folgeseiten bei der Exploration oder der Suche nach einer bestimmten Information zu finden. Um dies bewerkstelligen zu können, müssen die von der aktuellen Seite erreichbaren Seiten erkannt und aus ihnen die passende ausgewählt werden. Dies wiederum erfordert eine zumindest grob adäquate Repräsentation der Struktur des Hypertextes, denn ohne diese Information kann häufig nicht beurteilt werden, an welcher Stelle in der Hierarchie des Gesamttextes sich die aktuelle Seite befindet und in welchem Bezug zu dieser Position mögliche Folgeseiten stehen. Ohne entsprechende Strukturierungshilfen, wie z. B. Übersichtskarten ist die zusätzliche Belastung durch diese Aufgaben während der Rezeption evtl. so groß, daß die Rezeption des Inhaltes darunter leidet.

Zwei wichtige theoretische Gegenargumente können hier angeführt werden, obwohl das Grundproblem sicherlich damit nicht zu lösen sein dürfte - die empirische Bewährung steht auch hier noch aus, in der Regel wird auf zukünftige Untersuchungen mit besserer Kontrolle der entsprechenden Bedingungen und größeren Stichproben verwiesen.

Zunächst leidet auch das Konstrukt des „lost in hyperspace“ unter vergleichbaren Einschränkungen wie die weiter oben diskutierten Metaphern der „Navigation“ oder der „kognitiven Plausibilität“. Denn auch die Annahme eines möglichen Orientierungsverlusts beruht auf der Voraussetzung, daß die Texteinheiten des Hypertexts und deren (technische) Verknüpfung in der Form von sogenannten Links die zentralen Faktoren bei der Bildung einer Repräsentation des Textsinnes darstellen - eine Annahme, die aus psychologischer, insbesondere psycholinguistischer Sicht sicherlich zu hinterfragen ist. Vermutlich würden sich auch problemlos Situationen herstellen lassen, in denen Versuchspersonen mit traditionellen linearen Texten in ähnlicher Weise Desorientierung erleben, wie es so häufig beim Umgang mit Hypertexten berichtet wird. In gewisser Weise ist die Ableitung zu erwartender Orientierungsprobleme aus einer derartigen kognitionspsychologisch naiven Modellie-

zung des Gegenstandsbereichs ähnlich plausibel wie die oben bereits diskutierte Annahme kognitiver Plausibilität aufgrund einer vermuteten Strukturparallelität des semantischen Gedächtnisses mit einem Hypertext. Eine empirisch fundierte, theoretisch aber nicht auf derartige Theoriekonzeptionen festgelegte Überprüfung des Konzeptes „lost in hyperspace“ wäre also angeraten, zumal auch hier die bisherige Befundlage die Kritik an der theoretischen Ableitung deutlich zu stützen scheint.

Es gibt allerdings noch einen weiteren psychologischen Grund, am Modell der Desorientierung Zweifel anzumelden. So weist Schulmeister (1997) daraufhin, daß Kahn und Landow (1992, zit. n. Schulmeister, 1997) in einer vergleichenden Untersuchung über verschiedene Hypertexte zu dem interessanten Befund kamen, daß Experten/Expertinnen offensichtlich keine Probleme mit Situationen haben, für die gewöhnlich ein Orientierungsverlust angenommen wird. Die erfahrenen User in dieser Studie hatten sogar bisweilen eher Vergnügen an solch unklaren Situationen ohne eindeutige Handlungsalternativen. In der Charakterisierung des Browsingverhaltens gibt es für diesen Fall bereits ein von Kuhlen (1991) eingeführtes Konzept: „serendipitous browsing“. Damit wird der Lustgewinn beim ziellosen „Umhersurfen“ in Hypertexten beschrieben und die Freude an der zufälligen Entdeckung interessanter Inhalte. Ein Konzept, daß offensichtlich die Sorge vor der Desorientierung des überforderten Users/der Userin konterkariert (zumindest für diese Art der Auseinandersetzung mit Hypertexten im Zusammenhang mit klar strukturierten Retrieval-Tasks stellt der Orientierungsverlust vermutlich keinen großen Lustgewinn dar). Daß Inkohärenz mitnichten automatisch den Verständnisprozeß bei der Textrezeption hindern muß, sondern in Maßen sogar motivationssteigernd wirken kann, ist indes in der Forschung zur Psychologie der Textverarbeitung bereits lange bekannt (Groeben, 1982). Auch Ohler und Nieding (1997) warnen vor zu schnellen Versuchen, angeblich essentiell behindernde Probleme aus dem Weg räumen zu wollen:

Als Forschungsperspektive für die WWW-Benutzung folgt, daß man nicht vorschnell Benutzungsoberflächen zu optimieren versuchen sollte, sondern statt dessen genauer die Textur der Interaktion untersucht. Auch Fehlerkorrekturen sind Teil des produzierten Textes und selbst mit seinem 'Lost in Hyperspace' erzählt der Benutzer eine Geschichte, die insofern kohärent ist, als er mit seinen Versuchen der Reorientierung den linearen Text seiner Interaktionen 'weitschreibt', auch das Thema in der Geschichte lautet: 'I am in trouble'. (Ohler & Nieding, 1997, S. 177f)

Auf einen ähnlichen Aspekt in der Interpretation fehlerhaften Verhaltens (als solches kann der Orientierungsverlust ja beschrieben werden) weist Rauterberg (1997) in einem grund-

legenden Artikel zu diesem Thema („About faults, errors and other dangerous things“) hin, wenn er feststellt, daß Fehler und deren Behebung vermutlich einen größeren Beitrag zur Bildung eines adäquaten mentalen Modells herstellen, als vermeintlich korrektes Verhalten.

Im Design dieser Untersuchung, aber auch bei der Beurteilung bestimmter Ergebnisse sollte infolgedessen die Komplexität des Rezeptionsprozesses bedacht und auch die theoretisch/empirische Beleuchtung des Konstrukts der Desorientierung fokussiert werden.

2.2.3. Interaktionsstile im Design der Mensch-Maschine-Schnittstelle

2. 2. 3. 1. Vorbemerkung

Im folgenden sollen zwei ausgewählte Problemfelder der Forschung zur Mensch-Maschine-Kommunikation (Human-Computer-Interaction) in Grundzügen dargestellt werden. Die Auswahl orientiert sich dabei an der vorliegenden Fragestellung und erhebt keinen Anspruch auf eine vollständige Darstellung des Feldes. Statt dessen wird in Anlehnung an die Klassifikation von Shneiderman (1998) auf die Interaktionsstile eingegangen, die auch im Zusammenhang mit der Bearbeitung von Hypertexten eine Rolle spielen können.

2. 2. 3. 2. Menu-Selection

Ein wesentlicher Fortschritt in der Entwicklung moderner Betriebssysteme für Personal-Computer bestand in der Entwicklung graphischer Repräsentationen der zur Verfügung stehenden Kommandos eines Programmes, die anstelle komplizierter Befehlssyntax verwendet werden konnten (Shneiderman, 1998; Norman, 1991). Zunächst wurden in diesem Zusammenhang aufklappende Menüs in die Benutzerschnittstellen eingebaut, die mit Tastaturbefehlen oder mit der Maus bedient werden konnten. Bei einer großen Anzahl möglicher Programmbefehle ergeben sich hier bereits Probleme der Anordnung und Visualisierung der Daten, die sich in allen anderen Ansätzen zur Darstellung komplexer Informationen wiederfinden lassen.

Shneiderman (1998) faßt die Ziele von Menüsystemen folgendermaßen zusammen: „The primary goal for menu, form-fillin, and dialog-box designers is to create a sensible, comprehensible, memorable, and convenient organization relevant to the users’s task“ (Shneiderman 1998, S. 237)

Im wesentlichen gibt es zwei Probleme, die bei der Entwicklung von Menüsystemen bewältigt werden müssen: Die Informationen müssen in eine sinnvolle Ordnung gebracht werden und die Darstellung sollte so gewählt werden, daß die Informationen möglichst schnell und leicht zugänglich sind.

Daß eine Sortierung der Informationen nach inhaltlich sinnvollen Kriterien brauchbarere Interfaces zur Folge hat als eine Darbietung ohne Struktur ist naheliegend, aber keineswegs trivial. Dies wird insbesondere dann deutlich, wenn man zwei variable Faktoren berücksichtigt, die in der Benutzung derartiger Systeme von größter Bedeutung sein dürften, aber sicherlich verschiedenste Ausprägungen annehmen können: Die Struktur der Aufgabe und die kognitive Aktivität des Users. Beides kombiniert, ergibt eine nahezu beliebige Vielfalt an sinnvollen Konstellationen und Anordnungen von Befehlen in Menüsystemen.

Trotz dieser erheblichen Komplexität der Anforderungen („there is no perfect menu structure, that matches every person’s knowledge of the application domain“ Shneiderman 1998, S. 251) zeigen Untersuchungen die Wirksamkeit allgemeiner Prinzipien. So arbeiten User offenbar effektiver mit Menüstrukturen, die in die Breite gehen, und weniger effektiv, wenn die gleichen Informationen in tiefer verschachtelten, dafür aber „schmaleren“ Menüs angeboten werden (Shneiderman 1998).

Da mit zunehmender Informationsmenge aber dennoch auch in der Tiefe verschachtelte Menüs konstruiert werden müssen, ergibt sich auch hier ein Orientierungsproblem, denn mit zunehmender Verschachtelung wird es für den Benutzer immer schwieriger festzustellen, in welcher Relation zu den anderen Einträgen der momentan selektierte Menüeintrag steht.

Hier wurden graphische Repräsentationen von Menüsystemen entwickelt, die gleichzeitig möglichst viele Elemente eines Menüs am Bildschirm anzeigen, und zwar sowohl in der Breite als auch in der Tiefe. Diese „menu maps“ werden insbesondere häufig im Rahmen von Internet-Benutzungsoberflächen für das World-Wide-Web verwendet. Die Grenze einer solchen graphischen Karte ist allerdings ebenfalls relativ schnell erreicht, da es nicht möglich ist, beliebig viele Items (Menüeinträge) am Bildschirm gleichzeitig in lesbarer Form zu präsentieren.

Eine Lösung für dieses Grundproblem stellen spezielle Verzerrungstechniken und „fisheye-views“ dar (vgl. Abschnitt 2. 2. 6. 3).

Um das Problem der unterschiedlichen Nutzergruppen (z. B. Experten/innen und Novizen/innen) und der aufgabenabhängigen Nutzung (vgl. Abschnitt 2.2.5) wenigstens ansatzweise

zu lösen, wurden zunehmend multiple Zugangsweisen verwendet, die nach Möglichkeit fließende Übergänge aufweisen sollten (z. B. die Möglichkeit der Verwendung von Shortcuts alternativ zum manuellen Aufruf von Menüeinträgen). Prinzipiell wird das Problem so vermutlich nicht befriedigend zu lösen sein, und komplexere Möglichkeiten adaptiver Systeme (auch wenn bisherige diesbezügliche Ergebnisse eher ernüchternd zu sein scheinen) wären in Erwägung zu ziehen (so verwendet z. B. das Betriebssystem „Windows98“ einen internen Zähler, um wenig verwendete Menüeinträge auszublenden allerdings ein sicherlich noch sehr rudimentärer Ansatz der Adaptivität).

2. 2. 3. 3. *Direct Manipulation*

Als weiterer genereller Modus des Designs der Benutzerschnittstelle ist für die gegenwärtige Fragestellung die „direct-manipulation“-Technik relevant. Zentral bei dieser Darstellungstechnik ist der Umstand, daß sozusagen auf eine Ebene der Abstraktion/Repräsentation verzichtet wird, und stattdessen die zu bearbeitenden Objekte direkt manipuliert werden können. So ist es z. B. in modernen Textverarbeitungsprogrammen möglich, die Änderungen des Layouts, der Schriftgröße oder eingefügte Graphiken direkt im Bearbeitungsmodus am Bildschirm zu sehen, wohingegen in früheren Versionen das Ergebnis einer Layout-Veränderung häufig erst beim Ausdruck zu erkennen war. Noch eine Generation früher wurden ganze Texte nur im Kommandozeilen-Modus bearbeitet, d.h. es war immer nur eine Zeile des Textes am Bildschirm zu sehen.

Als Erklärung für die (empirisch gesicherten) Vorteile derartiger Systeme führt Shneiderman (1998) Einsichten aus der Psychologie des komplexen Problemlösens an. Denn auch diese kam zu dem Ergebnis, daß die Art der Repräsentation eines Problems entscheidenden Einfluß auf die Lösungsgüte haben kann, wobei graphische Visualisierungstechniken sich häufig als vorteilhaft erwiesen (vgl. auch Hussy, 1993).

Wichtig ist allerdings eine Einschränkung, die in diesem Vergleich bereits anklingt. Aufwendige Visualisierungen sind besonders vorteilhaft im Zusammenhang mit komplexen Problemen. So sind durchaus Umstände denkbar, unter denen eine aufwendige Visualisierung zu einer Abnahme der Effizienz und der Zufriedenheit führen können, wenn z. B. Personen, die sowohl bezüglich der Instrumente als auch hinsichtlich der Domäne als Experten gelten, einfache Probleme mit derartigen Interfaces lösen sollen.

For experienced users, a tabular textual display of 50 document names may be more appropriate than only 10 graphic document icons with the names abbreviated to fit the

icon size (Shneiderman 1998, S. 204).

Als weitere mögliche Nachteile von „direct-manipulation-systems“ nennt Shneiderman (1998):

- User müssen die Bedeutung der zugrundeliegenden Metapher und ihrer Elemente (z. B. Icons) lernen;
- die graphische Repräsentation kann unscharf sein und zu falschen Schlüssen führen;
- erfahrene User können Computer häufig mit Tastaturbefehlen wesentlich schneller bedienen als mit der Hilfe einer Maus.

Die genauere Untersuchung der Vor- und Nachteile von „direct-manipulation-systems“ im Gegensatz zu Menüsystemen (im weiteren Sinne) wird nicht zuletzt Gegenstand der vorliegenden Untersuchung sein.

2.2.4. Probleme bei der Visualisierung komplexer Informationen

2. 2. 4. 1. Zusammenhang mit Menüsystemen

Auf den ersten Blick besteht kein relevanter Zusammenhang zwischen Menüsystemen von Computerprogrammen und den Visualisierungstechniken für Web-Seiten und Hypertextsystemen. Wenn man allerdings von der engen Definition eines Menüsystems als graphischer Darstellung der verfügbaren Befehle eines Computersystems abläßt und den Vergleich auf der Basis einer weiteren Definition vornimmt, die Menüsysteme als Visualisierungen komplexer hierarchischer Informationsräume begreift, werden die Parallelen evident.

Tatsächlich sind sowohl die Probleme als auch gewisse Lösungsansätze sichtlich verwandt, und zwar insofern, daß es sowohl bei Übersichtskarten und graphischen Browsern in Hypertextsystemen als auch in den verschiedenen Varianten von Menüsystemen um das Problem einer adäquaten Visualisierung hierarchischer Daten geht.

Der Unterschied liegt darin, daß im Falle des Hypertexts die darzustellenden Daten weniger strukturiert und in der Regel auch umfangreicher sind. Zudem spielt hier der Kontext, in den die jeweiligen Items (d.h. Hypertext-Knoten) eingebettet sind eine größere Rolle als in Menüsystemen, da es sich beim Hypertext ja immer noch primär um einen Text handelt.

Zusätzlich zu den genannten Parallelen der beiden Forschungsbereiche ergibt sich eine inhaltlich schwerwiegende Parallele durch ein beiden zugrundeliegendes Grundproblem, welches theoretisch bereits erörtert wurde. Sowohl Theorien zur Entwicklung und Verbes-

serung von Menüsystemen als auch solche zur Erforschung der Verarbeitung komplexer Hypertextsysteme rekurren häufig auf ein theoretisches Rahmenkonzept, in welchem nicht selten die Theorie mentaler Modelle (im weitesten Sinne) eine Rolle spielt. Konkret geht es also häufig darum, angemessene Strategien und Techniken zu finden (z. B. passende Metaphern oder landmarks zu finden), die die Bildung eines adäquaten mentalen Modells beim User befördern. Darin unterscheiden sich die beiden Forschungsgebiete nicht.

Eine gewisse Überlappung in den folgenden kurzen Ausführungen zur Visualisierung komplexer Informationen mit den vorangegangenen Kapiteln zu Menüsystemen ist also durchaus intendiert.

2. 2. 4. 2. *Fokussieren und Kontext*

In der Visualisierung umfangreicher Informationsmengen wird schon seit langem mit graphischen Übersichten operiert (Schaffer et al., 1996; Furnas & Bederson, 1995). Relativ schnell wurde allerdings deutlich, daß ein Problem derartiger Übersichtskarten darin besteht, daß sie entweder zu generell sind, und damit wichtige Details vermissen lassen, oder zu speziell („local maps“) und damit wichtige Informationen über den Gesamtkontext verlorengelassen - auch der Versuch mit mehreren Fenstern zu arbeiten, in denen beide Arten von Informationen dargeboten werden, litt an der zusätzlichen Beschränkung des Bildschirmplatzes und erwies sich als wenig tauglich (Schaffer et al., 1996; Nielsen, 1990).

Zusätzlich ergibt sich mit solchen Visualisierungen ein Problem, das bereits in Abschnitt 2.2.2 für Hypertexte und unter Abschnitt 2.2.4 für Menüsysteme angesprochen wurde: Auch hier wurde bei Versuchspersonen das Problem des Orientierungsverlusts festgestellt, das naheliegenderweise in Zusammenhang mit dem Verlust des globalen Kontexts auftritt. Die Anforderung bestand demnach darin, eine Visualisierung zu erzeugen, die nach Möglichkeit sowohl Kontext- als auch Detailinformationen liefern konnte. Hierzu wurden zahlreiche Varianten von verschiedenen Forschungsgruppen und Firmen entwickelt (eine sehr gute Übersicht bietet die „Online Library of Information Visualization Environments“ der Universität Maryland unter folgender WWW-Adresse: „<http://otal.umd.edu/Olive/>“). Die Variationen der verschiedenen Realisierungen der Ansprüche in Programmen bestehen im wesentlichen in der Anzahl der Dimensionen der Darstellung (eindimensional bis multidimensional), der Art der gewählten Metapher und der Auswahl an dargestellten Informationen (im wesentlichen: hierarchische Strukturen vs. Netzstrukturen).

Eine zentrale Innovation wurde erstmals von Furnas (1986) in seinem Artikel „Generalized fisheye views“ vorgestellt. Furnas hatte zusätzlich zu den bisher eingesetzten Zoom-Techniken eine Darstellungsform entwickelt, die es ermöglichte, gleichzeitig auf Detailinformationen zu fokussieren und dennoch den umliegenden Kontext weiterhin darzustellen. Um dies zu ermöglichen, mußten die vom Fokus weiter entfernt liegenden Informationen verzerrt bzw. gestaucht werden. Die Überlegenheit einer solchen Darstellung über einfache Zoomtechniken ohne Verzerrung konnte empirisch gesichert werden (McDonald & Stevenson, 1998; Schaffert et al., 1996; Smith, Newman & Parks, 1997; Zizi & Beaudoin-Lafon, 1994).

Einige Elaborationen dieser Technik wurden zusätzlich mit der Theorie der Navigation durch landmarks kombiniert und im World-Wide-Web angewendet (Mukherjea & Hara, 1997), allerdings bisher ohne empirische Überprüfung des Effektes einer derartigen Unterstützung.

Andere Entwicklungen (insbesondere für hierarchische und stark strukturierte Daten) versuchen das Ziel, Kontext und Fokussieren zugleich zu ermöglichen, und zwar durch die Kombination von speziellen Komprimierungs-/Aggregationstechniken und mit einer ausgefeilten Zoom-Technik zu erreichen (z.B. „InfoZoom“; Koenemann, 1999; „Table Lens“; Rao & Card, 1994).

Eine Weiterentwicklung der Grundgedanken von Furnas wird in 2. 2. 6. 3 mit dem Hyperbolic-Tree-Browser vorgestellt.

2.2.5. Verschiedene „information needs“ und aufgabenabhängige Suchstrategien

Für die Analyse von Informationsverarbeitungsprozessen im Zusammenhang mit Hypertexten ist es zunächst von Bedeutung, die Abhängigkeit der gewählten Problemlösestrategien von den „information needs“ (Fowler, Bradley & Fowler, 1992; Nielsen, 1989; Rauterberg & Fjeld, 1998) zu berücksichtigen, aber auch Abgrenzungen von Untersuchungen im Bereich von Information-Retrieval-Systemen vorzunehmen.

Die Klassifikation umfangreicher Hypertext-Systeme im Hinblick auf die Art der Informationssuche, die diese erfordern, ist keineswegs trivial. So beinhalten Hypertextsysteme deutlich Elemente von Volltextumgebungen und wären damit von datenbankbasierten Retrieval-Systemen abzugrenzen, die in der Regel über ein Such-Interface abgefragt werden. Andererseits ist der Umfang der Informationen und verschiedenen Seiten des World-Wide-Web inzwischen in einem Ausmaß gestiegen, daß die Informationssuche mit einer explora-

tiven Browsing-Strategie alleine nicht zielführend sein kann. In den meisten Fällen wird man deshalb eine Kombination aus verschiedenen Strategien beobachten können, die vom Abfragen von Suchmaschinen mit booleschen Operatoren bis hin zur Lektüre einzelner Seiten am Bildschirm die verschiedensten Strategien der Auswahl und Bewertung von Information beinhaltet.

Unabhängig davon kann aber dennoch eine Grobklassifikation verschiedener, aufgabenabhängiger Strategien getroffen werden. So schlagen Smith et al. (1997) eine Klassifikation von „searching tasks“ und „exploratory tasks“ vor. Die Unterscheidung zwischen „Searching“- und „Browsing“-Strategien findet sich in vielen Arbeiten. Catledge und Pitkow (1995) unterscheiden zusätzlich zu den genannten noch Serendipitous-Browsing-Strategien, die sich dadurch kennzeichnen lassen, daß der User zufällig auf neue Informationen stößt und sich von den angebotenen Verzweigungen „treiben“ läßt. Im engeren Sinn ist dies sicherlich nicht mehr als „Strategie“ zu bezeichnen, mit dem populären Begriff des „Surfens“ im World-Wide-Web ist allerdings offensichtlich genau diese Art der Informationsverarbeitung gemeint. Im übrigen besteht ein interessantes Spannungsfeld zwischen dieser Art der Suchstrategie und dem häufigen Phänomen des Orientierungsverlustes.

Für den Bereich der Such- und Problemlösestrategien im generellen Umgang mit Hypertextsystemen gibt es desweiteren eine Reihe von Studien, die einen Zusammenhang zwischen diesen und bestimmten Lernstrategien empirisch sichern konnten (Verheij, Stoutjesdijk & Beishuizen, 1996). Dabei konnten „surface processors“ und „deep processors“ anhand ihrer Navigationsstrategie unterschieden werden. In Studien zur allgemeinen Akquisition von Wissen aus Hypertexten konnten vergleichbare Strategien, nämlich „depth-first“ und „breadth-first“ festgestellt werden (de Bra, Houben, de Vocht & Kornatzky, 1994), wobei sich die „depth-first“-Strategie als deutlich überlegen erwies. Im übrigen stimmt dieser Befund mit Analysen überein, die Newell (1990) generell für Problemlösestrategien beschrieben hat. Auch hier zeigt sich die depth-first-Strategie (bzw. genauer „modified progressive deepening“ Newell 1990, S. 11) anderen Strategien gegenüber überlegen.

Wichtig für den Zusammenhang mit der vorliegenden Studie ist allerdings, daß es sich bei den genannten Strategien um solche handelt, die Personen anwenden, wenn sie einen Hypertext im weitesten Sinne „lesen“, sich also ohne spezielles Ziel durch die Seiten bewegen. Davon abzuheben ist die Situation, in der Personen gezielt Informationen in einem Hypertext suchen („searching tasks“).

2.2.6. Lösungsversuche

2. 2. 6. 1. Vorbemerkung zur Auswahl der Lösungsversuche

Da in der vorliegenden Arbeit nicht der Raum für eine erschöpfende Darstellung aller Lösungsversuche der in den vorangegangenen Kapiteln dargestellten Probleme gegeben ist, soll nur anhand von den beiden hier in der Untersuchung verwendeten Applikationen *WebToc* und *Hyperbolic-Tree* dargestellt werden, inwiefern diese den Versuch unternehmen, die genannten Probleme jeweils zu lösen.

2. 2. 6. 2. *WebToc*

WebToc wurde 1997 von David Nation an der Universität Maryland entwickelt (Nation, 1998; Nation, Plaisant, Marchionini & Komlodi, 1997).

Das Programm stellt den Inhalt der zugrundeliegenden WebSite in der Form eines hierarchischen Inhaltsverzeichnis dar, dessen Kategorien auf- und zugeklappt werden können. Als zusätzliche Funktionen kann dieses Inhaltsverzeichnis recherchiert werden, und weitere Informationen über die Anzahl von Dateien in einem Unterverzeichnis (d.h. einer Kategorie) und deren Größe werden mit farbigen Balken angezeigt (vgl. Abbildung 4).

Die Grundlage dieser Darstellung ist die Dateistruktur der Dateien auf dem Webserver, die über ein mitgeliefertes Java-Programm analysiert und in entsprechender Form abgespeichert wird.

In der Implementation des Programmes wird in der Regel eine Aufteilung des Bildschirms in ein kleineres linkes Fenster, welches *WebToc* enthält und ein größeres rechtes Fenster, in welchem die jeweiligen Seiten der Web-Site dargestellt werden, realisiert.

Aus der Sicht der zuvor diskutierten Probleme im Design von Menüsystemen kann *WebToc* nur eine teilweise befriedigende Lösung der gegebenen Probleme bescheinigt werden. So stellt *WebToc* zwar ein gut bedienbares lineares Menüsystem zur Verfügung. Allerdings ist das Prinzip, möglichst breite Menüstrukturen zu realisieren, bereits verletzt. Zudem muß das Anwendungsfenster bei größeren Hierarchien nach unten gescrollt werden, was sich in vielen Studien als generelles *usability*-Problem erwiesen hat. Erfolgversprechende Ansätze der Direct-Manipulation-Systeme wurden nicht umgesetzt.

Unter dem Blickwinkel der Theorie mentaler Modelle kann das System hingegen als relativ gelungen gelten. Mit der Metapher des Inhaltsverzeichnis wurde an gut verfügbares Alltagswissen aus dem Print-Bereich angeknüpft. Die Metapher gestattet es aber gleichzeitig,

relativ genau und umfassend wichtige Relationen und Items des Zielbereichs angemessen abzubilden. Einzig der Faktor der Neugierevokation ist durch die relativ konventionelle Metapher und die Umsetzung des Interfaces nicht gut gelungen.

Im Hinblick auf die Faktoren produktbezogener *usability* (s.u. 2. 2. 7. 3) kann für *WebToc* die Erfüllung der Kriterien „Minimize user memory load“, „Consistency“, „help and documentation“ (insbesondere im Hinblick auf die eigens für die Untersuchung erstellte Online-Hilfe in deutscher Sprache) angenommen werden. Wie auch für den *Hyperbolic-Tree* (s.u.) muß allerdings festgestellt werden, daß die zentrale Forderung nach „Shortcuts“ (Tastenkürzeln zur alternativen Auslösung von Programmaktionen) ebensowenig umgesetzt wurde, wie eine Unterstützung im Falle von Fehlbedienungen. Kritisch für die Bedingung „Consistency“ könnte sich der Umstand erweisen, daß Mausklicks auf Einträgen des *WebToc*-Verzeichnisses je nach Hierarchieebene gänzlich unterschiedliches Verhalten zur Folge haben (entweder wird der darunterliegende Zweig geöffnet, oder - wenn es keinen solchen mehr gibt - die entsprechende Seite im rechten Fenster geöffnet). Die weitere Beurteilung soll im Rahmen der Analyse der Gesamtergebnisse erfolgen (vgl. dazu 7). Hinsichtlich des Aspektes „technology variety“ im Rahmen der Forderungen zur *universal usability* (vgl. 2. 2. 7. 2) werden ebenfalls die experimentellen Ergebnisse weiteren Aufschluß geben.

Die Autoren/Autorinnen führen aufgrund erster Studien mit dem Instrument eine leichte Erlernbarkeit und die prinzipielle Vertrautheit mit der Visualisierung des Programmes als Vorteil an und konnten leichte Vorteile gegenüber statischen Inhaltsverzeichnissen feststellen (Nation, 1998; Nation, Plaisant, Marchionini & Komlodi, 1997).

2. 2. 6. 3. Hyperbolic-Tree-Browser

Der Hyperbolic-Tree-Browser wurde in den Xerox-Forschungslaboratorien entwickelt und erstmals 1995 der Öffentlichkeit vorgestellt (Lamping & Rao, 1996; Lamping, Rao & Pirolli, 1995).

Das Programm versucht (ausgehend von einer Inspiration durch die Graphik von M.C. Escher, siehe Abbildung 2), das Problem von Fokus und Kontext bei großen Mengen (hierarchischer) Informationen zu lösen, indem die im Fokus liegenden Objekte (oder Knoten) zentral, gleichzeitig aber alle anderen vorhandenen Informationen und deren Vernetzungsstruktur dargestellt werden. Dies wird durch eine zunehmende visuelle Kompression der

entfernt liegenden Knoten erreicht - ein Prinzip, das erstmals von Furnas (1986) als „fish-eye-view“ beschrieben wurde. Theoretisch liegt dieser Art der Darstellung komplexer Hierarchien ein Abkehr von Prinzipien der Euklidischen Geometrie zugunsten der hyperbolischen Geometrie zugrunde (vgl. Lamping, Rao & Pirolli, 1995).

Der Fokus kann dabei einfach per Mausklick auf das interessierende Objekt geändert werden. In diesem Fall wird der Wechsel in die neue Konfiguration mit einem animierten Übergang dargestellt, um das Phänomen des Orientierungsverlusts zu vermeiden („animated transition between different views of a structure can maintain object constancy can help the user assimilate the changes across views“, Lamping, Rao & Pirolli, S. 8). Als weitere Maßnahme gegen dieses Phänomen wurden spezielle Mechanismen in den Darstellungsalgorithmus eingebaut, die ein übermäßiges Rotieren von Objekten während eines Fokuswechsels verhindern sollen.

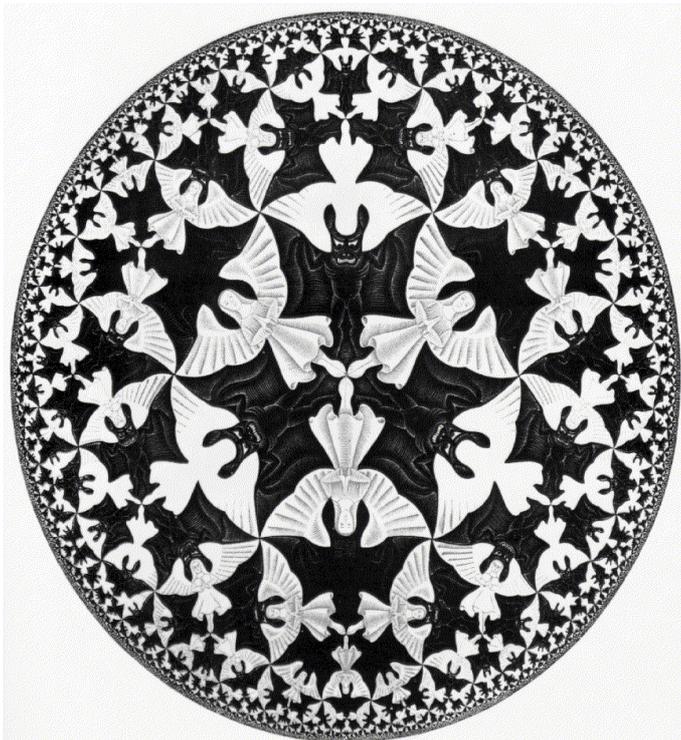


Abbildung 2: *Inspiration für die Visualisierung des Hyperbolic-Tree-Browsers: Graphik von M.C. Escher*

Theoretisch können mit dieser Art der Darstellung eine ganze Reihe zuvor besprochener Probleme in der Visualisierung umfangreicher Informationsmengen gelöst werden. Bezüglich der Probleme, die im Zusammenhang mit dem Design von Menüsystemen bekannt

sind, löst der Hyperbolic-Tree-Browser das Problem einer möglichst breiten Menüstruktur durch die Verzerrung von Menüeinträgen an den Rändern. So werden immer alle „Menüeinträge“ (bzw. Knoten) einer Ebene angezeigt - wenn sie im Focus sind, auch in lesbarer Form, ansonsten gestaucht und nur schematisch zu erkennen.

Desweiteren stellt das Programm eine Synthese aus einem Direct-Manipulation-System und einem Menüsystem dar. Aus der Sicht der Theorie mentaler Modelle kann angenommen werden, daß eine Metapher gefunden wurde (Baumstruktur), die in der Lage ist, viele Relationen des Zielbereiches zu repräsentieren, die zugleich aber an vertrautes Wissen anknüpfen kann. Das zweite metaphorische Element in der Darstellung ist das Prinzip der verzerrten Darstellung im „fisheye-view“. Auch hier scheint es gut zu gelingen, mit der gewählten Darstellungsmetapher sowohl die technischen Anforderungen (möglichst viele Informationen auf dem Bildschirm darstellen zu können) zu erfüllen als auch eine Metapher zu finden, die an breit verfügbares Alltagswissen anknüpft (sozusagen „der Blick durch den Spion“). Ob diese Anforderungen, insbesondere jene an die gewählten Metaphern, allerdings tatsächlich erfolgreiche Umsetzungen darstellen, können letztlich nur empirische Analysen zeigen.

Hinsichtlich der Einhaltung definierter Standards produktbezogener *usability* (s.u. 2. 2. 7. 3) ergibt sich in der Beurteilung des *Hyperbolic-Tree* ein durchwachsenes Bild, soweit dies vorab beurteilt werden kann: Zunächst kann die Anwendung für sich in Anspruch nehmen, die Forderungen „Minimize user memory load“, „Feedback“, „help and documentation“ (unter Berücksichtigung der für dieses Experiment angefertigten Hilfefunktion), „simple and natural dialogue“ und „speak the Users’s language“ zu erfüllen, wobei alle genannten Faktoren nach der empirischen Überprüfung zugeführt werden müssen. Eine wesentliche Forderung, die es ermöglichen soll, multiple Zugangsweisen für unterschiedlich geübte Benutzer zu schaffen, nämlich sogenannte „Shortcuts“ sind allerdings nicht implementiert. Ebenso ist im Zusammenhang mit der Forderung „Consistency“ und „Minimize user memory load“ die Frage zu stellen, ob sich der fehlende History-Mechanismus nicht sehr zu Lasten der Erfüllung dieser Forderungen auswirkt. Ein weiterer Problembereich sind die Forderungen „Good error messages“ und „Prevent Errors“. Wenn z.B. mit einem versehentlichen Mausklick der Fokus an einer nicht gewünschten Stelle gesetzt und der gesamte Bildschirm daraufhin umgruppiert wird, ist dies sicherlich als „Fehler“ zu klassifizieren. Das Programm bietet aber keine Möglichkeit, einen solchen Schritt rückgängig zu machen.

Andererseits kann entlastend angeführt werden, daß es sich beim *Hyperbolic-Tree* um ein Werkzeug zur Navigation in internetbasierten Hypertexten handelt, nicht aber um ein stationäres Programm, mit dem Benutzer permanent arbeiten. Zusätzlich gibt es technische Beschränkungen (z.B. Ladezeit), die die Realisierung aufwendiger Features behindern. Deshalb ist es zunächst vertretbar, trotz dieser Mängel bezüglich *usability* von einem Vorteil in der Nutzung des Tools auszugehen; näheres kann nur die empirische Überprüfung erbringen. Die Beurteilung der Aspekte von *universal usability*, insbesondere die Forderung nach einer Beachtung der unterschiedlichen technischen Standards kann ebenfalls nur angemessen empirisch überprüft werden.

In einer aktuellen Evaluationsstudie (Pirolli, Card & van der Wege, 2000), die den Einfluß der Navigationsbedingung (*Hyperbolic-Tree* vs. Microsoft-Dateimanager) bei verschiedenen Suchaufgaben untersuchte, kamen die Autoren zu dem Schluß, der *Hyperbolic-Tree* sei einem vergleichbaren Visualisierungswerkzeug ohne „fisheye“-Technik bei der Beantwortung von Retrieval-Fragen überlegen (in Bezug auf die Bearbeitungszeit). Weitere Diskussion dieser Ergebnisse im Zusammenhang mit der vorliegenden Untersuchung soll in Kapitel 7 erfolgen.

2.2.7. Usability-Forschung

2. 2. 7. 1. Vorbemerkungen

The goal of universal access to information and communications services is compelling. Enthusiastic networking innovators, business leaders, and government policymakers see opportunities and benefits from widespread usage. But even if they succeed and the economies of scale bring low costs, computing researchers will still have much work to do. They will have to deal with the difficult question: How can information and communications services be made usable for every citizen? Designing for experienced frequent users is difficult enough, but designing for a broad audience of unskilled users is a far greater challenge. Shneiderman (2000, S. 85)

Forschung zur Human-Computer-Interaction ist ohne *usability*-Forschung eigentlich nicht denkbar. Die Übersetzung von *usability* in „Benutzerfreundlichkeit“ ist etwas unscharf, da mit dem Terminus „Freundlichkeit“ Aspekte wie Effektivität und Erlernbarkeit nicht angemessen abgebildet werden, die andere Alternative „Softwareergonomie“ hingegen bildet einen Teilbereich der *usability* nicht ab, der im folgenden dargestellt wird (*universal usability*) - aus diesem Grund wird im folgenden der englische Terminus weiterverwendet.

Insbesondere aber vor dem Hintergrund der psychologischen Erkenntnis der Unvorhersehbarkeit menschlicher Rezeptionsprozesse und der kognitiven Aktivität von Usern beim Umgang mit ihrer Umwelt erscheint es nochmals zwingend, Versuche der „menschenge-rechten“ Gestaltung von Schnittstellen zwischen Mensch und Computer einer möglichst va-liden Untersuchung mit realen Usern in einem realen Umfeld zuzuführen.

2. 2. 7. 2. *usability*-Ziele

Die Ziele der *usability*-Forschung können in einer zweistufigen Hierarchie dargestellt wer-den, *universal usability* (Shneiderman, 2000) und produktbezogene *usability* (Nielsen, 1993)

universal usability

Mit dem Konzept der *universal usability*, das Ben Shneiderman (2000) in einem Grundsatz-artikel aus dem U.S. Communications Act von 1934 ableitet (dort „universal access“), kann eine Gruppe von Metazielen der *usability*-Forschung definiert werden, die die Umsetzung des in diesem Gesetz festgelegten Demokratieprinzips des universellen Zugangs zu Infor-mationen zum Ziel haben. Shneiderman (Shneiderman 2000) nennt als Ziel die folgenden drei Gesichtspunkte, die sich im besonderen auf den Zugang zu Informationen im Internet beziehen:

- technology variety,
- user diversity,
- [die Überbrückung von] gaps in user knowledge.

Mit der Berücksichtigung der *technology variety* soll erreicht werden, daß Personen mög-lichst unabhängig von der verwendeten Software, der Qualität des Netzzuganges und der Leistungsfähigkeit des Computers in gleicher Weise Zugang zu den Informationen erhalten sollen. Berücksichtigt man die schnelle Entwicklung immer leistungsfähigerer Computer und die daraus resultierende Bandbreite vorhandener Geräte unterschiedlichster Ausstat-tung und Leistung, wird deutlich, welche enormen Konsequenzen diese Forderung bereits impliziert.

Den Aspekt *user diversity* formuliert Shneiderman selbst so: „Accommodating users with different skills, knowledge, age, gender, disabilities, disabling conditions (mobility, sun-

light, noise), literacy, culture, income, and so forth.“ (Shneiderman 2000, S. 86). Auch die Berücksichtigung dieses Aspekts ist keineswegs einfach. Insbesondere muß versucht werden, das Problem zu lösen, daß erfahrene Benutzer/Benutzerinnen andere Unterstützung benötigen als unerfahrene; allerdings sollte die Lösung nicht so aussehen, daß einfach versucht wird, den kleinsten gemeinsamen Nenner zu finden. Vielmehr gestatten es Computersysteme in ungleich besserer Weise als traditionelle Lern- und Informationssysteme, adaptive Umgebungen herzustellen, die es tatsächlich gestatten, unterschiedlichste Usergruppen in gleicher Weise zu fördern. Hinsichtlich dieses Zieles könnte sich der Computer und das Internet als überaus fruchtbares Forschungsfeld erweisen.

Als dritten Gesichtspunkt nennt Shneiderman *gaps in user knowledge*. Das Ziel besteht darin, die Lücken im Wissen des Users zu schließen. Problematisch wird dies, wenn man bedenkt, daß es sicherlich mindestens so viele individuelle Lücken gibt, wie User des Systems, d.h. man sollte bei der Bearbeitung dieses Zieles wiederum die kognitive Produktivität des Menschen in Rechnung stellen. Eine zusätzliche Problematik bzw. eine zusätzliche Herausforderung erwächst aus dem Umstand, daß die Erfüllung dieses Zieles in einem dynamischen Prozeß zu leisten ist, d.h. die Lücken im Wissen des Users ändern sich mit der Benutzung des Systems. Auch an dieser Stelle zeigt sich das enorme Potential von computerbasierten Anwendungen im Vergleich zu traditionellen Medien, da Computerprogramme theoretisch ein Maximum an Flexibilität und Prozeßorientierung gestatten.

Produktbezogene usability

Im Hinblick auf konkrete Design- und Entwicklungsprozesse hat Nielsen (1993) 5 Prinzipien formuliert:

- learnability,
- efficiency,
- memorability,
- errors,
- satisfaction.

Auch wenn sich diese Entwicklungsprinzipien in die Hierarchie der zuvor dargestellten Prinzipien von Shneiderman (2000) subsumieren lassen, ergeben sich auch interessante Bereiche, in denen keine vollständige Überlappung festgestellt werden muß. So taucht im Sy-

stem von Shneiderman das Konzept der „Satisfaction“ an keiner Stelle auf, andererseits fehlt bei den Prinzipien Nielsens die Problematik unterschiedlicher technischer Gegebenheiten. Zumindest letzteres läßt sich auf den Fokus der Klassifikationsversuche zurückführen: Shneiderman formulierte seinen Prinzipien im Hinblick auf Informationssysteme im Internet, während Nielsen eher allgemein den Prozeß der Softwareergonomie im Blick hatte.

2. 2. 7. 3. Kriterien für „usable systems“

Die Prinzipien benutzerfreundlicher Systeme (Nielsen, 1993) sollen hier nur kurz aufgeführt werden. Im Rahmen der Diskussion (7) wird auf einige der genannten Prinzipien noch näher eingegangen werden.

- simple and natural dialogue,
- speak the User's Language,
- minimize User memory load,
- consistency,
- feedback,
- shortcuts,
- clearly marked exits,
- good error messages,
- prevent Errors,
- help and Documentation.

2. 2. 7. 4. usability-testing

Um schließlich feststellen zu können, ob die Umsetzung der genannten Prinzipien in einer aktuellen Systemimplementation als geglückt gelten kann, sind kontrollierte Techniken des *usability-testing* anzuwenden.

Für die Auswahl der richtigen Prozedur müssen eine Reihe von Faktoren berücksichtigt werden: Kosten, Relevanz und Einsatzgebiet der Anwendung, Zeitaufwand, Zuverlässigkeit der Ergebnisse und schließlich das gewünschte Maß der Generalisierbarkeit der Ergebnisse.

Shneiderman (1998) diskutiert folgende Liste der möglichen Untersuchungsdesigns:

- expert reviews,
- *usability* testing and laboratories,
- surveys,
- acceptance tests,
- evaluation during active use,
- controlled psychological oriented experiments.

Je nach Ziel und verfügbaren Mitteln sollte eines dieser Designs ausgewählt werden. Zur Beurteilung der Vor- und Nachteile kann auf die umfangreichen diesbezüglichen Erkenntnisse der psychologischen Forschung rekurriert werden (vgl. Abschnitt 4.2).

2. 2. 7. 5. Zusammenfassung

Usability spielt offensichtlich sowohl auf einer globalen, ja fast politischen Ebene eine Rolle als auch - naheliegenderweise - für die Beurteilung bestimmter qualitativer Dimensionen einer Anwendung. Ob diese Kriterien jeweils gut umgesetzt oder überhaupt eingehalten wurden, kann sowohl über Experten-Beurteilungen als auch in Form von Befragungen der Anwender/Anwenderinnen eines Systems (und zwar nicht nur mit reaktiven Verfahren, wie z.B. Fragebögen) erfasst werden. Zusätzlich gibt es die Möglichkeit, operationale Kriterien abzuleiten, die als Maß objektiver *usability*-Komponenten einer Messung zugänglich gemacht werden können.

Idealerweise sollten alle drei Säulen der Erhebung von *usability* immer ergänzend angewendet werden aus forschungsökonomischen Gründen dürfte dies allerdings selten tatsächlich umzusetzen sein.

In der vorliegenden Untersuchung wird - ebenfalls aus Ökonomiegründen - nur ein eingeschränktes Spektrum der möglichen Datengewinnungsverfahren zur *usability* zum Einsatz kommen (nähere Ausführungen dazu siehe unter 4.4.5). Wünschenswert wäre eine zusätzliche nonreaktive Erhebung z.B. in der Form von Interviews, um abschätzen zu können, ob die Kriterien der standardisierten Verfahren und die abgeleiteten Maße objektiver *usability* tatsächlich alle relevanten Dimensionen abbilden können.

Anstelle von Experten/innen-ratings zur Einhaltung der Aspekte produktbezogener *usability* ist deren Umsetzung oben im Rahmen der Vorstellung der „Lösungsversuche“, die zugleich die experimentellen Bedingungen darstellen, diskutiert worden (vgl. Abschnitt 2.2.6).

Die Beurteilung bestimmter Aspekte der *universal usability* (s.o.) hingegen kann in der Tat über abgeleitete operationale Kriterien (vgl. insbesondere 6.15) und im Rahmen einer Gesamtbeurteilung der Erhebungen erfolgen.

3. Fragestellung und Hypothesen

3.1 Fragestellung der Untersuchung

Mit der vorliegenden Untersuchung soll versucht werden, eine Antwort auf die Frage zu finden, ob durch geeignete Hilfsmittel die Bildung eines mentalen Modells in Bezug auf eine komplexe Informationslandschaft („information landscape“) gefördert werden kann und infolgedessen verschiedene Arten von Suchaufgaben mit höherer Effizienz und subjektiver Zufriedenheit gelöst werden können.

3.2 Hypothesen

3.2.1. Hypothese zum Einfluß der angebotenen Navigationshilfen

Es wird vermutet, daß die eingesetzten Hilfsmittel zur Visualisierung einen positiven Einfluß sowohl auf die Performanz als auch auf die Anzahl richtig gelöster Antworten im Umgang mit komplexen Informationen haben. Dabei wird angenommen, daß diese Wirkung umso stärker ausfallen wird, je deutlicher die dem Hilfsmittel zugrundeliegende Visualisierungstechnik die Bildung eines mentalen Modells fördern kann. Konkret soll die Annahme überprüft werden, ob die Bildung eines möglichst adäquaten mentalen Modells durch die Präsentation geeigneter Metaphern (Dutke, 1994; McDonald & Stevenson, 1998; Rautenberg, 1995; Jih & Reeves, 1992) gefördert kann. Dabei wird davon ausgegangen, daß die stärksten Verbesserungen von der Anwendung *Hyperbolic-Tree* ausgehen werden, gefolgt von *WebToc*.

Zusätzlich besteht die Annahme, daß sich der Einfluß der Navigationshilfen stärker zeigen wird, wenn für die Beantwortung der Fragen komplexe Suchoperationen und eine Orientierung über die Gesamtstruktur der angebotenen Informationen erforderlich sind (vgl. Abschnitt 2.2.5).

Desweiteren wird angenommen, daß diese Zunahme der Effektivität im Umgang mit komplexen Informationen mit einer besseren Beurteilung der Programmumgebung hinsichtlich deren *usability* einhergeht. Auch hier wird - nicht zuletzt aufgrund der Annahme einer Korrelation von objektiver Leistung und subjektive Beurteilung der dazu verwendeten Programmumgebung - davon ausgegangen, daß User des *Hyperbolic-Tree* die höchsten Werte für *usability* angeben werden, gefolgt von *WebToc*-Anwendern/Anwenderinnen und schließlich den Versuchspersonen in der Gruppe *ohne Hilfe*.

3.2.2. Hypothesen zu Kovariaten

3. 2. 2. 1. *Einfluß des Faktors spatial ability*

Es wird zusätzlich angenommen, daß Personen mit stärker ausgeprägtem räumlichen Vorstellungsvermögen (*spatial ability*) bei der Bearbeitung solcher Fragen, zu deren Lösung die Betrachtung übergeordneter Strukturen der dargebotenen Informationen hilfreich sein kann, gegenüber Personen mit schwach ausgeprägtem räumlichen Vorstellungsvermögen im Vorteil sind (vgl. 2.1.7. "Zusammenhang von *spatial ability* und Navigationskompetenz").

3.2.3. Interaktionshypothesen

3. 2. 3. 1. *Interaktion von Fragentyp und Navigationshilfe*

Es wird davon ausgegangen, daß der unter 3.2.1 formulierte Effekt der Navigationshilfe im Zusammenhang mit Struktur-Fragen in stärkerem Maße zu beobachten sein wird als bei der Bearbeitung von Retrieval-Fragen, da bei der Beantwortung von Struktur-Fragen in stärkerem Maße Wissen über die Position von Informationen in Relation zu anderen Informationen und zur Gesamtstruktur herangezogen werden müssen.

4. Methode

4.1 Skizzierung der empirischen Vorgehensweise

Es ist vorgesehen, einen einheitlichen Versuchs-Hypertext so zu modifizieren, daß er in den drei UV-Stufen mit den jeweiligen Navigationshilfen zur Bearbeitung angeboten werden kann. Die dynamische räumliche Karte kann dabei in einem eigenen Fenster aufgerufen werden. Die Nutzung dieser Karte ist allerdings für die Navigation im Hypertext nicht obligatorisch.

Die Versuchspersonen sollen den verschiedenen Navigationsvarianten randomisiert zugeteilt werden. Die Erhebung der *spatial ability*, der *usability* und weiterer Zusatzinformationen soll im Anschluß an die Beantwortung der Navigationsaufgaben ebenfalls am Bildschirm erfolgen.

4.2 Generelle methodische Probleme von Online-Untersuchungen

Da die spezifischen Vor- und Nachteile von Untersuchungen im World-Wide-Web sicherlich noch nicht zum Standardrepertoire empirischer Forschung gehören, soll hier kurz auf die generelle Problematik derartiger Forschung eingegangen werden.

Online-Forschung ist eine sehr junge psychologische Disziplin, die ersten Experimente wurden 1995 durchgeführt (Reips, 2000). Hinzu kommt die Tatsache, daß das Internet selbst eine vergleichsweise junge Errungenschaft des Computerzeitalters darstellt und zudem einem enormen Wachstumsprozeß unterworfen ist, der sowohl bezüglich der Userpopulation als auch hinsichtlich der Inhalte und nicht zuletzt entsprechend der zugrundeliegenden Technik ständigem Wandel unterworfen ist - ein denkbar ungünstiges Feld aus methodologischer Perspektive. Hinzu kommen die überaus geringen Möglichkeiten der versuchsplanerischen Kontrolle und die vielschichtige Problematik selbstselektierender Stichproben (Reips, 2000; Bandilla, 1999).

Andererseits stellt die Möglichkeit, psychologische Untersuchungen im Internet durchzuführen, nicht nur eine interessante Herausforderung dar, sondern beinhaltet vor allem auch vielfältige Vorteile gegenüber herkömmlicher Forschung, insbesondere dann, wenn man den Vergleich nicht im Hinblick auf theoretische Konzeptionen herkömmlicher Forschung, sondern im Hinblick auf die gängige Praxis vollzieht. Vor diesem Hintergrund wird deutlich, daß die Problematik selbstselektierender Stichproben unter Umständen weniger Einfluß auf die Interpretierbarkeit der Ergebnisse haben könnte als die Tatsache, daß rund 80%

aller psychologischer Studien mit Studierenden der Psychologie durchgeführt wurden (Smart, 1966, zit. n. Reips, 2000).

Web-Experimente gestatten es hingegen, mit sehr geringem Kostenaufwand in kurzer Zeit relativ große Stichproben zu erlangen. Reips (2000) nennt weitere Vorteile, wovon viele den Aspekt der externen Validität betreffen:

- bessere Generalisierbarkeit aufgrund von Untersuchungssituation und Untersuchungszeit
- weniger Störvarianz aufgrund künstlicher externaler Motivation (z. B. durch Versuchspersonenstunden)
- größere statistische Power aufgrund größerer Stichproben
- geringere Störeffekte durch Versuchsleitereinflüsse (keine Interaktion zwischen Vp und VL möglich)
- höhere Transparenz durch generelle Zugangsmöglichkeit zum Experiment
- bessere Möglichkeiten zur Kontrolle der Einhaltung ethischer Standards
- bessere Generalisierbarkeit auf die Gesamtbevölkerung (im Vergleich zur Stichprobe aus der Gruppe der Psychologiestudierenden)

Unabhängig von diesen offensichtlichen Vorteilen fallen die methodischen Probleme doch so stark ins Gewicht, so daß die psychologische Online-Forschung der letzten fünf Jahre sich noch stark auf die Überprüfung empirisch gesicherter Ergebnisse im Online-Experiment konzentriert hat. In vielfältigen Untersuchungen konnte auf diesem Wege die Validität psychologischer Online-Forschung im Internet belegt werden. So listen Krantz & Dalal (2000) allein neun experimentelle und vier Korrelationsstudien auf, die sich diesem Aspekt der neuen Forschungsrichtung gewidmet haben. Sie kommen zu dem Ergebnis, daß „in all cases, there seems to be a surprising match between laboratory and Web versions of surveys, scales, and experimental variables“ (S. 56). In jüngeren Studien konnte z.B. selbst unter sehr restriktiven Bedingungen die psychometrische Äquivalenz der Ergebnisse von Paper and Pencil und Online-Fassung bestimmter diagnostischer Instrumente belegt werden (Richter, Naumann & Noller, 1999; Naumann, Richter & Noller, 2000).

Auch wenn die angeführten allgemeinen Gründe bereits ausreichend wären, um die Durchführung von Internet-Experimenten zu rechtfertigen, bestehen im Zusammenhang mit der vorliegenden Fragestellung noch spezifische Argumente für diese Untersuchungsform.

Bei der überwiegenden Zahl bisherigen Online-Untersuchungen handelte es sich um Fragestellungen, die nicht primär mit Techniken oder Inhalten des Internets zu tun hatten. Im Gegensatz dazu geht es in der vorliegenden Untersuchung ja gerade um die Beurteilung von Navigationshilfen für die Arbeit im und mit dem World-Wide-Web.

Eine solche Fragestellung vor dem Hintergrund der vorgebrachten Vorteile der Online-Forschung „ins Labor zu verfrachten“ wäre also weder inhaltlich/forschungslogisch noch aus methodologischer Perspektive geboten.

4.3 Untersuchungsmaterial

Das Untersuchungsmaterial sollte mehrere Kriterien möglichst optimal erfüllen. Unter dem Gesichtspunkt der ökologischen Validität sollte eine tatsächlich existierende Web-Site herangezogen werden. Ein weiteres Kriterium war der Umfang der Informationen. Hier sollte eine möglichst umfangreiche und komplexe Seite mit strukturierten Informationen zu einem Themenbereich verwendet werden. Die Struktur der Informationen sollte eine hierarchische Gliederung zulassen. Da neue Ansätze der Visualisierung und der Unterstützung von Navigation gegen die bisher gebräuchliche Art der Navigation im Internet getestet werden sollten, war zusätzlich die Art der Vernetzung und der angebotenen Navigationshilfen in der Untersuchungs-Seite von Bedeutung. Diese sollten möglichst mit den Standardmöglichkeiten der HTML-Sprache auskommen, und nicht bereits selbst neuartige (oder sehr eigenständige) Konzepte umzusetzen versuchen. Schließlich war es zusätzlich von Bedeutung, eine Site zu finden, die thematisch auf breites Interesse stoßen würde.

Nach umfangreichen Recherchen wurde schließlich die Web-Site des deutschen Bundestages (<http://www.bundestag.de>) für geeignet befunden und für die Untersuchung aufbereitet. Die Wahl fiel auf diese Site, weil sie über eine Navigationsstruktur verfügt, die den genannten Anforderungen entspricht, der Umfang ausreichend ist, und nicht zuletzt, weil die zu Beginn genannte (siehe Kapitel 1) emanzipatorische Perspektive der gesamten Untersuchung an einer solchen Site, die der Information der Bundesbürger/innen über die Aktivitäten eines wichtigen politischen Organs dient, besonders deutlich werden kann.

Da die gesamte Site sich allerdings als zu umfangreich erwies (>13.000 Seiten) und die anvisierten Navigations-Tools nicht in der Lage waren, eine derart große Informationsmenge zu verarbeiten, wurden bestimmte Teilläste aus der bestehenden Struktur für die Zwecke der Untersuchung gelöscht und auf diese Elemente verweisende Querverweise aus anderen Be-

reichen der Site ebenfalls entfernt (z. B. die englische Version der Site, die privaten Homepages der Abgeordneten usw.).

Schließlich resultierte eine Version mit 3474 einzelnen Dateien (html-Seiten) und insgesamt 2520 Graphik-Dateien. Interaktive Elemente der Original-Site, die eine Kommunikation mit einem speziellen Programm auf dem Webserver benötigen, wurden ebenfalls entfernt, da eine Nachbildung dieser Programmstrukturen aus urheberrechtlichen und technischen Gründen nicht möglich gewesen wäre. Auf jeder der verbleibenden Seiten wurde eine Fußzeile eingefügt, in der darauf hingewiesen wurde, daß die jeweilige Seite zu Forschungszwecken kopiert wurde. Abgesehen von diesen Änderungen wurden keine weiteren Modifikationen an den Seiten vorgenommen.

4.4 Operationalisierung der Hypothesen - verwendete Erhebungsverfahren

4.4.1. UV1: Präsentation. Unterschiedliche Navigationshilfen

4. 4. 1. 1. Klassische Präsentation mit statischer Überblicksseite und einer Navigationsleiste

In dieser Faktorstufe *ohne Hilfe* wird den Vpn die WebSite des deutschen Bundestages ohne weitere Modifikationen präsentiert. D.h., es stehen nur die Navigationshilfen zur Verfügung, die von den Seitenbetreibern selbst vorgesehen wurden (vgl. Abbildung 3). Die von den Betreibern dieser Web-Site vorgesehenen Navigationshilfen sind in den Kontext der jeweiligen Seite eingebettete Links und eine kleine Navigationsleiste am oberen Rand jeder Seite, die wiederum Links zu den Oberkategorien der Startseite anbietet.



Abbildung 3: Startseite des Experiments in der Faktorstufe „ohne Hilfe“

4. 4. 1. 2. Lineares Inhaltsverzeichnis (realisiert mit WebToc)

In dieser Faktorstufe *WebToc* wird zusätzlich zu den Informationen der Bundestags-Web-Site ein weiteres Fenster am linken Bildschirmrand eingeblendet, in welchem mit *WebToc* (Nation, Plaisant, Marchionini, & Komlodi, 1997) ein hierarchisches Inhaltsverzeichnis der Site dargestellt wird (vgl. Abbildung 4). Innerhalb dieses Inhaltsverzeichnisses können Kapitel auf- und zugeklappt werden, außerdem gibt es die Möglichkeit, einen Suchbegriff in ein Textfeld einzugeben. Wenn dieser Begriff auf einer Ebene des Inhaltsverzeichnisses gefunden werden kann, wird die entsprechende Ebene automatisch „ausgeklappt“. Alle Einträge dieses Verzeichnisses repräsentieren Links, die auf Seiten der Bundestags-Web-Site verweisen. Wenn dementsprechend auf einen Eintrag zweifach geklickt wird, wird im rechten Fenster die zugehörige Seite im Volltext geladen.

Zusätzlich zu diesen Navigationsmöglichkeiten bietet das Programm weitere Informationen in Form verschiedenfarbiger Balken an. Hier werden sowohl die Anzahl von Dateien in ei-

nem (Unter-)Verzeichnis als auch die Dateigröße aller Dateien im jeweiligen (Unter-)Verzeichnis angezeigt.

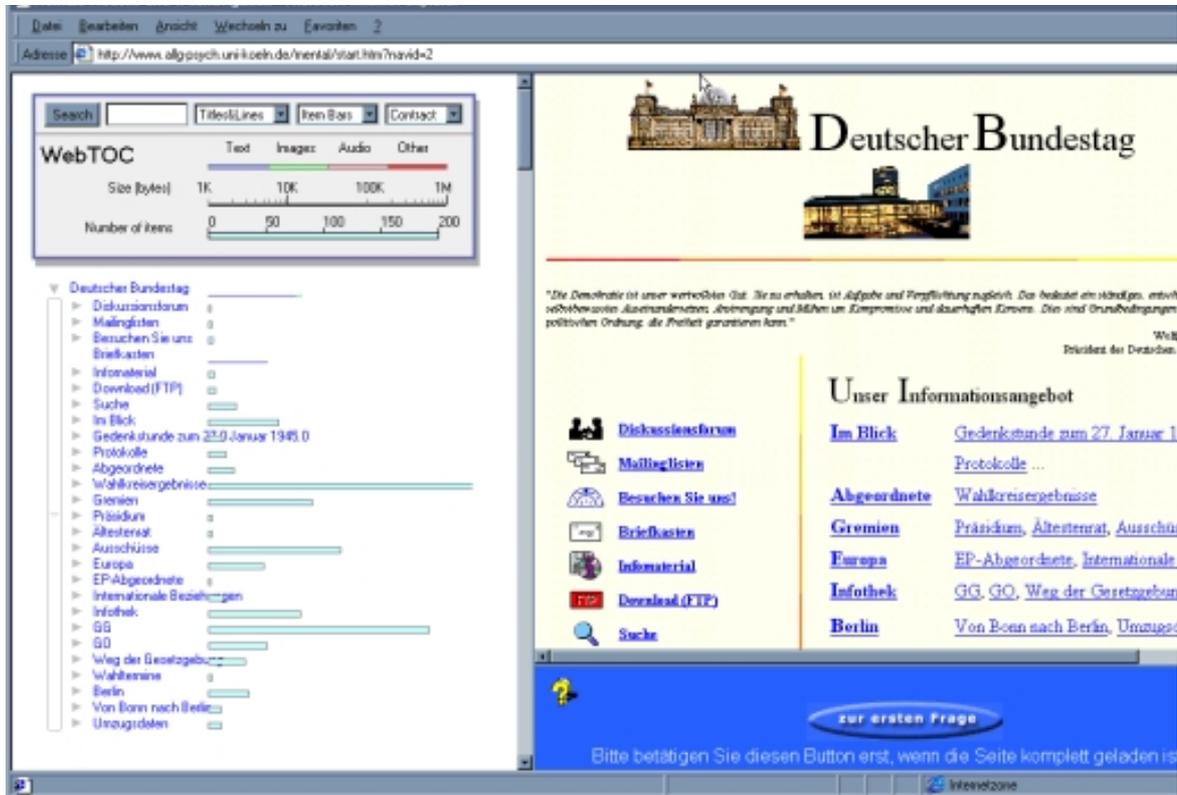


Abbildung 4: Startseite des Experiments in der Faktorstufe WebToc

4. 4. 1. 3. Dynamischer graphischer Browser (realisiert mit Hyperbolic-Tree)

In der dritten Faktorstufe *Hyperbolic-Tree* wird den Vpn eine Visualisierung der Inhalte der Bundestags-Web-Site angeboten, die nur noch wenig mit derjenigen der ursprünglichen Site gemein hat (vgl. Abbildung 5). Mit dem *Hyperbolic-Tree* (Lamping & Rao, 1996; Lamping, Rao & Pirolli, 1995) wird dem Nutzer eine Visualisierung der Informationen in der Form eines graphischen Browsers angeboten (vgl. 2. 2. 6. 3. "Hyperbolic-Tree-Browser").

Innerhalb dieses „Baumes“ können einzelne Knoten fokussiert werden. Auch dieses Programm beinhaltet eine Suchfunktion, bei der die Treffer farblich in der Baumdarstellung hervorgehoben werden.

Die Interaktion mit dem Textbestand der Bundestag-Web-Site wird in dieser Bedingung dann hergestellt. Wenn der Benutzer auf einen Knoten (Link) doppelklickt wird die zuge-

2.,,Wie hoch war die Wahlbeteiligung (Bundestagswahl) im Wahlkreis Berlin-Reinickendorf 1998?“

3.,,Wie lautet die Abkürzung für die Parlamentarische Versammlung der Nato?“

Diese Fragen wurden so formuliert, so dass die Antworten auf jeweils einer Seite der Site gefunden werden konnten. Die Informationen konnten durch einfache Retrieval-Techniken und direkte Suche in der angebotenen Site-Struktur aufgefunden werden.

4. 4. 2. 3. AV1b: komplexe Struktur-Fragen

Zur Operationalisierung dieses Elements wurden 3 Fragen formuliert:

1.,,Welches Bundesland stellt als nächstes den/die Bundesratspräsidenten/in?“

2.,,Ist in der Geschäftsordnung des Bundestages etwas zu Diäten geregelt?“

3.,,Welche der Parteien hat den höchsten Frauenanteil im Bundestag?“

Zur Beantwortung dieser Fragen waren komplexe Vergleichs- und Suchprozesse über große Teile der Site notwendig. So bestand z.B. die Schwierigkeit bei Frage 1 darin, daß die Site hauptsächlich Informationen zum/zur Vorsitzende/n des Bundestages anbot, nicht aber zur entsprechenden Person des Bundesrates. Dadurch war die Möglichkeit einfacher Retrieval-Techniken eingeschränkt, und die Lösung der Frage konnte erheblich vereinfacht werden, wenn Informationen der Site-Struktur hinzugezogen wurden

Für Frage 2 bestand die Schwierigkeit darin, daß die gesuchte Information nur im Ausschlußverfahren gewonnen werden konnte. Die richtige Antwort lautete "Nein", aber es konnten dennoch Fundstellen zur Frage der Diätenregelung erzielt werden. Für die richtige Beantwortung der Frage war dann entscheidend, daß es der Versuchsperson gelang, diese Fundstellen in der Hierarchie der Sitestruktur zu lokalisieren, um zu erkennen, daß sämtliche Fundstellen außerhalb des Zweiges der Geschäftsordnung gelegen waren.

Auch für Frage 3 war die Antwort schnell zu finden, wenn Organisationsmerkmale der Informationen in der Gesamt-Site in die Suche einbezogen wurden.

4.4.3. AV2: Eigenschaften des Navigationspfades (aus Logfile-Analysen)

4. 4. 3. 1. Vorbemerkung: Einschränkungen

Aus technisch-methodischen Gründen muß hinsichtlich dieser AV eine Vorbemerkung gemacht werden.

Das wesentliche Merkmal von Logfile-Daten in einer psychologischen Untersuchung ist die vollständige Aufzeichnung der Aktivitäten des Users. Zwar ist es so, daß die meisten der häufig genannten Probleme von Logfile-Daten im Zusammenhang mit Online-Studien (Catledge & Pitkow, 1995; Pitkow, 1997; Pirolli, Pitkow & Rao, 1996; Cooley, Bamshad & Srivastava, 1999) bei dieser Untersuchung durch eine spezielle Technik (Session-Tracking) ausgeschaltet werden konnten und die Logfiles eindeutig einzelnen Usern zugeordnet werden können. Allerdings verursacht das Design der Untersuchung ein Problem, welches die Validität der erhobenen Daten z.T. erheblich einschränken dürfte. Das Problem besteht darin, daß (aus technischen Gründen) die Aktionen des Users innerhalb des jeweiligen Hilfsprogrammes (*WebToc* oder *Hyperbolic-Tree*) nicht aufgezeichnet werden konnten. Deshalb kennzeichnen in diesen Bedingungen die im Logfile resultierenden Daten nur den Teil der Aktionen des Users, der sich auf die Interaktion mit den Volltext-Seiten der Bundestags-Web-Site beziehen.

Unter Berücksichtigung dieser Einschränkung können die Daten allerdings dennoch zur Interpretation herangezogen werden, insbesondere für die Faktorstufe 1 der UV *Navigationshilfe*, da hier vollständige Logfiles vorliegen.

Um wenigstens ansatzweise zusätzlichen Aufschluß über das Ausmaß der nicht aufgezeichneten Aktionen innerhalb des jeweiligen Tools zu erhalten, wurde protokolliert, wann ein Zugriff auf eine Seite ausserhalb des Tools von diesem aus erfolgte.

4. 4. 3. 2. AV2a: Anzahl besuchter Knoten

Bei dieser AV wird die Anzahl besuchter Knoten über die ganze Untersuchung zur Analyse herangezogen. Da die Bestandteile „Einführung“ und die diversen Seiten zur Abschlußbefragung sowie zur Erhebung der *usability* hinsichtlich der Anzahl besuchter Knoten invariant sind, kann die Varianz dieses Gesamtwertes auf die Phase der Untersuchung während der Beantwortung der Treatment-Fragen zurückgeführt werden.

4. 4. 3. 3. AV2b Anzahl vom Tool aus aufgerufener Knoten

Zusätzlich zur Gesamtzahl besuchter Knoten soll die Anzahl der Knoten gezählt werden, die vom jeweiligen Hilfsprogramm aus aufgerufen wurden. Da diese Möglichkeit nur für die Stufen *WebToc* und *Hyperbolic-Tree* des Faktors *Navigationshilfe* bestand, ist dieses Maß eher zu exploratorischen Zwecken zu verwenden.

4. 4. 3. 4. AV2c: Stratum

Zur Charakterisierung der Linearität eines Hypertextes wurde das aus der Graphentheorie entlehnte Maß des „Stratums“ verwendet (de Bra & Houben, 1994; de Vocht, 1994; Botafogo, Rivlin, & Shneiderman, 1992). McEneaney (1999) modifizierte die zugrundeliegenden Berechnungen dergestalt, daß das Verfahren auch zur Beschreibung einzelner Pfade von Usern in einem Hypertext herangezogen werden kann. Mit dieser Maßzahl kann erfaßt werden, inwieweit der zugrundeliegende Graph (d.h. Logfile eines Users) aus linearen Sequenzen besteht.

Um das Stratum zu errechnen, wird aus der Sequenz der einzelnen Seitenaufrufe in der Logfile eine Distanzmatrix erstellt, die die Anzahl der Seitenaufrufe beinhaltet, die benötigt werden, um von einer Seite des Hypertextes/Pfades zu einer anderen zu gelangen.

4. 4. 3. 5. AV2d: Compactness

Ebenfalls von Botafogo et al. (1992) und McEneaney (1999) wurde aus der Matrixrechnung das Maß der „Compactness“ als Indikator für die Vernetztheit eines Hypertexts, bzw. eines Pfades durch einen solchen vorgeschlagen.

Auch dieses Maß wird aus einer anhand der in den Logfiles gegebenen Sequenzen gebildeten Distanzmatrix abgeleitet.

4. 4. 3. 6. AV2e Aufruf bestimmter Seitentypen

Mit einem speziellen Analyseprogramm (*Log-Pat*; Noller, 2000) soll anhand der Logfiles das Auftreten bestimmter Seitentypen, und zwar solcher von Typ „Index-Seiten“ und vom Typ „Such-Seiten“, gezählt werden.

4.4.4. AV3: Lösungsgüte

4. 4. 4. 1. AV3a: Anzahl richtiger Antworten

Alle für den Faktor „Fragentyp“ gebildeten Suchfragen lassen sich eindeutig beantworten. Die Antworten werden überprüft und für jede Vp geht die Anzahl richtiger Antworten in die Zählung ein.

4. 4. 4. 2. AV3b: Bearbeitungsdauer

Für jede der sechs Fragen des Faktors „Fragentyp“ wird individuell für jede Vp die Bearbeitungszeit erfaßt.

4.4.5. AV4: usability

Zur Erhebung der subjektiven Einschätzungen von *usability* wurde mit dem *Computer System Usability Questionnaire (CSUQ; Lewis, 1995)* ein standardisiertes Verfahren herangezogen.

Dieses Instrument wurde speziell zur Erhebung der subjektiven Einschätzungen von *usability* in einer Scenario-basierten Untersuchungssituation entworfen. Der Fragebogen besteht in der ursprünglichen Fassung aus 19 Items, die auf einer siebenstufigen Skala mit den Polen „strongly agree“ und „strongly disagree“ beantwortet werden sollen.

Lewis (1995) berichtet eine 3-faktorielle Struktur des Instruments, mit den Faktoren „System-Usefulness“, „Information-Quality“ und „Interface-Quality“.

Das Instrument verfügt über sehr gute Kennwerte bezüglich der Reliabilität (Alpha: Gesamtskala=.97; Faktor „System-Usefulness“=.96, Faktor „Information-Quality“=.91, Faktor „Interface-Quality“=.91). Aus diesem Grund erscheint es relativ unproblematisch, das Instrument auf die Anforderungen der Untersuchungssituation anzupassen und auf einige Items zu verzichten.

Für eine Verwendung im Rahmen der vorliegenden Untersuchungen waren in zweifacher Hinsicht Modifikationen vonnöten: Die Items mussten ins Deutsche übersetzt werden und nach Möglichkeit sollte aus dem gegebenen Itempool eine Auswahl getroffen werden, um die Gesamtbearbeitungszeit der Untersuchung nicht zu sehr anwachsen zu lassen.

Bezüglich der Auswahl konnten zunächst solche Items ausgeschlossen werden, die sich mit Aspekten der Programmschnittstelle befassen, welche in der hier untersuchten Umgebung

nicht vorhanden sind (z. B. Items die sich auf die Qualität der Fehlermeldungen des Systems bezogen), wie auch Items, die Aspekte der Interaktion mit dem Programm zu erfassen versuchen, die aufgrund der speziellen Erhebungssituation in dieser Untersuchung nicht relevant waren (z.B. Lerneffekte).

Es resultierte eine Fassung mit elf Items, die ins Deutsche übersetzt wurden (siehe Anhang C).

4.4.6. Kovariate: *Spatial Ability*

Da *spatial ability* bzw. „räumliches Vorstellungsvermögen“ schon seit vielen Jahren in der psychologischen Forschung, insbesondere im Zusammenhang mit der Messung von Intelligenz, erfaßt wird, erschien es auch hier sinnvoll, auf ein standardisiertes Verfahren zurückzugreifen.

Die Wahl fiel auf einen Subtest des Wilde-Intelligenz-Test (Jaeger & Althoff, 1983), und zwar die Skala „Abwicklungen“. Diese wird dem Faktor „Räumliches Vorstellen“ zugeordnet.

Auch hier mußte aus untersuchungstechnischen Gründen eine möglichst kurze (und dennoch reliable) Variante gefunden werden. Aus diesem Grund wurde die Version des Subtest mit der höchsten berichteten Reliabilität (Form B; $\alpha=.86$) herangezogen und anhand der Spearman-Brown-Formel (Lienert & Raatz, 1998) errechnet, um wieviele Items der Subtest gekürzt werden kann, wenn als untere Grenze ein (theoretischer) Reliabilitätswert von mindestens .70 angesetzt wird.

Aufgrund dieser Berechnung wurde eine Auswahl von 8 der ursprünglichen 20 Items gebildet (vgl. Anhang E), wobei zufällig Items entlang des Schwierigkeitsgradienten ausgewählt wurden.

4.4.7. Explorative Fragen zum Experiment

Auf einer vierstufigen Skala mit den Extrempolen „trifft zu“ und „trifft nicht zu“ sollten die Vpn folgende Fragen beantworten:

- „Das Experiment hat mir gut gefallen“;
- „Es fiel mir leicht, die Fragen zu beantworten“;
- „Ich hatte manchmal das Gefühl, nicht mehr zu wissen, wo ich bin“;

- „Ich beschäftige mich viel mit Politik und war deshalb schon vorab mit der Materie vertraut“.

Hierbei sollten die ersten drei der genannten Fragen dazu dienen, weiteren Aufschluß über bestimmte Aspekte des Experiments zu erhalten. Mit der vierten Frage sollte hingegen überprüft werden, ob bereits systematisches Vorwissen vorhanden war, das die Bearbeitung der Fragen eventuell erleichtert haben könnte.

4.5 Der Versuchsplan im Überblick

Beim resultierenden Design handelt es sich um einen einfaktoriellen, multivariaten Zufallsgruppenversuchsplan.

Im Überblick sieht der Plan wie folgt aus:

Tabelle 1: Versuchsplan

UVA1 (keine Hilfen)	AV1, AV2, AV3, AV4
UVA2 (Inhaltsverzeichnis, realisiert mit <i>WebToc</i>)	AV1, AV2, AV3, AV4
UVA3 (dynamische Karte, realisiert mit <i>Hyperbolic-Tree</i>)	AV1, AV2, AV3, AV4

Anmerkungen:

UVA: Navigationshilfe; AV1: Fragentyp; AV2: Logfile-Maße (Anzahl besuchter Knoten, Compactness, Stratum); AV3: Lösungsgüte; AV4: *usability*

4.6 Stichprobe

4.6.1. Vorbemerkungen

Der Spielraum für die Beeinflussung der Zusammensetzung der Stichprobe ist im Rahmen von Online-Untersuchungen sehr eng. So stellt z. B. Bandilla (1999) dar, wie eine ganze Reihe unterschiedlicher Einflußfaktoren dazu führen, daß in Web-Untersuchungen immer nur (mehrfach) selbstselegierende Stichproben zu erzielen sind. Dieser Umstand mindert die Möglichkeiten des Rückschlusses auf die zugrundeliegende Grundgesamtheit erheblich (vgl. die Erörterung unter Abschnitt 4.2).

Infolgedessen liegt der hier vorgelegten Studie eine selbstselegierende anfallende Stichprobe zugrunde.

Um die Einschätzung der Zusammensetzung der Grundgesamtheit aller Internet-Nutzer im deutschsprachigen Raum zu ermöglichen - insbesondere auch mit Rückbezug auf die Gesamtbevölkerung - sei kurz auf die Ergebnisse der fünften Erhebungswelle des GFK-Online-Monitors verwiesen, die zu folgenden Ergebnissen kam (GFK, 2000): Bei der telefonischen Befragung einer bevölkerungsrepräsentativen Stichprobe von 8000 Personen im Alter von 15 bis 69 Jahren gaben 29.9% der Befragten an, das Internet regelmäßig zu nutzen. Hochgerechnet auf die Gesamtbevölkerung (in der genannten Altersgruppe 53 Millionen) verfügen etwas mehr als 24 Millionen Bundesbürger/Bundesbürgerinnen bereits über einen Zugang zum Internet, was einer Steigerung von über 50% innerhalb eines Jahres entspricht. Die häufig festgestellte Ungleichverteilung von Männern und Frauen besteht immer noch, scheint sich aber zunehmend zu nivellieren - in der aktuellen Erhebung waren 38.9% der Internet-Nutzer Frauen. Nach wie vor sind junge Personen mit höherem Bildungsniveau und hohem Einkommen überproportional vertreten, die Autoren/Autorinnen berichten aber auch diesbezüglich einen Trend zu einer Angleichung an soziodemographische Gegebenheiten in der Gesamtbevölkerung.

Für die psychologische Online-Forschung sind diese Trends sicherlich hocheffizient, da die Qualität der auf diesem Wege zu gewinnenden Stichproben sicherlich mit den genannten Entwicklungen steigt und damit eine Generalisierbarkeit auf die Gesamtbevölkerung plausibler wird.

Für die vorliegende Untersuchung ist dieser Rückschluß auf die Gesamtbevölkerung nicht intendiert, da es ja um eine Bewertung von Hilfsmitteln zur Informationssuche im Internet geht, die für Personen ohne Internet-Zugang keine Rolle spielen (von gewissen Offline-Einsatzmöglichkeiten z. B. in Bibliotheken abgesehen). Interessant ist im Hinblick auf diese Trends allerdings die Tatsache, daß sich das Profil der Internet-User sicherlich - insbesondere im Hinblick auf das Verhältnis von Experten/innen zu Novizen/innen - verändern wird, und daß der einleitend bereits genannte Trend zur Nutzung des Internet auch von staatlichen Organisationen mit der steigenden Nutzung und Verfügbarkeit des Netzes sich fortsetzen wird.

4.6.2. Zusammensetzung der Untersuchungsstichprobe

Die Stichprobe besteht aus 185 gültigen Datensätzen. Von den 177 Personen, die bereit waren Ihr Geschlecht anzugeben, waren 57 (30.8%) weiblichen, und 120 (64.9%) männlichen Geschlechts. Diese Zusammensetzung steht prinzipiell im Einklang mit anderen Studien

zur Zusammensetzung der Population der Internetnutzer/innen, der Frauenanteil liegt allerdings deutlich unter dem der oben zitierten aktuellen GFK-Erhebung (vgl. 4.6).

Hinsichtlich der Erfahrung mit dem Internet wurden zwei Kennwerte erhoben, die Jahre der Nutzung des Internet und die Anzahl von Stunden der Internetnutzung pro Woche.

Hier ergaben sich folgende Verteilungen: Bezüglich der Stunden der Internetnutzung pro Woche gaben die Vpn Werte im Bereich von 0 - 100 Stunden an, der Mittelwert beträgt 12.8 Stunden (SD=13.33). Hinsichtlich der Jahre der Internetnutzung liegen die Werte in einem Bereich von 0 - 96 Jahren, der Mittelwert beträgt 4.51 Jahre (SD=7.47). Der Extrem-Wert von 96 Jahren stellt allerdings sicherlich eine ungültige Angabe dar. Ohne Berücksichtigung dieses Wertes liegt der Wertebereich zwischen 0 - 16 Jahren. Aus Tabelle 2 ist ersichtlich, wie diese Stichprobenkennwerte im Vergleich mit großen aktuellen Erhebungen zur Demographie der Internetnutzer/innen aussehen.

Tabelle 2: Vergleich der Stichprobenkennwerte mit anderen Studien

Kennwert/Studie	<i>eigene Studie</i>	<i>ARD/Enigma(2000)</i>	<i>Forsa (2000)</i>
Stichprobengröße	185	1000	11003
Nutzung Internet/ Stunden proWoche	12,8	9,6	6,3
Nutzung Internet/Jahre	4,5	2,3	1,7
Geschlecht % (m/f)	67/32	65/35	68/31

Die Verteilung der verschiedenen Betriebssysteme zeigt einen Anteil von 93.5 Prozent für die Microsoft-Betriebssysteme (Windows95, Windows98, Windows-NT und Windows2000). Die Betriebssysteme Linux (3.8%) und MacOS(1.1%) nehmen nur einen marginalen Anteil ein, ebenso die Restkategorie der anderen Betriebssysteme (1.6%).

Da für diese Untersuchung vergleichbare technische Bedingungen von erheblicher Bedeutung sind, wurde auch die Bildschirmauflösung, die am Computer der jeweiligen Vpn eingestellt war, erfasst. Diesbezüglich ist die am häufigsten vertretene Kategorie eine Auflösung von 1024 x 768 Bildschirmpunkten (47%), gefolgt von einer Auflösung von 800 x 600 Punkten (38.4%). Mit deutliche geringeren Werten waren die Kategorien für höhere Auflösungen als 1024 x 768 Punkte (13%) und 640 x 480 Punkte (1.6%) besetzt.

4.7 Ablauf der Untersuchung

Die Untersuchung wurde vollständig als Online-Experiment im World-Wide-Web konzipiert. Nachdem eine Versuchsperson die Start-Seite der Untersuchung aufgerufen hatte (zur Akquise der Versuchspersonen vgl. Abschnitt 5), wurde per Zufallsgenerator eine Navigationsbedingung ausgewählt. Auf der Startseite wurde kurz das Anliegen der Untersuchung dargestellt und der institutionelle Rahmen angegeben. Wenn die (automatische) Prüfung der technischen Voraussetzungen ohne Fehler absolviert wurde (im Falle einer Verletzung der Voraussetzungen wurde eine Hinweisseite mit Instruktionen zur Behebung eingeblendet), wurde der Einführungsteil des Experiments geladen. Hier wurden kurz allgemeine Instruktionen zum Ablauf des Experiments gegeben, gefolgt von speziellen Instruktionen zur jeweiligen Navigationshilfe (siehe Anhang A). Für die beiden Stufen *Hyperbolic-Tree* und *WebToc* der Navigationsbedingung wurde an dieser Stelle zusätzlich ein kurzes einführendes Tutorial geladen und die Bedienelemente des Zusatzprogrammes erläutert.

Wenn dieser Einführungsblock absolviert war, wurde die Startseite der Web-Site des deutschen Bundestages eingeblendet. Am unteren Rand des Bildschirms erschien zusätzlich ein Fenster mit der ersten Frage. Es wurde permanent ein Schalter angeboten, bei dessen Betätigung in einem separaten Fenster nochmals die Hilfe zur Bedienung der Programmumgebung eingeblendet wurde.

Die sechs Fragen der AV „Fragentyp“ wurden in abwechselnder Reihenfolge präsentiert (Fragen 1, 3 und 5 *Retrieval-Fragen*, Fragen 2, 4 und 6 *Struktur-Fragen*).

Nachdem die sechs Fragen bearbeitet wurden, erschien ein mehrseitiger Abschlußfragebogen, der das Instrument zur Erhebung von *usability*, das Instrument zur Erhebung von *spatial ability* und einige weitere Fragen beinhaltete (in der genannten Reihenfolge).

Abschließend wurde der Versuchsperson eine individuelle Auswertungsseite präsentiert, auf der die erzielten Ergebnisse im Vergleich mit den Ergebnissen der bisherigen Versuchspersonen in der jeweiligen Bedingung dargestellt wurden.

4.8 Kontrolle von Störvariablen

Wie bereits allgemein in Abschnitt 4.2 ausgeführt, ist aufgrund des Designs der Untersuchung als Online-Experiment mit vielfachen Störeinflüssen zu rechnen.

Mögliche Effekte durch unterschiedliche technische Voraussetzungen und individuelle Einstellungen am Computer der Versuchspersonen wurden im wesentlichen durch zwei Maßnahmen kontrolliert. Zum einen wurde durch spezielle Programmieretechniken eine

standardisierte Umgebung für die Ausführung des Experiments hergestellt, d.h. das Fenster des Betrachtungsprogrammes („Browser“) wurde automatisch auf volle Größe eingestellt. Ebenso wurden bestimmte Standardnavigationselemente eingeblendet. Um die trotz dieser Maßnahmen zu erwartende Variabilität des Merkmals „Bildschirmauflösung“ zu kontrollieren, wurde diese Variable für jede Versuchsperson erfaßt. Desweiteren wurden das eingesetzte Betriebssystem und die verwendete Browser-Version abgespeichert.

Weitere Personenmerkmale, die einen systematischen Einfluß auf die zu erhebenden Variablen haben könnten, wurden per Nachbefragung erhoben, im Besonderen waren dies Daten zur Intensität der Internetnutzung, zum Geschlecht und zum thematischen Vorwissen bezüglich des Versuchsmaterials.

An einigen Stellen wäre eine Standardisierung der Bearbeitungsdauer aus methodischer Sicht wünschenswert gewesen (z. B. Einführung, Untertest für *spatial ability*). Da im Rahmen einer Online-Untersuchung aber nicht ausreichend Möglichkeit vorhanden ist, den Sinn derartiger rigider Maßnahmen zu erläutern, wurde hier auf eine derartige Beschränkung verzichtet. Stattdessen wurden die Bearbeitungszeiten für die verschiedenen Arbeitsschritte individuell differenziert aufgezeichnet und in spätere Auswertungen einbezogen.

Um einen möglichen systematischen Effekt durch Lernprozesse während der Beantwortung der Fragen zu verhindern, wurden die Fragen des Faktors „Fragentyp“ in abwechselnder Reihenfolge präsentiert.

Ein weiteres Problem im Zusammenhang mit Online-Untersuchungen stellen Störeffekte durch spezielle Anreize dar, die den Versuchspersonen für die Teilnahme an der Untersuchung in Aussicht gestellt werden (vgl. Gräf & Heidingsfelder 1999). Um derartige Effekte zu verhindern und nach Möglichkeit internal motivierte Teilnehmer/innen zu rekrutieren, wurde auf externe Belohnungen gänzlich verzichtet. Stattdessen wurde den Teilnehmern/innen ein individuelles Ergebnisfeedback in Aussicht gestellt, welches am Ende der Untersuchung auch jeweils präsentiert wurde.

Ein weiterer möglicher Störeinfluß besteht in der durch die experimentellen Bedingungen verursachten künstlichen Verkleinerung des zur Verfügung stehenden Bildschirms für die Suche und Darstellung von Informationen durch die Präsentation der zusätzlichen Navigationshilfen. Da es sich im Falle einer tatsächlichen Auswirkung dieses Effektes nur um einen zugunsten der H₀ handeln kann (weil in der Treatment-Bedingung der Bildschirm verkleinert wird), kann dieser potentielle Störeinfluß vernachlässigt werden.

5. Durchführung der Untersuchung

Die Untersuchung wurde vom 16.04.2000 bis zum 15.05.2000 im World-Wide-Web auf einem Server des Lehrstuhls für Allgemeine Psychologie und Kulturpsychologie der Universität zu Köln durchgeführt.

Die „Akquise“ von Versuchspersonen kann bei Online-Experimenten im Internet naturgemäß nur indirekt erfolgen. Zu diesem Zweck wurden Hinweise auf die Untersuchung auf einschlägigen Meta-Sites beantragt, wie z. B. in der entsprechenden Kategorie einer zentralen Site für Online-Forschung (<http://www.online-forschung.de>), oder in einer Rubrik einer zentralen Site für Psychologie (<http://www.psychologie.de>). Desweiteren erfolgte eine Ankündigung des Experimentes in verschiedenen Newsgroups, deren Zielgruppe im weiteren Sinn mit der Thematik des Experimentes in Verbindung gebracht werden konnte. Schließlich wurden nach den gleichen Kriterien diverse Mailinglisten ausgesucht und eine eMail mit weitgehend identischem Wortlaut abgesetzt; in den meisten Fällen wurde dabei nur die Anrede an die Liste angepasst, zweimal war hingegen eine Umformulierung aufgrund der Bekanntschaft des Autors in der Liste geboten (siehe Anhang B).

In dieser Zeit fanden 763 Zugriffe auf die Startseite statt, 414 Personen betätigten den ersten „Start-Button“. Insgesamt ergaben sich 185 vollständige Datensätze (zur Analyse des Dropouts siehe Abschnitt 6.15).

6. Darstellung der Ergebnisse

6.1 Vorstrukturierung

Im folgenden sollen die gewonnenen Daten im Hinblick auf die Forschungshypothesen untersucht werden. Zunächst wird die psychometrische Qualität der Skalen zur Erfassung der AV Fragentyp berichtet und entsprechende Analysen durchgeführt. Danach folgen verschiedene varianzanalytische Untersuchungen zur Untersuchung des Einflusses der Navigationshilfen auf die abhängigen Variablen. Der Schwerpunkt liegt hierbei auf den Performanz-Maßen und der Lösungsgüte. Die Analysen sollen zusätzlich multivariat und unter Einbeziehung der Kovariaten *spatial ability* vollzogen werden. Da aus untersuchungstechnischen Gründen in dieser Variablenkombination geringere Fallzahlen resultieren, werden getrennte univariate Analysen vorgeschaltet. Beziehen sich die Effekte auf mehr als 2 Faktorstufen, wird zusätzlich ein Post-Hoc-Vergleich mit einem alpha-adjustierten Verfahren (Bonferonni) durchgeführt und berichtet.

Die Daten zur AV *usability* sollen zunächst ebenfalls hinsichtlich der psychometrischen Qualität und der Reproduzierbarkeit der berichteten Skalenstruktur des Instruments untersucht werden. Es folgen varianzanalytische Berechnungen im Hinblick auf die 3 Komponenten der *usability*, die das Verfahren messen soll.

Die AV Logfilemaße setzt sich aus verschiedenen Kennwerten zusammen, für welche einzeln Analysen berichtet werden. Aufgrund untersuchungstechnischer Einschränkungen wird hier zusätzlich eine exploratorische Analyse innerhalb einzelner Faktorstufen präsentiert.

In einem weiteren Schritt erfolgt die Analyse der zusätzlichen Einzelfragen, die nach Ablauf des Experimentes zur Beantwortung präsentiert wurden.

Abschließend erfolgen eine Reihe exploratorischer Analysen zu weiteren Einflußgrößen, und zwar zur Interneterfahrung, zur Bearbeitungszeit des Einführungstutorials und zu möglichen Lerneffekten über den Verlauf der Untersuchung in Abhängigkeit der angebotenen Navigationshilfen.

Zuletzt werden Analysen zum Dropout der Untersuchung präsentiert, die sich auf einen größeren Datensatz beziehen und ebenfalls exploratorischen Charakter haben.

6.2 Technische Vorbemerkungen

Alle durchgeführten Signifikanztests werden auf einem Alpha-Niveau von 5% interpretiert, sofern nichts anders angegeben wird.

6.3 Untersuchung auf Homogenität über die Zellen

Trotz des asymmetrischen Dropouts und der daraus resultierenden ungleichen Zellenverteilung, ergibt eine Analyse über wesentliche Kennwerte zur Beschreibung der Stichprobe (Geschlecht, Jahre der Internetnutzung, Stunden Internetnutzung pro Woche, Bildschirmauflösung) keinen signifikanten Unterschied über die drei Navigationsbedingungen. Die Mittelwertsunterschiede werden auch dann nicht signifikant, wenn mit einem Alpha-Niveau von 20% getestet wird (weil in diesem Fall die H_0 die Forschungshypothese ist). Für den multivariaten Vergleich ergibt sich $F(8,166)=.80$, $p=.61$, der univariate Vergleich für die Variable *Auflösung* ergibt $F(2,166)=1.58$, $p=.21$, für *Stunden-Internetnutzung pro Woche* $F(2,166)=.61$, $p=.54$, für *Jahre-Internetnutzung* $F(2,166)=.68$, $p=.50$ und schließlich für die Variable „thematisches Vorwissen“ $F(2,166)=.16$, $p=.85$. Aus diesem Grund werden (unabhängig von der gesonderten näheren Analyse dieses systematischen Dropouts unter 6.15) die folgenden Analysen anhand des kompletten Datensatzes gerechnet.

6.4 Beurteilung der Voraussetzungen für eine varianzanalytische Untersuchung

Zu den wesentlichen Voraussetzungen einer Varianzanalyse gehören die der Varianzhomogenität und die der Normalverteilung der Daten (Bortz, 1993; Litz, 2000). Um die Erfüllung dieser Voraussetzungen zu überprüfen, wurden für die einzelnen abhängigen Variablen jeweils anhand des Kolmogorov-Smirnov Tests die Normalverteilungsannahme und anhand des Levene-Tests die Annahme homogener Varianzen überprüft.

Tabelle 3: Test der Varianzhomogenität

	Levene Statistic	df1	df2	Sig.
FRAG_RET	9,909	2	167	,000
FRAG_STR	6,722	2	182	,002
FRAG_GES	11,030	2	182	,000
DAUE_RET	13,349	2	182	,000
DAUE_STR	2,399	2	182	,094
DAUE_GES	11,208	2	182	,000
USAB_GES	,961	2	174	,385
SYS_USE	,520	2	174	,595
INFO_QUA	,600	2	174	,550
INTE_QUA	,945	2	174	,391

Anmerkungen: FRAG_RET - richtige Antworten Retrieval; FRAG_STR - richtige Antworten Struktur; FRAG_GES - richtige Antworten Gesamt; DAUE_RET - Bearbeitungsdauer Retrieval; DAUE_STR - Bearbeitungsdauer Struktur; DAUE_GES - Bearbeitungsdauer Gesamt; USAB_GES *usability* Gesamt; SYS_USE - System-Usefulness; INFO_QUA - Information Quality; INTE_QUA - Interface-Quality

Wie aus Tabelle 3 ersichtlich ist, erfüllen nur die Variablen *Strukturfragen* (DAUE_STR) und *usability* (USAB_GES) sowie die drei *usability*-Faktoren „system-usability“ (SYS_USE), „information-quality“ (INFO_QUA) und „interface-quality“ (INTE_QUA) die Voraussetzung der Varianzhomogenität.

In ähnlicher Weise werden bei einem Teil der Variablen die Voraussetzungen bezüglich der Verteilungsform der Daten nicht erfüllt. Die Ergebnisse aus Tabelle 4 zeigen, daß nur für die Werte für Gesamtbearbeitungsdauer (DAUE_GES) und *Strukturfragen* (DAUE_STR) die Annahme der Normalverteilung aufrechterhalten werden kann.

Tabelle 4: Test der Normalverteilungsannahme: Bearbeitungsdauer/Anzahl richtiger Fra-

One-Sample Kolmogorov-Smirnov Test

		FRAG_R ET	FRAG_ST R	FRAG_G ES	DAUE_R ET	DAUE_ST R	DAUE_ GES
N		170	185	185	185	185	185
Normal Parame- ters	Mean	2,4882	1,7568	4,0432	228,6541	306,9081	535,56
	Std. Deviation	,7556	1,1084	1,8469	151,8266	189,6277	297,17
	Most Extreme Difference						
	Absolute	,357	,225	,217	,120	,058	,079
	Positive	,249	,144	,145	,120	,058	,079
	Negative	-,357	-,225	-,217	-,070	-,055	-,039
Kolmogorov-Smirnov Z		4,652	3,056	2,947	1,636	,786	1,079
Asymp. Sig. (2-tailed)		,000	,000	,000	,009	,568	,195

a. Test distribution is Normal.

b. Calculated from data.

gen

Anmerkungen: FRAG_RET - richtige Antworten Retrieval; FRAG_STR - richtige Antworten Struktur;

FRAG_GES - richtige Antworten Gesamt; DAUE_RET - Bearbeitungsdauer Retrieval; DAUE_STR - Bearbeitungsdauer Struktur; DAUE_GES - Bearbeitungsdauer Gesamt

Für die AV *usability* zeigen die Ergebnisse aus Tabelle 5, daß lediglich für den Aspekt „System-usability“ (SYS_USE) die Annahme der Normalverteilung der Daten zurückgewiesen werden muß.

Tabelle 5: Test der Normalverteilungsannahme für usability-Faktoren

		One-Sample Kolmogorov-Smirnov Test			
		USAB_G ES	SYS_USE	INFO_QUA	INTE_QUA
N		177	177	177	177
Normal Parameters ^{a,b}	Mean	2,9168	12,1864	9,1469	7,8701
	Std. Deviation	1,5941	6,6464	5,0913	5,1335
Most Extreme Differences	Absolute	,090	,110	,088	,083
	Positive	,045	,073	,064	,083
	Negative	-,090	-,110	-,088	-,079
Kolmogorov-Smirnov Z		1,193	1,466	1,176	1,102
Asymp. Sig. (2-tailed)		,116	,027	,126	,176

a. Test distribution is Normal.

b. Calculated from data.

Bortz (1993) weist darauf hin, daß die „Voraussetzungen der Varianzanalyse mit wachsendem Umfang der untersuchten Stichproben an Bedeutung verlieren“ (S. 263). Als kritische Grenze für Stichprobenumfänge nennt er Werte kleiner als 10 pro Versuchsbedingung. Aufgrund dieser Befundlage soll im folgenden bei den univariaten Varianzanalysen zusätzlich der Kruskal-Wallis-Test als nonparametrisches Verfahren zur Absicherung der Ergebnisse hinzugezogen werden.

6.5 Univariate Analysen der Retrieval bzw. Strukturfragen in Abhängigkeit der Navigationshilfe

6.5.1. Psychometrische Kennwerte der Skalen

Wie in Abschnitt 4.4.2 bereits ausgeführt, sollten für die AV1 (Fragetyp) zwei Aspekte erfasst werden: *Retrieval-Fragen* und komplexe *Strukturfragen*. Zur psychometrischen Qualität der hierzu gebildeten Skalen ist vorab zu sagen, daß aufgrund der speziellen Anforderungen einer Online-Untersuchung relativ wenige Fragen pro Skala (3) gebildet wurden.

Für die Subskala *Retrieval-Fragen* zeigt sich ein Alpha von .62, für die Subskala *komplexe Strukturfragen* ein Wert von .47.

Einzelne Itemanalysen können Tabelle 6 entnommen werden. Für die Retrieval-Skala fällt sowohl hinsichtlich des Schwierigkeitsniveaus als auch im Hinblick auf die Trennschärfe Item 5 durch schlechte Werte auf.

Bei den Items der Struktur-Skala konnte ein sehr einheitliches Schwierigkeitsniveau erzielt werden, allerdings läßt hier generell die Trennschärfe der einzelnen Items zu wünschen übrig - die insgesamt schlechte interne Konsistenz dieser Skala bringt diesen Befund ebenfalls zum Ausdruck.

Tabelle 6: Itemanalysen

Item	Skala	Schwierigkeit	Trennschärfe	α if Item deleted	α der Skala
1	Retrieval	.81	.58	.42	.62
3	Retrieval	.86	.64	.39	.62
5	Retrieval	.60	.34	.95	.62
2	Struktur	.57	.33	.30	.47
4	Struktur	.60	.28	.39	.47
6	Struktur	.59	.27	.41	.47

Um die Zuordnung der sechs Fragen zu den postulierten beiden Skalen zu überprüfen (insbesondere aufgrund der schlechten Reliabilität) wurde zusätzlich anhand einer konfirmatorischen Faktorenanalyse überprüft, ob die zweifaktorielle Lösung mit einer Skala *Retrieval-Fragen* und einer Skala „Struktur-Fragen“ einer einfaktoriellen Lösung überlegen ist.

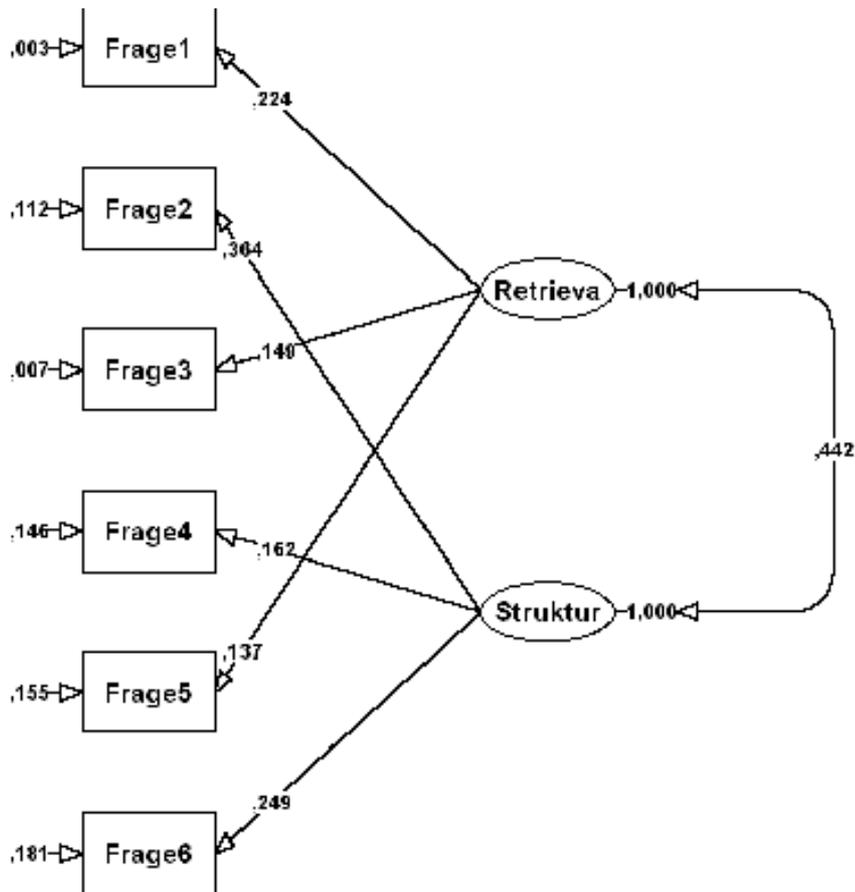


Abbildung 6: 2-faktorielles Modell

Die Indikatoren für die Modellgüte (Tabelle 7) zeigen eine gute Passung des zweifaktoriellen Modells, das sich auch der einfaktoriellen Lösung gegenüber als klar überlegen erweist. Diese Beobachtung wird durch den signifikanten Unterschied des Chi-Quadrat-Wertes gestützt $\chi^2_{\text{Diff}}(1, N=185)=22.38, p<.01$.

Tabelle 7: Indizes der Modellgüte für die konkurrierenden Meßmodelle

Index	2-faktorielle Lösung	1-faktorielle Lösung
Chi-Quadrat	$\chi^2(8, N=185)=16.48, p<.04$	$\chi^2(9, N=185)=38.78, p<.01$
RMSEA	.075	.134
sRMR (standardisierter RMR)	.0041	.0744

Tabelle 7: Indizes der Modellgüte für die konkurrierenden Meßmodelle

Index	2-faktorielle Lösung	1-faktorielle Lösung
GFI	.97	.936
AGFI	.927	.850
NFI	.943	.866
PNFI	.503	.520

6.5.2. Univariate Varianzanalysen: Anzahl richtiger Antworten

Retrieval-Fragen

Für die Bedingung der AV, *Retrieval-Fragen* ergeben sich signifikante Mittelwertsunterschiede $F(2,170)=6.68$, $p<.01$, $\eta^2=.08$: *ohne Hilfe* ($M=2.4$, $SD=.71$), *WebToc* ($M=2.7$, $SD=.54$) und *Hyperbolic-Tree* ($M=2.1$, $SD=1.05$).

Die Ergebnisse des Kruskal-Wallis-Tests ergeben ebenfalls einen signifikanten Mittelwertsunterschied, $\chi^2(2, N=185)=10.51$, $p<.01$.

In den Post-Hoc-Tests (Bonferroni) zu dieser Konfiguration zeigt sich wiederum ein signifikanter Mittelwertsunterschied zwischen *WebToc* und *Hyperbolic-Tree* ($p<.01$). Auch hier bleiben die anderen Post-Hoc-Vergleiche nicht signifikant.

Komplexe Struktur-Fragen

Für die Komponente AV1b, „komplexe Struktur-Fragen“, zeigen sich im Gegensatz dazu keine signifikanten Mittelwertsunterschiede zwischen den Gruppen, $F(2,185)=1.63$: *ohne Hilfe* ($M=1.7$, $SD=1.19$), *WebToc* ($M=1.9$, $SD=.95$) und *Hyperbolic-Tree* ($M=1.5$, $SD=1.15$).

Auch die Ergebnisse des Kruskal-Wallis-Tests ergeben für diese Variable keinen signifikanten Mittelwertsunterschied, $\chi^2(2, N=185)=2.47$.

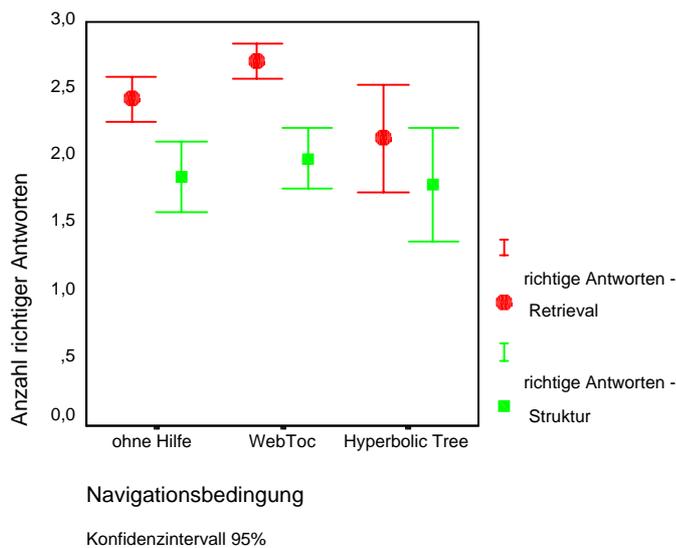


Abbildung 7: richtige Antworten in Abhängigkeit der Navigationshilfe.

6.5.3. Univariate Varianzanalysen: Bearbeitungsdauer

Für die inferenzstatistische Analyse der Bearbeitungsdauer in den einzelnen Stufen der UVs ist ein wichtiger Aspekt bei der Auswahl der zu analysierenden Datensätze zu beachten, und zwar der Zusammenhang von Bearbeitungsdauer und abgebrochenen Fragen. Wird dieser Zusammenhang nicht berücksichtigt, dürften die Daten evtl. stark verzerrt in die Analyse eingehen, da Vpn mit geringer Bearbeitungsdauer diese evtl. nur deshalb erzielt haben, weil sie einen Teil der Fragen überhaupt nicht bearbeitet haben. Da eine entsprechende Antwortalternative zur Ablehnung der Frage nicht vorgesehen war, soll versucht werden, durch eine Vorabselektion der Datensätze dem Problem entgegenzuwirken. Um ein plausibles Kriterium für diese Vorabselektion zu finden, wurde für jede Frage anhand der empirischen Werteverteilung der Antwortzeiten ein individuelles Cut-Off-Kriterium definiert, wobei als Anhaltspunkt die Werte unterhalb des ersten Anstiegs der Verteilungskurve von der Analyse ausgeschlossen wurden, wenn die zugehörige Frage nicht richtig beantwortet wurde.

Anhand dieses Selektionskriteriums resultieren für den Bereich der Retrieval-Fragen 170 Datensätze, für die Strukturfragen 161.

Retrieval-Fragen

Deutliche Unterschiede ergeben sich für die Komponente AV1a, *Retrieval-Fragen: ohne Hilfe* ($M=226.56$ sec., $SD=124.49$), *WebToc* ($M=210.43$, $SD=97.01$) und *Hyperbolic-Tree* ($M=369.66$, $SD=337.52$). Der Unterschied ist signifikant $F(2,170)=15.62$, $p<.01$.

Im Bonferonni Test zeigt sich hier ein ähnliches Muster wie für die Gesamtskala. Es ergeben sich demzufolge signifikante Mittelwertsunterschiede in den einzelnen Paarvergleichen zwischen *WebToc* und *Hyperbolic-Tree* ($p<.01$), als auch zwischen *ohne Hilfe* und *Hyperbolic-Tree* ($p<.01$). Der Unterschied zwischen *ohne Hilfe* und *WebToc* ist nicht signifikant.

Struktur-Fragen

Wiederum weitgehend parallel zu den vorherigen Analysen zur Anzahl richtiger Antworten scheint auch bei der Bearbeitungsdauer in der Bedingung der UV2b, „komplexe Struktur-Fragen“, kein mehr als zufälliger Unterschied zu resultieren, die Mittelwertsunterschiede werden nicht signifikant, $F(2,161)=.05$: *ohne Hilfe* ($M=339.14$ sec., $SD=188.53$), *WebToc* ($M=335.81$, $SD=145.46$) und *Hyperbolic-Tree* ($M=349.34$, $SD=175.46$).

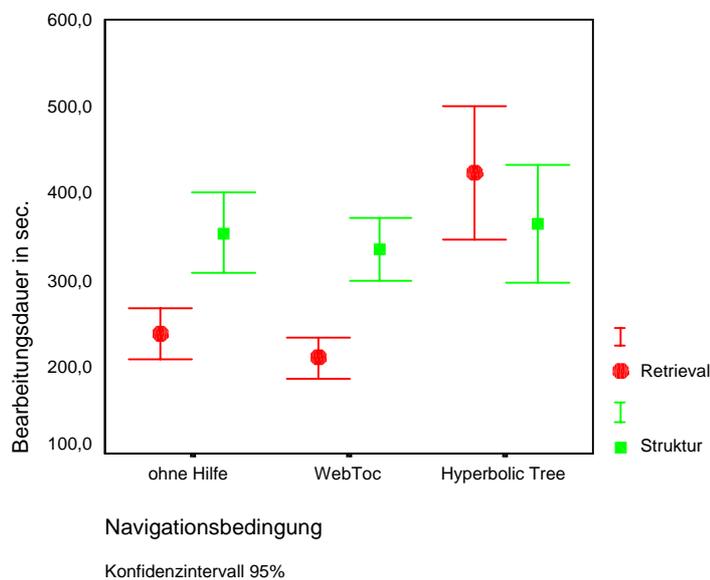


Abbildung 8: Bearbeitungsdauer

6.6 Analysen zur Kovariaten *spatial ability*

6.6.1. Vorbemerkung

Aufgrund eines technischen Problems während der Untersuchung konnten die Antworten für zwei Items dieses Instruments nicht erfasst werden. Aus diesem Grund gehen in die folgenden Analysen nur sechs der ursprünglichen acht Items ein.

6.6.2. Psychometrische Kennwerte der Skalen zur Erhebung von *spatial ability*

Die Skala erreicht einen unter Berücksichtigung der erheblichen Kürzung im Vergleich zum ursprünglichen Instrument (4.4.6) befriedigenden Wert von Alpha .68.

6.6.3. Verteilung der Werte für *spatial ability*

Insgesamt liegen für 155 Vpn Werte für dieses Instrument vor (die Differenz zu den oben genannten Zahlen über die Gesamtstichprobe ergibt sich durch Abbrecher, die für diese Variable stärker ins Gewicht fallen, da der Test für *spatial ability* am Ende der Untersuchung erfolgte). Die Verteilung der Werte ist aus Abbildung 9 zu entnehmen.

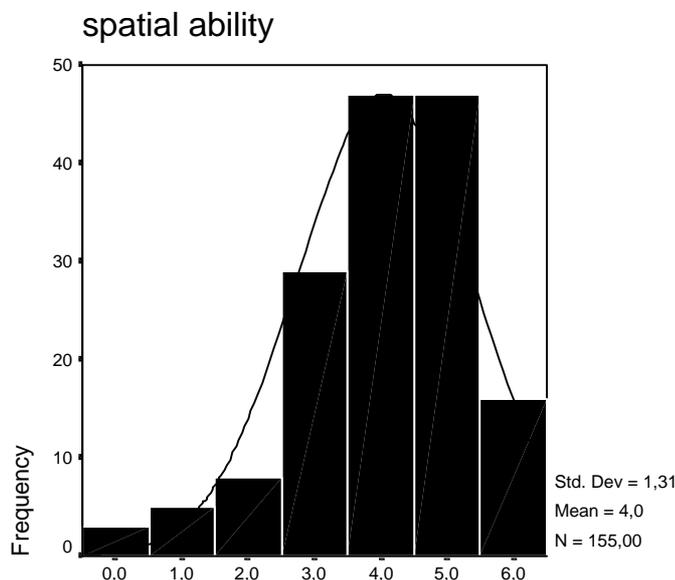


Abbildung 9: Verteilung der Werte für *spatial ability*

Da es sich bei dem Instrument zur Erfassung dieser Komponente ursprünglich um einen Test handelt, der unter eindeutiger Zeitbeschränkung durchgeführt werden soll (vgl. Abschnitt 4.4.6) und diese Versuchsbedingung aus untersuchungspragmatischen Gründen hier

nicht eingehalten werden konnte, interessiert zusätzlich die Bearbeitungsdauer für diesen Test, sowie der Zusammenhang von Testwerten und der Bearbeitungsdauer.

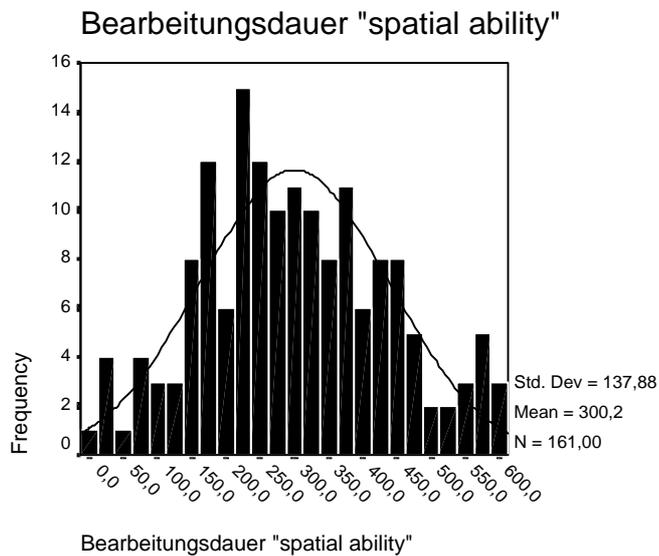


Abbildung 10: Bearbeitungsdauer für den Untertest *spatial ability*

Es ist aus mehreren Gründen ein Zusammenhang zwischen der Bearbeitungsdauer und den Testleistungen zu erwarten, vor allem auch deshalb, weil der Autor des Wilde-Intelligenz-Test, aus dem dieser Subtest verändert entnommen wurde, zwei Varianten mit unterschiedlichen Vorgabewerten für die Bearbeitungszeit vorgibt (Jäger & Althoff, 1983). Aus diesem Grund soll mit einer linearen Regressionsrechnung überprüft werden, ob der entsprechende Zusammenhang zwischen Bearbeitungsdauer und Testleistung nachgewiesen werden kann. Inhaltlich zentral ist dann allerdings die Frage, ob bei einem hohen Zusammenhang die Testleistung überhaupt noch als Indikator für *spatial ability* herangezogen werden kann. Sollte der Zusammenhang z. B. den maximalen (theoretischen) Wert von 1 erreichen, würde die Variable *spatial ability* vollständig über die Bearbeitungsdauer erklärt und könnte somit durch diese ersetzt werden.

Die Regressionsanalyse zeigt tatsächlich den erwarteten Zusammenhang ($R^2 = .22$, $F(1,160) = 44.82$, $p < .01$).

Deskriptiv ist dieser Zusammenhang aus Abbildung 11 ersichtlich, allerdings legt die Abbildung auch die Annahme zusätzlicher Einflußfaktoren nahe.

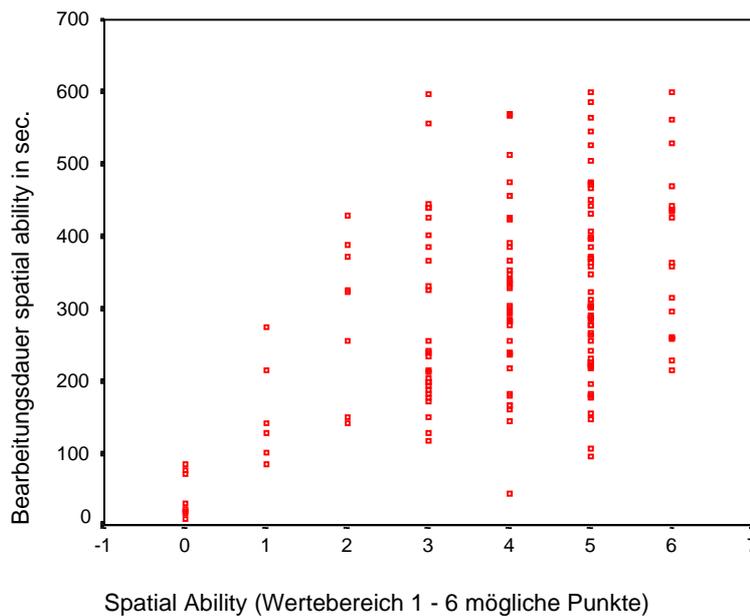


Abbildung 11: Zusammenhang von *spatial ability* und der Bearbeitungsdauer für den entsprechenden Test

Alle Analysen scheinen also trotz der methodischen Bedenken aufgrund der fehlenden Zeitkontrolle für diesen Untertest nahezulegen, daß *spatial ability* durchaus im folgenden als Kovariate in die Analysen einbezogen werden sollte.

6.7 Multivariate Analysen unter Einbeziehung der Kovariaten *spatial ability*

Die Ergebnisse der multivariaten Analysen bzw. der Kovarianzanalyse sind für das Untersuchungsdesign zentral, insbesondere unter der zusätzlichen Einbeziehung der Kovariaten *spatial ability*.

Die Analysen werden aus den in Abschnitt 6.5.3 genannten Gründen anhand der vorab selektierten Daten durchgeführt. Dadurch, daß für die Erhebung der Kovariaten *spatial ability* nicht für alle Vpn Erhebungsdaten vorliegen, reduziert sich die Anzahl der in diesem Vergleich repräsentierten Daten nochmals (N=145).

Für die Beurteilung der Ergebnisse der multivariaten Analysen soll jeweils als Maß Pillais Trace herangezogen werden (zur Begründung vgl. Bortz 1993, S. 674), weitere Kennwerte werden in Tabellen mitgeteilt und bei starken Abweichungen im Text besprochen.

Für die Haupteffekte wird jeweils zusätzlich als Maß für die Effektstärke η^2 angegeben und unter Bezug auf Bortz und Döring (1995; S. 568) als klein, mittel oder groß eingestuft.

6.7.1. Werte für Retrieval-Fragen

Die deskriptiven Werte für die abhängigen Variablen in diesem Vergleich sind aus Tabelle 8 zu entnehmen.

Tabelle 8: Deskriptive Statistiken

Navigationsbedingung		Mean	Std. Deviation	N
richtige Antworten - Retrieval	ohne Hilfe	2,5075	,6366	67
	WebToc	2,7759	,4207	58
	Hyperbolic Tree	2,2400	,9256	25
	Total	2,5667	,6494	150
Bearbeitungsdauer - Retrieval	ohne Hilfe	228,6716	125,1983	67
	WebToc	201,6724	85,8545	58
	Hyperbolic Tree	410,0400	199,3426	25
	Total	248,4600	146,8752	150

Wie aus Tabelle 9 ersichtlich wird, hat die Kovariate *spatial ability* im multivariaten Vergleich keinen signifikanten Effekt ($F(2,150)=2.75$). Auf den Faktor *Navigationshilfe* ist dagegen ein signifikanter Effekt zurückzuführen, $F(4,150)=14.20$; $p<.01$. Die Stärke des Effekts ist mit $\eta^2=.16$ als klein bzw. unter Berücksichtigung der zusätzlichen Prüfverfahren als klein bis mittel einzustufen.

Tabelle 9: Multivariate Vergleiche - Retrieval

Effect		Value	F	df	Error df	Sig.	η^2
Intercept	Pillai's Trace	,586	102,575 ^a	2	145	,000	,586
	Wilks' Lambda	,414	102,575 ^a	2	145	,000	,586
	Hotelling's Trace	1,415	102,575 ^a	2	145	,000	,586
	Roy's Largest Root	1,415	102,575 ^a	2	145	,000	,586
spatial ability	Pillai's Trace	,037	2,755 ^a	2	145	,067	,037
	Wilks' Lambda	,963	2,755 ^a	2	145	,067	,037
	Hotelling's Trace	,038	2,755 ^a	2	145	,067	,037
	Roy's Largest Root	,038	2,755 ^a	2	145	,067	,037
Navigationshilfe	Pillai's Trace	,326	14,202	4	292	,000	,163
	Wilks' Lambda	,680	15,438 ^a	4	290	,000	,176
	Hotelling's Trace	,463	16,676	4	288	,000	,188
	Roy's Largest Root	,445	32,503 ^b	2	146	,000	,308

a. Exact statistic

b. The statistic is an upper bound on F that yields a lower bound on the significance level.

Die univariaten Mittelwertsvergleiche zeigen einen signifikanten Effekt der Kovariaten *spatial ability* auf die Anzahl der richtigen Antworten, $F(1,150)=5.37$; $p<.03$, nicht jedoch im Hinblick auf die Bearbeitungsdauer, $F(1,150)=.36$. Im Vergleich dazu ergeben die univariaten Analysen für den Faktor *Navigationshilfe* einen signifikanten Effekt sowohl für die Variable „Anzahl richtiger Antworten“ (vgl. Abbildung 12), $F(2,150)=7.81$; $p<.01$, $\eta^2=.09$ (kleiner Effekt) als auch für die Bearbeitungsdauer (vgl. Abbildung 13), $F(2,150)=23.89$; $p<.01$, $\eta^2=.24$ (mittlerer Effekt).

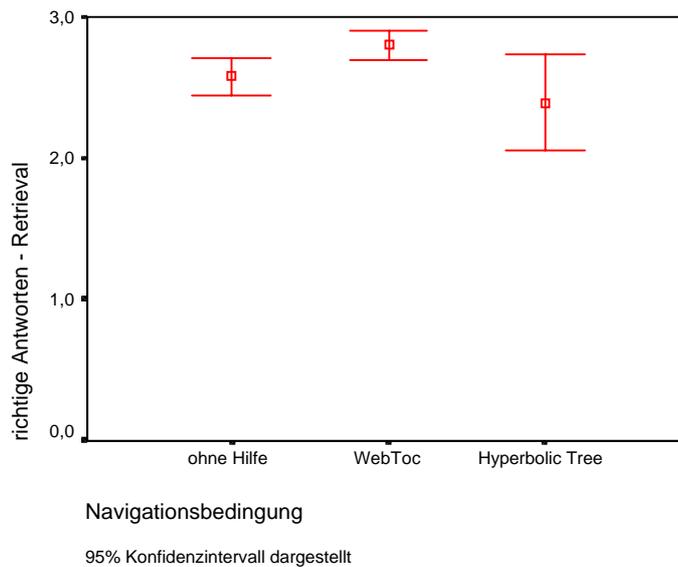


Abbildung 12: richtige Antworten - Retrieval

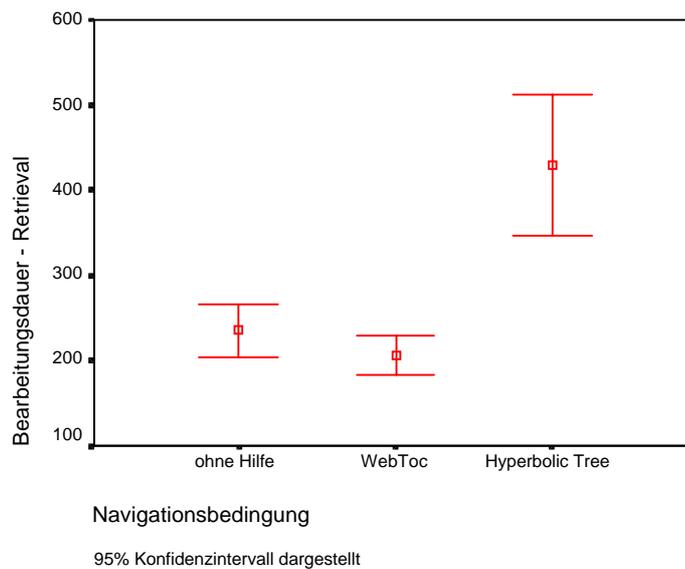


Abbildung 13: Bearbeitungsdauer - Retrieval

Die paarweisen Mittelwertsvergleiche mit dem Bonferonni-Test legen folgenden Zusammenhang für diese Kombination nahe:

Im Hinblick auf die „Anzahl richtiger Antworten“ zeigen sich signifikante Mittelwertsunterschiede zwischen *WebToc* und *ohne Hilfe* ($p < .05$), wie auch zwischen *WebToc* und *Hyperbolic-Tree* ($p < .01$) (vgl. Abbildung 12)

In Bezug auf die abhängige Variable „Bearbeitungsdauer“ zeigen sich ebenfalls signifikante Unterschiede in den Werten für die Bedingung *Hyperbolic-Tree* und *WebToc* ($p < .01$) sowie für die Bedingungen *Hyperbolic-Tree* und *ohne Hilfe* ($p < .01$). Die Bedingungen *WebToc* und *ohne Hilfe* unterscheiden sich nicht signifikant (vgl. Abbildung 13).

6.7.2. Werte für Struktur-Fragen

Die deskriptiven Werte für die abhängigen Variablen in diesem Vergleich sind aus Tabelle 10 zu entnehmen.

Tabelle 10: Deskriptive Statistiken

Navigationsbedingung		Mean	Std. Deviation	N
richtige Antworten - Struktur	ohne Hilfe	2,0156	1,0615	64
	WebToc	1,9310	,9150	58
	Hyperbolic Tree	2,0870	,9960	23
	Total	1,9931	,9895	145
Bearbeitungsdauer - Struktur	ohne Hilfe	336,5313	193,7437	64
	WebToc	331,0172	146,7397	58
	Hyperbolic Tree	373,0435	163,2203	23
	Total	340,1172	170,9618	145

Wie aus Tabelle 11 ersichtlich wird, wirkt sich die Kovariate *spatial ability* im multivariaten Vergleich signifikant aus ($F(4,145)=42.66; p<.01$). Auf den Faktor *Navigationshilfe* scheint hingegen kein signifikanter Mittelwertsunterschied zurückzuführen sein ($F(4,145)=.39$).

Tabelle 11: Multivariate Vergleiche - Gesamt

Effect		Value	F	df	Error df	Sig.	η^2
Intercept	Pillai's Trace	,379	42,669 ^a	2	140	,000	,379
	Wilks' Lambda	,621	42,669 ^a	2	140	,000	,379
	Hotelling's Trace	,610	42,669 ^a	2	140	,000	,379
	Roy's Largest Root	,610	42,669 ^a	2	140	,000	,379
spatial ability	Pillai's Trace	,056	4,141 ^a	2	140	,018	,056
	Wilks' Lambda	,944	4,141 ^a	2	140	,018	,056
	Hotelling's Trace	,059	4,141 ^a	2	140	,018	,056
	Roy's Largest Root	,059	4,141 ^a	2	140	,018	,056
Navigations hilfe	Pillai's Trace	,011	,387	4	282	,818	,005
	Wilks' Lambda	,989	,385 ^a	4	280	,819	,005
	Hotelling's Trace	,011	,383	4	278	,821	,005
	Roy's Largest Root	,010	,723 ^b	2	141	,487	,010

a. Exact statistic

b. The statistic is an upper bound on F that yields a lower bound on the significance level.

Die univariaten Mittelwertsvergleiche zeigen einen signifikanten Effekt der Kovariaten *spatial ability* auf die Anzahl der richtigen Antworten, $F(2,145)=6.19; p<.02$, nicht jedoch im Hinblick auf die Bearbeitungsdauer, $F(2,145)=1.75$.

Die univariaten Analysen für den Faktor *Navigationshilfe* ergeben weder für die Variable „Anzahl richtiger Antworten“ (vgl. Abbildung 14), $F(2,145)=.08$ noch für die Bearbeitungsdauer (vgl. Abbildung 14) einen signifikanten Effekt, $F(2,145)=.70$.

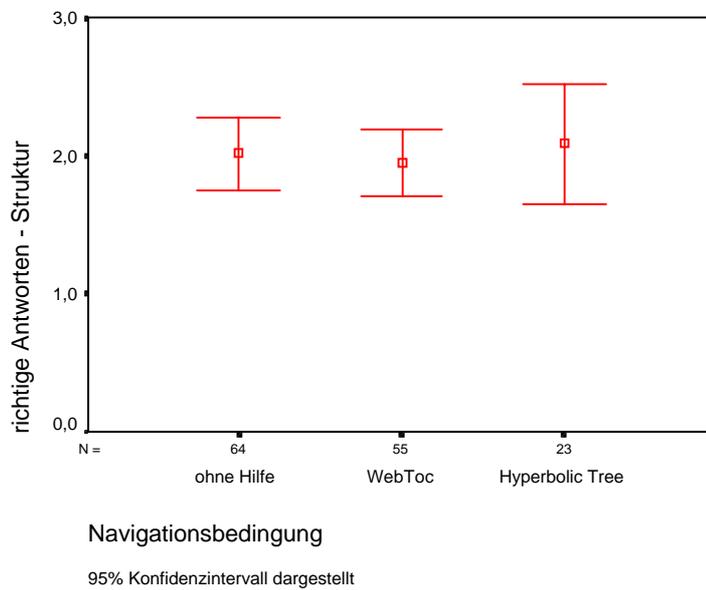


Abbildung 14: richtige Antworten - Struktur

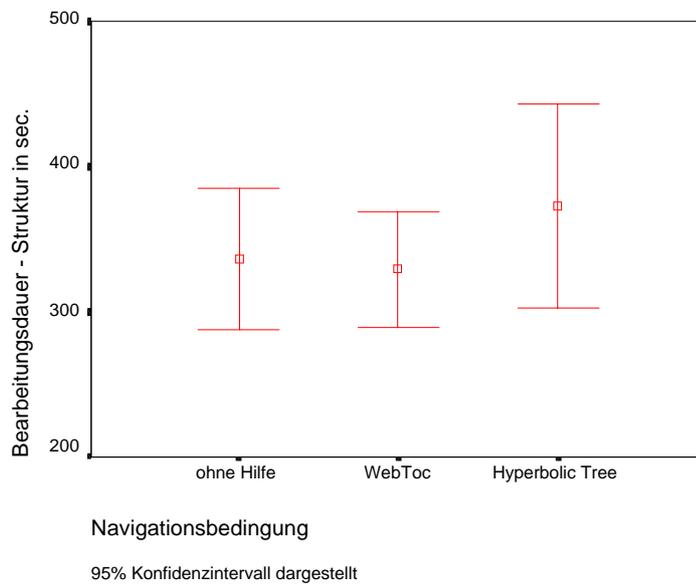


Abbildung 15: Bearbeitungsdauer - Struktur

6.7.3. Untersuchung der zwei Aspekte der AV *Fragentyp* in Abhängigkeit der Navigationsbedingung

Eine Analyse der AV *Fragentyp* in Abhängigkeit des Faktors *Navigationsbedingung* erbringt folgende Ergebnisse:

Anzahl richtiger Antworten

Tabelle 12: Deskriptive Statistiken richtige Antworten in Abhängigkeit der Navigationsbedingung und des Fragentyps

Navigationsbedingung		Mean	Std. Deviation	N
richtige Antworten - Retrieval	ohne Hilfe	2,4853	,6577	68
	WebToc	2,7288	,5518	59
	Hyperbolic Tree	2,1538	1,0077	26
	Total	2,5229	,7171	153
richtige Antworten - Struktur	ohne Hilfe	1,8971	1,1348	68
	WebToc	1,9153	,9152	59
	Hyperbolic Tree	1,8846	1,1073	26
	Total	1,9020	1,0436	153

Im multivariaten Vergleich zeigt sich ein signifikanter Effekt für die AV-Komponenten von *Fragentyp* ($F(1,153)=12.02$; $p<.01$, $\eta^2=.07$), die Interaktion dieser beiden Komponenten mit der Navigationsbedingung ist nicht signifikant ($F(1,150)=2.27$), ebenso verhält es sich für die Interaktion des Fragentyps mit der Kovariaten *spatial ability* ($F(1,150)=2.18$).

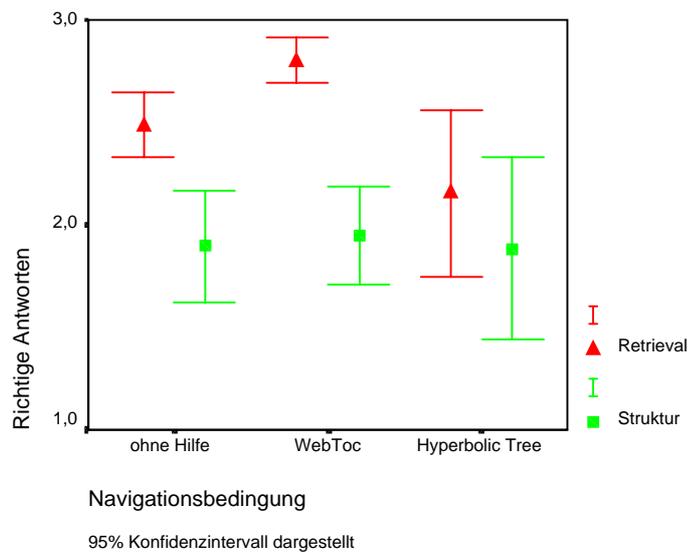


Abbildung 16: richtige Antworten in Abhängigkeit von Fragentyp und Navigationsbedingung

Betrachtet man die Mittelwertsunterschiede hinsichtlich des Between-Subject-Faktors *Navigationshilfe* und den Effekten der Kovariaten *spatial ability* zeigt sich nur für die Kovariate ein signifikanter Effekt ($F(1,153)=20.09$; $p<.01$). Der Faktor *Navigationshilfe* zeigt in dieser Konfiguration keinen signifikanten Effekt ($F(2,153)=2.64$).

Bearbeitungsdauer

Tabelle 13: Deskriptive Statistiken für die AV Bearbeitungsdauer in Abhängigkeit der Navigationsbedingung und des Fragentyps

Navigationsbedingung		Mean	Std. Deviation	N
Bearbeitungsdauer - Retrieval	ohne Hilfe	235,5313	123,1821	64
	WebToc	201,6724	85,8545	58
	Hyperbolic Tree	429,6957	192,7713	23
	Total	252,7862	146,3456	145
Bearbeitungsdauer - Struktur	ohne Hilfe	336,5313	193,7437	64
	WebToc	331,0172	146,7397	58
	Hyperbolic Tree	373,0435	163,2203	23
	Total	340,1172	170,9618	145

Im multivariaten Vergleich zeigt sich ein signifikanter Effekt für die AV-Komponenten von *Fragentyp*, $F(1,145)=5.08$; $p<.03$, $\eta^2=.03$, aber auch die Interaktion dieser Komponenten

mit der Navigationsbedingung ist signifikant, $F(2,145)=9.65$; $p<.01$. Die Interaktion der beiden Komponenten mit der Kovariaten *spatial ability* zeigt hingegen keinen signifikanten Effekt $F(1,145)=1.26$.

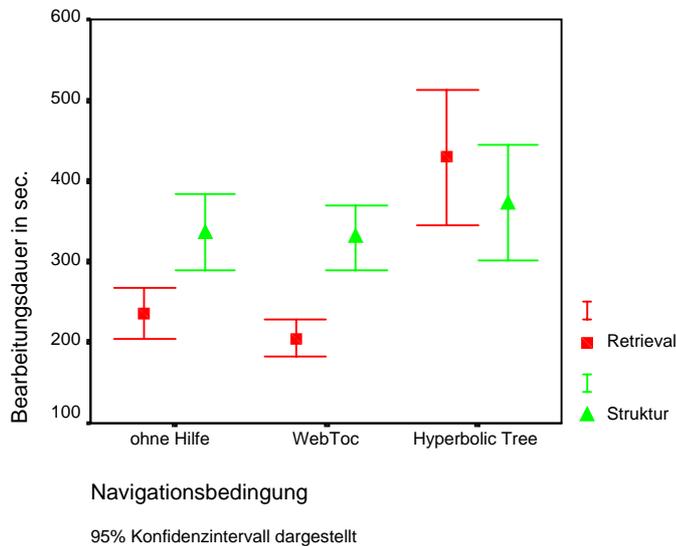


Abbildung 17: Bearbeitungsdauer in Abhängigkeit von Fragentyp und Navigationsbedingung

Betrachtet man nun den Mittelwertsunterschied hinsichtlich des Between-Subject-Faktors *Navigationshilfe* und der Kovariaten *spatial ability*, zeigt sich für die Kovariate kein signifikanter Effekt, $F(1,145)=1.14$. Der Faktor *Navigationshilfe* hingegen zeigt in dieser Konfiguration einen signifikanten Effekt, $F(2,145)=10.63$; $p<.01$.

6.8 Analyse der Daten des usability-Fragebogens

6.8.1. Reliabilität

Der *usability*-Fragebogen erreicht ähnliche Reliabilitätswerte, wie sie auch in der Literatur berichtet wurden (Lewis, 1995). Alpha beträgt .95.

6.8.2. Faktorielle Struktur

Wie unter 4.4.5 bereits ausgeführt, wurde das Instrument zur Erfassung von *usability* auf die speziellen Anforderungen dieser Untersuchung angepasst und um acht Items verkürzt. Zudem wurden die verbleibenden Items ins Deutsche übersetzt. Insofern ist es von beson-

derer Bedeutung, ob neben psychometrischen Kennwerten die faktorielle Struktur des Instruments in den Daten reproduziert werden kann.

Die vom Autor des Instruments postulierte dreifaktorielle Struktur (Lewis, 1995) konnte anhand der Daten dieser deutschsprachigen Adaptation zunächst mit einer exploratorischen Faktorenanalyse nicht reproduziert werden. Mit diesem Verfahren resultiert hingegen eine einfaktorielle Lösung (nach Varimax-Rotation), die 70.4 Prozent der Varianz aufklären kann.

Um Annahmen über die faktorielle Struktur näher untersuchen zu können, wurde zusätzlich eine konfirmatorische Faktorenanalyse durchgeführt. Hierbei wurde die oben genannte einfaktorielle Lösung und die berichtete Lösung mit drei Faktoren untersucht.

Zunächst wurde die einfaktorielle Lösung untersucht.

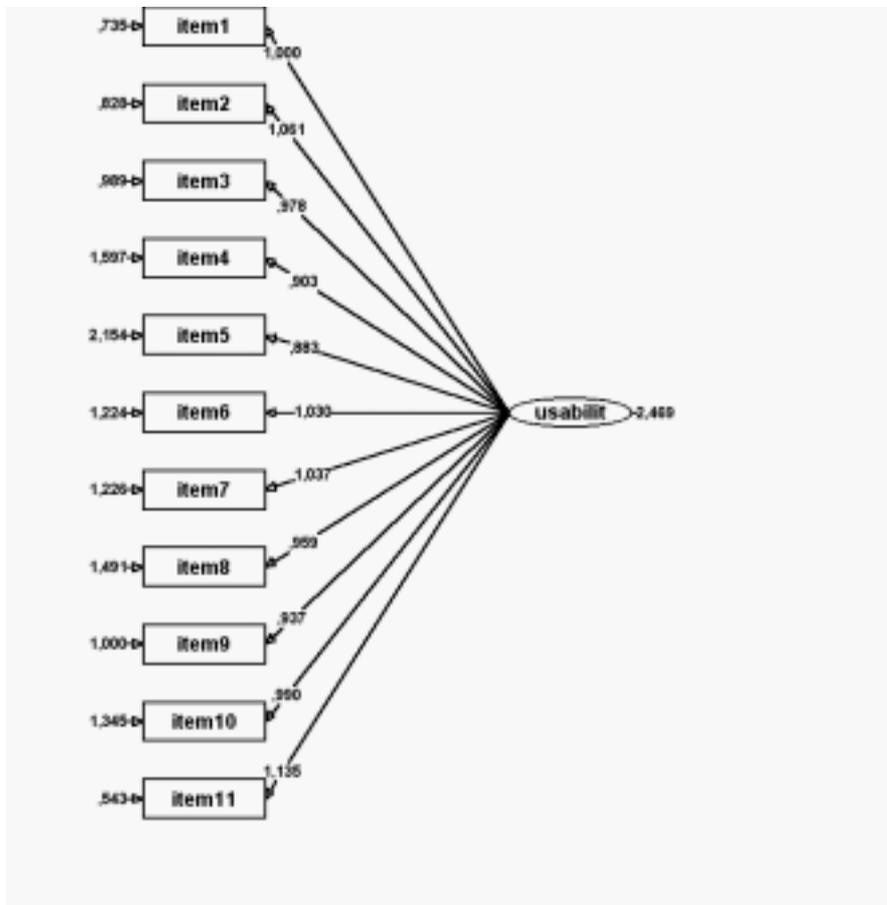


Abbildung 18: Einfaktorielle Lösung in der konfirmatorischen Faktorenanalyse

Die zugehörigen Kennwerte für die Güte des Modells ergeben sich wie folgt:

Tabelle 14: Indizes der Modellgüte für das 1-faktorielle Meßmodell

Index	Wert
Chi-Quadrat	$\chi^2(43, N=177)=288.47, p<.01$
RMSEA	.18
sRMR (standardisierter RMR)	.06
GFI	.73
AGFI	.59
NFI	.85

Für die Berechnung der Lösung des dreifaktoriellen Modells wurde eine Lösung für 10 Items gesucht, da Lewis (1995) für Item19 (hier Item11) keine Zuordnung zu einem der vorgeschlagenen Faktoren vornimmt.

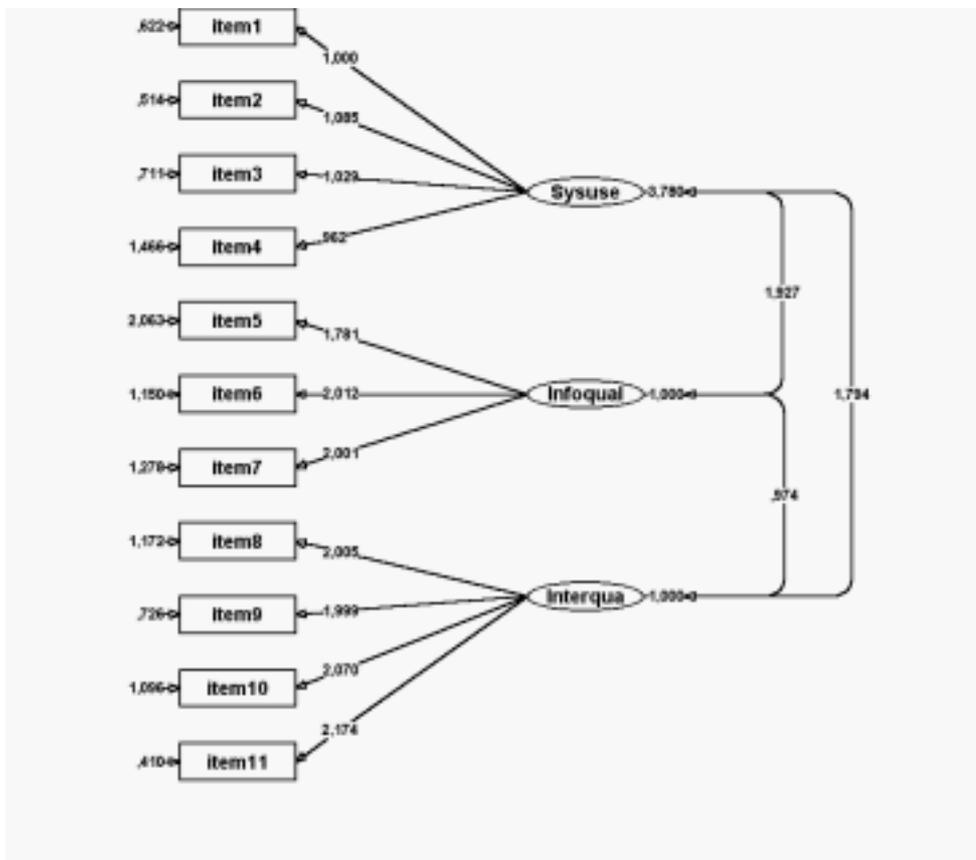


Abbildung 19: Meßmodell der 3-faktoriellen Struktur

Für dieses Modell wurden folgende Parameter der Anpassungsgüte errechnet:

Tabelle 15: Indizes der Modellgüte für das 3-faktorielle Meßmodell

Index	Wert
Chi-Quadrat	$\chi^2(40, N=177)=193.77, p<.01$
RMSEA	.148
sRMR (standardisierter RMR)	.034
GFI	.82
AGFI	.721
NFI	.92

Entgegen den Ergebnissen der exploratorischen Faktorenanalyse scheint die dreifaktorielle Lösung also der einfaktoriellen nach dieser Analyse klar überlegen zu sein. Das Modell mit drei Faktoren weist hinsichtlich aller relevanter Indizes eine bessere Passung auf, die Differenz der Chi-Quadrat-Werte ist signifikant, $\chi^2(3, N=177)=90.68, p<.01$. Die insgesamt relativ schlechten Werte für die Anpassung der Modelle an die Daten sind für konfirmatorische Faktorenanalysen offenbar nicht untypisch (vgl. Kelloway, 1998). Da es sich bei *usability* um eine multifacettierte AV handelt, wird allerdings dennoch analog zur Vorgehensweise des Autors des Fragebogens (Lewis, 1995) zusätzlich zu den Einzelkomponenten ein Summenscore in die Analysen einbezogen. Dies ist auch insofern zu rechtfertigen, als bei der drei-faktoriellen Lösung ein Item, welches offenbar eine Gesamtbeurteilung verlangt, nicht mit einbezogen wird (Item 19 der ursprünglichen Fassung: „Im Allgemeinen bin ich mit diesem System zufrieden“, vgl. Anhang C 2).

6.8.3. Ergebnisse für die unterschiedlichen Navigationsbedingungen

usability-Gesamt

Die Varianzanalyse zeigt einen Mittelwertsunterschied zwischen den Gruppen, der folgendermaßen dargestellt werden kann: *ohne Hilfe* ($M=3.08, SD=1.62$), *WebToc* ($M=3.05, SD=1.44$) und *Hyperbolic-Tree* ($M=2.26, SD=1.66$). Der Unterschied ist signifikant $F(2,177)=3.63, p<.05$.

In den Post-Hoc-Tests (Bonferroni) zeigt sich allein der Unterschied zwischen *ohne Hilfe* und *Hyperbolic-Tree* als signifikant (*Mittelwertsdifferenz*: .81; $p<.05$).

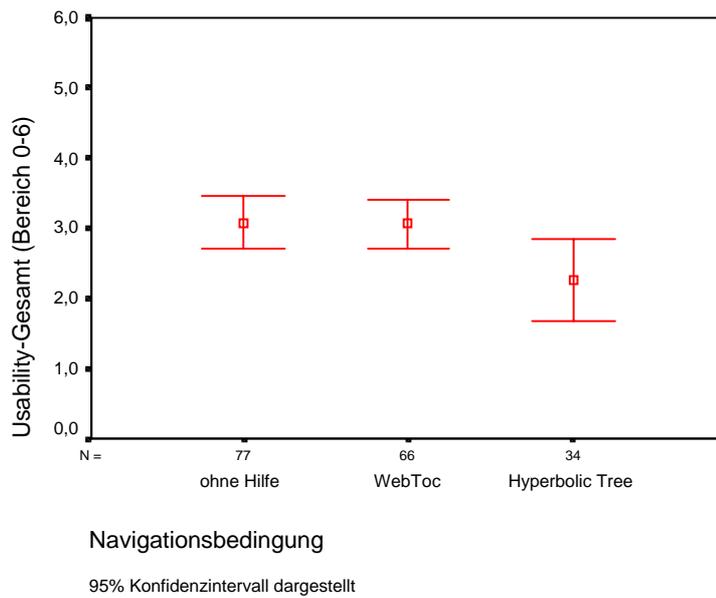


Abbildung 20: usability in Abhängigkeit der Navigationsbedingung

Tabelle 16: usability-Dimensionen in Abhängigkeit der Navigationsbedingung

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
SYS_USE	Between Groups	359,998	2	179,999	4,224	,016
	Within Groups	7414,849	174	42,614		
	Total	7774,847	176			
INFO_QUA	Between Groups	177,041	2	88,520	3,512	,032
	Within Groups	4385,140	174	25,202		
	Total	4562,181	176			
INTE_QUA	Between Groups	100,804	2	50,402	1,933	,148
	Within Groups	4537,207	174	26,076		
	Total	4638,011	176			

usability - Faktor „System Usefulness“

Die Varianzanalyse zeigt für diese Komponente der *usability* folgenden Mittelwertsunterschied zwischen den Gruppen: *ohne Hilfe* ($M=12.94$, $SD=6.68$), *WebToc* ($M=12.8$, $SD=6.16$) und *Hyperbolic-Tree* ($M=9.26$, $SD=6.84$). Der Unterschied ist signifikant $F(2,177)=4.22$, $p<.05$.

In den Post-Hoc-Tests (Bonferroni) erweist sich sowohl der Unterschied zwischen *ohne Hilfe* und *Hyperbolic-Tree* als signifikant ($p<.05$) als auch der Unterschied zwischen *WebToc* und *Hyperbolic-Tree* ($p<.05$).

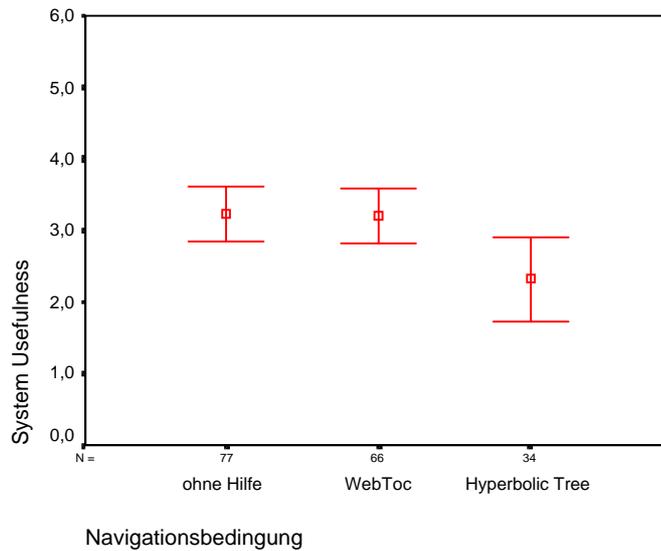


Abbildung 21: „System-Usefulness“ in Abhängigkeit der Navigationsbedingung

usability - Faktor „information quality“

Auch für diesen Faktor ergibt sich ein Mittelwertsunterschied zwischen den Gruppen, der folgendermaßen ausgeprägt ist: *ohne Hilfe* ($M=9.77$, $SD=5.05$), *WebToc* ($M=9.45$, $SD=4.65$) und *Hyperbolic-Tree* ($M=7.11$ $SD=5.59$). Der Unterschied ist signifikant $F(2,177)=3.51$, $p<.05$.

In den Post-Hoc-Tests (Bonferroni) zeigt sich allein der Unterschied zwischen *ohne Hilfe* und *Hyperbolic-Tree* als signifikant ($p<.05$).

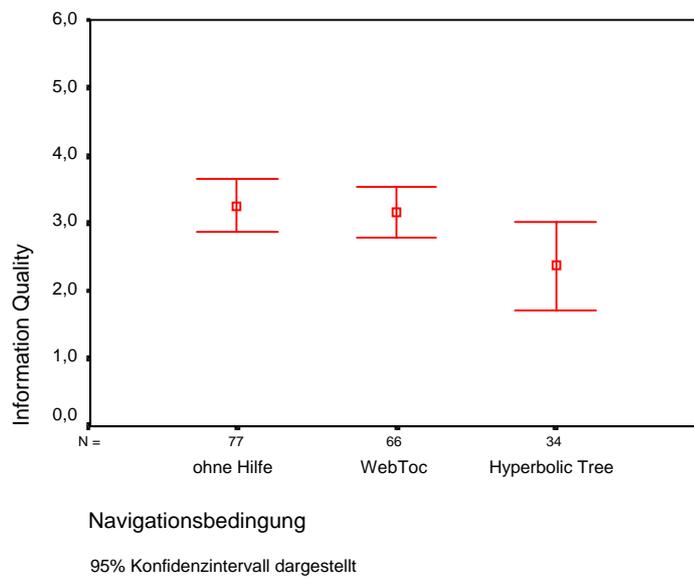


Abbildung 22: „Information-Quality“ in Abhängigkeit der Navigationsbedingung

usability - Faktor „Interface Quality“

Hinsichtlich des Faktors „Interface Quality“ gibt es folgende Mittelwertsunterschiede zu berichten: *ohne Hilfe* ($M=8.2$, $SD=5.36$), *WebToc* ($M=8.27$, $SD=4.77$) und *Hyperbolic-Tree* ($M=6.32$ $SD=5.13$). Der Unterschied ist nicht signifikant $F(2,177)=1.93$.

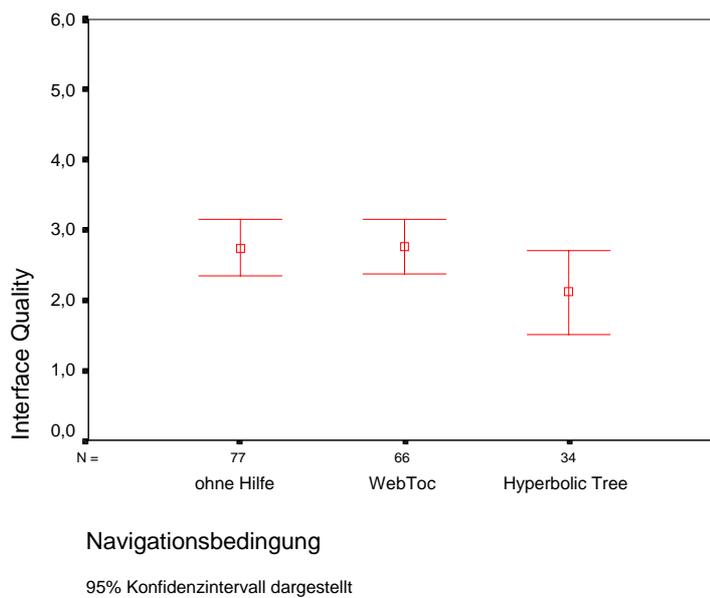


Abbildung 23: „Interface-Quality“ in Abhängigkeit der Navigationsbedingung

6.9 Analysen einzelner Kennwerte aus den Logfiles in Abhängigkeit von der Navigationsbedingung

Wie bereits unter 4.4.3 dargestellt, können die abhängigen Variablen aus den Logfile-Daten aufgrund untersuchungstechnischer/methodischer Bedenken nur bedingt interpretiert werden.

6.9.1. Unterschiede hinsichtlich Stratum und Compactness

Da nicht für alle Vpn Logfile-Daten vorliegen (zur Begründung vgl. 4.4.3), ergibt sich für die folgenden Analysen eine Stichprobengröße von $N=162$.

Die graphentheoretischen Maße Stratum und Compactness unterscheiden sich in Abhängigkeit der Navigationsbedingung wie folgt:

Tabelle 17: Deskriptive Statistiken zu Compactness und Stratum

		Descriptives					
		N	Mean	Std. Deviation	Std. Error	Minimum	Maximum
Compactness	ohne Hilfe	72	,6096	,1276	.0150	,34	,86
	WebToc	65	,5179	,1447	.0179	,34	,84
	Hyperbolic Tree	25	,4421	,1179	.0236	,34	,79
	Total	162	,5469	,1462	.0115	,34	,86
Stratum	ohne Hilfe	72	,4585	,2538	.0299	,16	1,00
	WebToc	65	,6197	,2831	.0351	,15	1,00
	Hyperbolic Tree	25	,8037	,2358	.0472	,16	1,00
	Total	162	,5764	,2890	.0227	,15	1,00

Sowohl für Compactness ($F(2,162)=17.20$; $p<.01$) als auch für Stratum ($F(2,162)=17.39$; $p<.01$) lassen sich signifikante Mittelwertsunterschiede zeigen (vgl. Tabelle 18).

Tabelle 18: Univariate Varianzanalysen in Abhängigkeit der Navigationsbedingung

		Sum of Squares	df	Mean Square	F	Sig.
Compactness	Between Groups	,613	2	,306	17,209	,000
	Within Groups	2,830	159	1,780E-02		
	Total	3,443	161			
Stratum	Between Groups	2,415	2	1,207	17,399	,000
	Within Groups	11,035	159	6,940E-02		
	Total	13,450	161			

In den multiplen Paarvergleichen (Bonferonni-Test) zeigt sich für die Variable „Stratum“ bei allen Kombinationen ein signifikanter Mittelwertsunterschied (vgl. Tabelle 19). Für die Variable „Compactness“ ergibt sich ebenfalls für alle Bedingungen ein signifikanter Unterschied, mit der Ausnahme der Kombination *Hyperbolic-Tree* und *WebToc*.

Tabelle 19: multiple Mittelwertsvergleiche (Post-Hoc-Analyse)

Bonferroni						
Dependent Variable	(I) Navigationsbedingung	(J) Navigationsbedingung	Mean Difference (I-J)	Std. Error	Sig.	
Compactness	ohne Hilfe	WebToc	.0917*	.0228	,000	
		Hyperbolic Tree	.1675*	.0310	,000	
	WebToc	ohne Hilfe	-.0917*	.0228	,000	
		Hyperbolic Tree	.0758	.0314	,051	
	Hyperbolic Tree	ohne Hilfe	-.1675*	.0310	,000	
		WebToc	-.0758	.0314	,051	
Stratum	ohne Hilfe	WebToc	-.1612*	.0451	,001	
		Hyperbolic Tree	-.3453*	.0612	,000	
	WebToc	ohne Hilfe	.1612*	.0451	,001	
		Hyperbolic Tree	-.1841*	.0620	,010	
	Hyperbolic Tree	ohne Hilfe	.3453*	.0612	,000	
		WebToc	.1841*	.0620	,010	

*. The mean difference is significant at the .05 level.

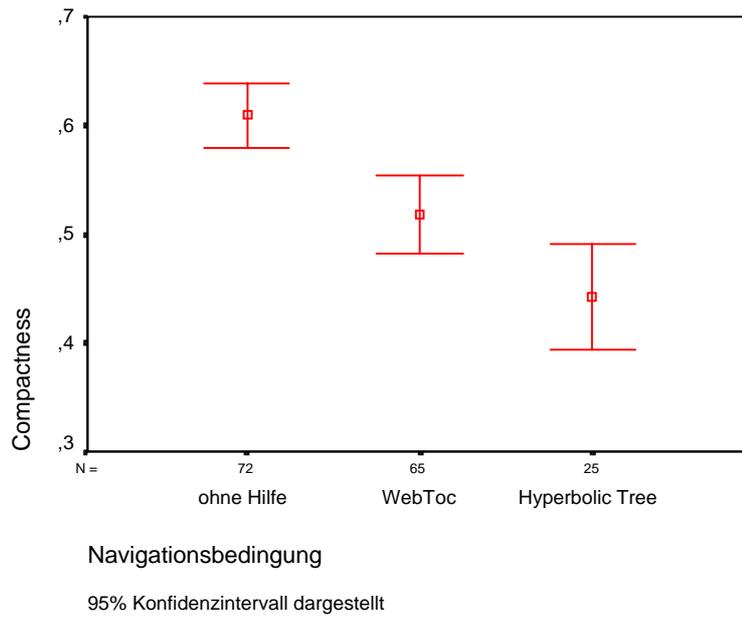


Abbildung 24: Compactness in Abhängigkeit der Navigationsbedingung

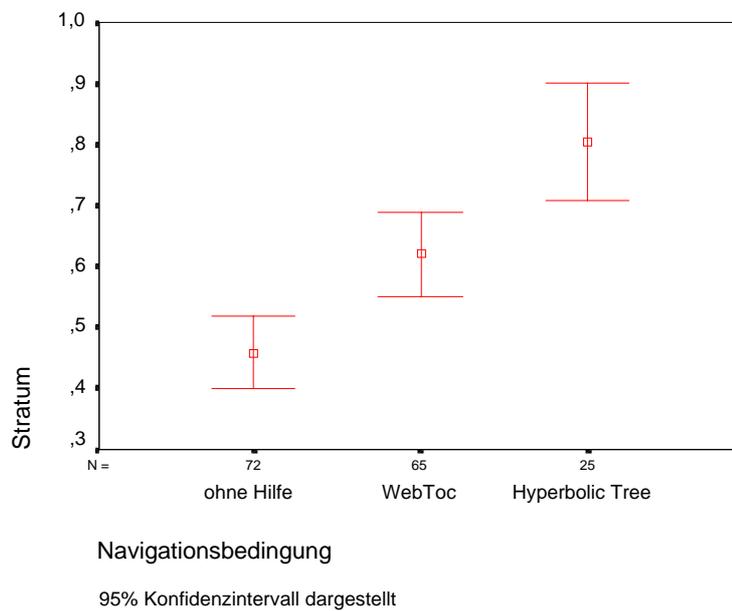


Abbildung 25: Stratum in Abhängigkeit der Navigationsbedingung

6.9.2. Anzahl besuchter Knoten

Tabelle 20: Deskriptive Statistiken zur Anzahl der besuchten Knoten

log_knoten

	N	Mean	Std. Deviation	Std. Error	Minimum	Maximum
ohne Hilfe	81	31,1235	16,8370	1,8708	1,00	85,00
WebToc	68	26,5294	14,3927	1,7454	,00	79,00
Hyperbolic Tree	36	14,7778	13,9512	2,3252	1,00	51,00
Total	185	26,2541	16,4906	1,2124	,00	85,00

Die Varianzanalyse ergibt einen signifikanten Mittelwertsunterschied für diese Dimension, $F(2,185)=13.98$; $p<.01$.

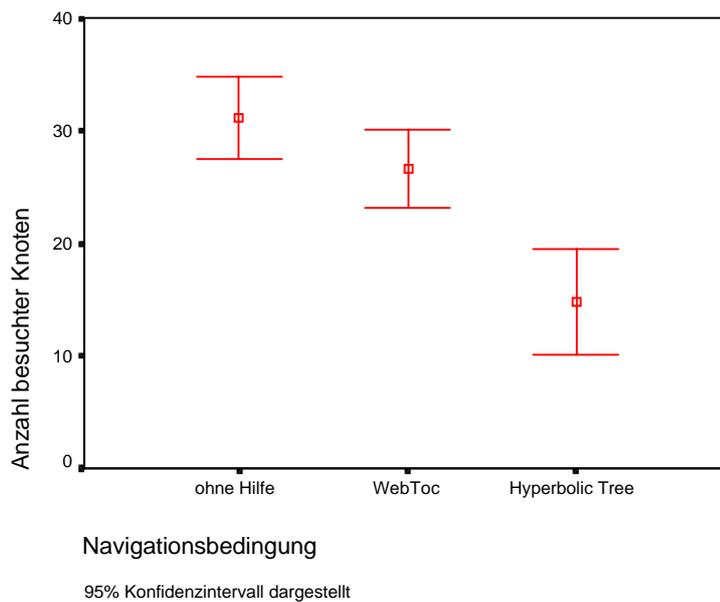


Abbildung 26: Anzahl besuchter Knoten in Abhängigkeit der Navigationsbedingung

In den Post-Hoc Analysen mit dem Bonferonni-Test zeigt sich, daß sowohl für die Kombination „ohne-Hilfe“ - *Hyperbolic-Tree* ($p<.01$) als auch *WebToc* - *Hyperbolic-Tree* ($p<.01$) ein signifikanter Mittelwertsunterschied festgestellt werden kann. Die Mittelwerte der Bedingungen *ohne Hilfe* und *WebToc* unterscheiden sich hingegen nicht signifikant.

6.9.3. Aufruf der Seitentypen „Index“ und „Suche“

Die Gruppen der drei Stufen des Faktors *Navigationshilfe* unterscheiden sich signifikant hinsichtlich des Aufrufs bestimmter Seitentypen.

Im besonderen wurde der Aufruf der Startseite (Index) und der Aufruf von globalen Suchseiten der Bundestags-Website ausgezählt und anhand der Gesamtlänge der jeweiligen Logfile standardisiert.

Tabelle 21: Deskriptive Statistiken für Seitenaufrufe „Index“ und „Suchfunktion“

		N	Mean	Std. Deviation	Std. Error	Minimum	Maximum
Index-Seite	ohne Hilfe	84	.09202	.07547	.00823	.000	.280
	WebToc	71	.06263	.07481	.00888	.000	.286
	Hyperbolic Tree	37	.01311	.03802	.00625	.000	.176
	Total	192	.06595	.07512	.00542	.000	.286
Suchfunktion	ohne Hilfe	84	.18473	.14194	.01549	.000	.514
	WebToc	71	.15972	.13190	.01565	.000	.478
	Hyperbolic Tree	37	.07595	.11711	.01925	.000	.333
	Total	192	.15452	.13901	.01003	.000	.514

Wie aus Tabelle 22 ersichtlich wird, ist der Mittelwertsunterschied in beiden Fällen signifikant.

Die zugehörigen Post-Hoc-Vergleiche zeigen signifikante Unterschiede zwischen den Faktorstufen *Hyperbolic-Tree* und *WebToc*, *Hyperbolic-Tree* und *ohne Hilfe*, aber nicht zwischen *WebToc* und *ohne Hilfe*.

Tabelle 22: Univariate Varianzanalyse über Seitenaufrufe

		Sum of Squares	df	Mean Square	F	Sig.
Index-Seite	Between Groups	,161	2	8,060E-02	16,620	,000
	Within Groups	,917	189	4,850E-03		
	Total	1,078	191			
Suchfunktion	Between Groups	,307	2	,153	8,574	,000
	Within Groups	3,384	189	1,790E-02		
	Total	3,691	191			

6.10 Exploratorische Einzelanalysen der graphentheoretischen Logfilemaße für die

Stufen des Faktors Navigationshilfe

Aufgrund der oben diskutierten (vgl. 4.4.3) Einwände gegen eine Interpretation der Logfilemaße über alle Stufen des Faktors *Navigationshilfe* sollen zusätzlich exploratorische Analysen innerhalb der einzelnen Faktorstufen vorgenommen werden. Insbesondere die Bedingung *ohne Hilfe* ist in diesem Zusammenhang von Interesse, da hier vollständige Logfile-Daten für alle Vpn vorliegen.

In den folgenden Abschnitten soll dann beleuchtet werden, ob die abhängigen Variablen „Anzahl richtiger Antworten“ und „Bearbeitungsdauer“ systematisch mit den Ausprägungen für die Logfilemaße variieren (Korrelation) bzw. in welchem Umfang die Varianz dieser Werte von den Prädiktoren „Stratum“ und „Compactness“ aufgeklärt werden kann (Regressionsanalyse).

6.10.1. Faktorstufe *ohne Hilfe***Korrelationen (Pearson)**

Wie aus Tabelle 23 ersichtlich ist, ergeben sich für die Kombination der Variablen Com-

Tabelle 23: Korrelationen

Kombination	r (Pearson)	N	p
Compactness - richtige Antworten (Retrieval)	.35	70	<.01
Compactness - richtige Antworten (Struktur)	.29	70	<.02
Compactness - richtige Antworten (Gesamt)	.36	72	<.01
Stratum - richtige Antworten (Retrieval)	-.42	70	<.01
Stratum - richtige Antworten (Struktur)	-.38	72	<.01
Stratum - richtige Antworten (Gesamt)	-.45	72	<.01
Compactness - Bearbeitungszeit (Retrieval)	.15	70	n.s.
Compactness - Bearbeitungszeit (Struktur)	.21	66	n.s.
Compactness - Bearbeitungszeit (Gesamt)	.20	66	n.s.
Stratum - Bearbeitungszeit (Retrieval)	-.11	70	n.s.
Stratum - Bearbeitungszeit (Struktur)	-.13	66	n.s.
Stratum - Bearbeitungszeit (Gesamt)	-.10	66	n.s.

compactness und Stratum mit den verschiedenen Skalen für „Anzahl richtiger Antworten“ substantielle Korrelationen, die ohne Ausnahme signifikant sind.

Im Zusammenhang mit den Skalenwerten der Variable „Bearbeitungszeit“ ergeben sich dagegen keine signifikanten Korrelationen.

Um den Zusammenhang zwischen den Logfilemaßen und den Ausprägungen für „Anzahl richtiger Antworten“ näher zu beleuchten, soll die Hypothese überprüft werden, ob ein entsprechender Zusammenhang vorliegt

Lineare Regressionsanalyse

In die lineare Regressionsrechnung sollen die Variablen „Stratum“ und „Compactness“ als Prädiktoren eingehen. Als Kriterium werden die zwei Skalenwerte für „Anzahl richtiger Antworten“ herangezogen (Struktur- und Retrieval-Fragen).

Für die Variable *Retrieval-Fragen* ergibt sich ein R^2 von .19. Auf der Ebene der einzelnen Faktoren zeigt sich für den Faktor „Compactness“ ein Koeffizient von $\beta=-.42$ und für „Stratum“ ein Koeffizient von $\beta=-.81$.

Für die Variable „Struktur-Fragen“ ergibt sich ebenfalls ein R^2 von .19. Auf der Ebene der einzelnen Faktoren zeigt sich für den Faktor „Compactness“ ein Koeffizient von $\beta=-0.67$ und für „Stratum“ ein Koeffizient von $\beta=-1.02$.

6.10.2. Faktorstufe *WebToc*

Korrelationen

Für diese Faktorstufe werden nur zwei Korrelationen signifikant, und zwar die Werte für „Stratum“ und „Compactness“ jeweils in Zusammenhang mit der Bearbeitungsdauer für Retrieval-Fragen (Stratum*Bearbeitungsdauer-Retrieval: $N=65$; $r=-.29$; $p<.02$; Compactness*Bearbeitungsdauer-Retrieval: $N=65$; $r=.37$; $p<.01$;).

Lineare Regressionsanalyse

Aufgrund der Korrelationen sollen in die Regression die Variablen „Stratum“ und „Compactness“ als Prädiktoren eingehen. Als Kriterium wird nur der Skalenwert „Bearbeitungsdauer-Retrieval“ herangezogen.

Für diese Konstellation ergibt sich ein R^2 von .20. Auf der Ebene der einzelnen Faktoren zeigt sich für den Faktor „Compactness“ ein Koeffizient von $\beta=1.27$ und für „Stratum“ ein Koeffizient von $\beta=0.93$.

6.10.2.1. Faktorstufe Hyperbolic-Tree

Korrelationen

Für diese Faktorstufe ergibt sich nur eine signifikante Korrelation, und zwar für die Kombination von „Stratum“ und „Bearbeitungsdauer-Struktur“ (N=28; $r=-.45$; $p<.04$).

Lineare Regressionsanalyse

Gemäß den Ergebnissen der bivariaten Korrelationen sollen in die lineare Regressionsrechnung hier nur der Faktor „Stratum“ als Prädiktor und die Variable „Bearbeitungsdauer-Struktur“ als Kriterium eingehen.

Es ergibt sich ein R^2 von .21.

6.11 Analysen zu Bearbeitungsdauer und Anzahl richtiger Antworten in Abhängigkeit der Internet-Erfahrung

Das Design der vorliegenden Untersuchung als Online-Experiment ließ es aufgrund der starken Zeitbeschränkung nicht zu, Computer- bzw. Web-Literacy differenziert zu erheben. Dennoch soll versucht werden, anhand der erfassten Variablen „Stunden der Internetnutzung pro Woche“ und „Jahre der Internetnutzung“ explorativ Zusammenhänge zwischen deren Ausprägungen und den Werten der abhängigen Variablen zu untersuchen.

Hierzu wurde die Stichprobe anhand eines Mediansplit jeweils in zwei Gruppen aufgeteilt. Da eine symmetrische Aufteilung in zwei nahezu gleichgroße Gruppen aufgrund einer stark besetzten Mittelkategorie nicht möglich war, wurde eine kleinere Gruppe mit Werten unterhalb des Medians für „Stunden der Internetnutzung pro Woche“, und eine etwas größere Gruppe mit Werten in der Mediankategorie und oberhalb des Medians gebildet. Dies erscheint insofern plausibel, als aufgrund der ermittelten technischen und demographischen Kennwerte für die Stichprobe (vgl. 4.6) von einer insgesamt eher überdurchschnittlich qualifizierten Auswahl ausgegangen werden kann. Es resultierte folgende Verteilung (vgl. Tabelle 24):

Tabelle 24: Aufteilung der Gruppen für „Interneterfahrung“

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	wenig	72	38,5	39,3	39,3
	viel	111	59,4	60,7	100,0
	Total	183	97,9	100,0	
Missing	System	4	2,1		
Total		187	100,0		

6.11.1. Analyse in Abhängigkeit der Anzahl richtiger Antworten und der Bearbeitungsdauer

Retrievalfragen

Für den multivariaten Vergleich im Hinblick auf die Variable *Retrieval-Fragen* ergibt sich ein signifikanter Effekt für den Faktor *Navigationshilfe*, $F(4,149)=13.55$; $p<.01$, $\eta^2=.16$, aber auch für den Gruppenfaktor „Interneterfahrung“ $F(2,149)=4.26$; $p<.02$; $\eta^2=.05$. Weder die Kovariate *spatial ability* noch die Interaktion werden signifikant.

Bei den univariaten Vergleichen zeigt sich folgendes Bild: Die Navigationsbedingung hat sowohl auf die „Anzahl richtiger Fragen“, $F(2,149)=8.92$; $p<.01$ als auch auf die zugehörige Bearbeitungsdauer einen signifikanten Effekt, $F(2,149)=20.45$; $p<.01$. Die Kovariate „spatial ability“ zeigt auch hier keinen signifikanten Einfluß. Anders verhält es sich mit der Kovariaten „Interneterfahrung“. Diese zeigt für die „Anzahl richtiger Antworten“ einen signifikanten Effekt, $F(1,149)=7.09$; $p<.01$, nicht jedoch für die Bearbeitungsdauer. Die Interaktion ist auch hier nicht signifikant.

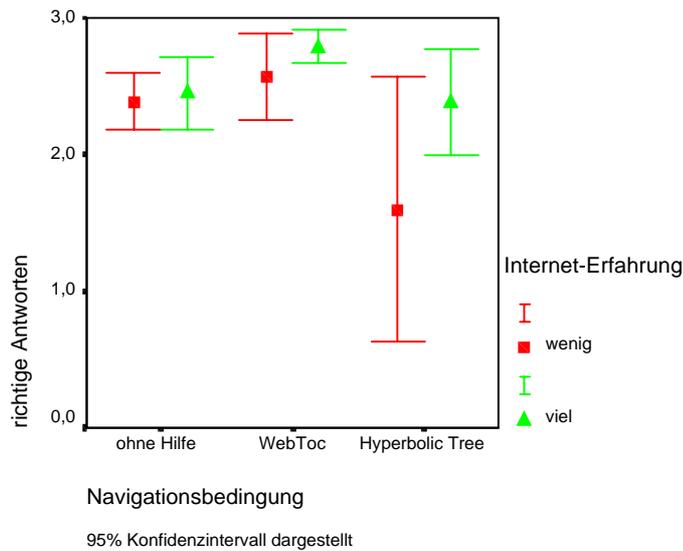


Abbildung 27: Anzahl richtiger Retrieval-Fragen in Abhängigkeit von Interneterfahrung und Navigationsbedingung

Die paarweisen Mittelwertsvergleiche mit dem Bonferonni-Test legen folgenden Zusammenhang für diese Kombination nahe:

Im Hinblick auf die „Anzahl richtiger Antworten“ zeigen sich signifikante Mittelwertsunterschiede zwischen *WebToc* und *Hyperbolic-Tree* ($p < .01$) wie auch zwischen *Hyperbolic-Tree* und *ohne Hilfe* ($p < .05$). Der Unterschied zwischen den Bedingungen *WebToc* und *ohne Hilfe* wird hingegen nicht signifikant.

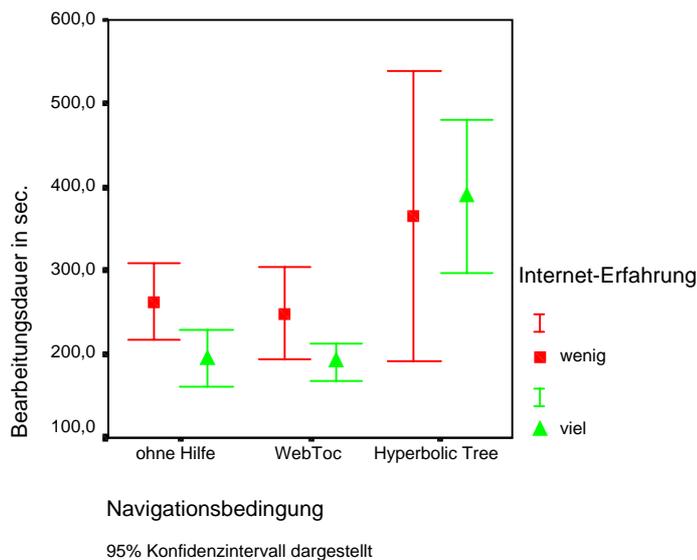


Abbildung 28: Bearbeitungsdauer für Retrieval-Fragen in Abhängigkeit von Interneterfahrung und Navigationsbedingung

Hier vermitteln die multiplen Mittelwertsvergleiche folgendes Bild:

Im Hinblick auf die Bearbeitungsdauer zeigen sich signifikante Mittelwertsunterschiede zwischen *Hyperbolic-Tree* und *ohne Hilfe* ($p < .01$) wie auch zwischen *Hyperbolic-Tree* und *WebToc* ($p < .01$). Der Unterschied zwischen den Bedingungen *WebToc* und *ohne Hilfe* wird hingegen nicht signifikant.

Strukturfragen

Für den multivariaten Vergleich im Hinblick auf die Variable „Struktur-Fragen“ zeigt sich für keinen beteiligten Faktor ein signifikanter Effekt - lediglich die Kovariate *spatial ability* zeigt hier einen signifikanten Effekt, $F(2,146)=4.00$; $p < .03$.

Die univariaten Analysen ergeben ebenfalls nur für die Kovariate *spatial ability* im Zusammenhang mit der Variablen „Anzahl richtiger Antworten“ einen signifikanten Effekt, $F(1,146)=6.05$; $p < .02$.

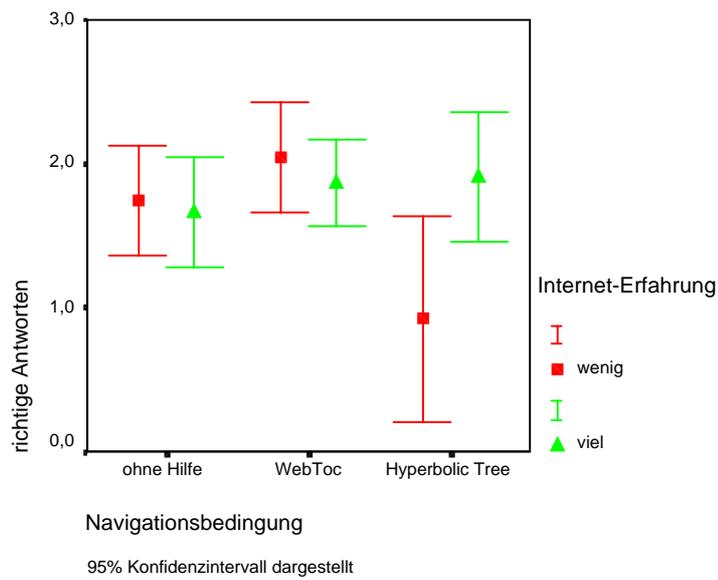


Abbildung 29: Anzahl richtiger Struktur-Fragen in Abhängigkeit von Interneterfahrung und Navigationsbedingung

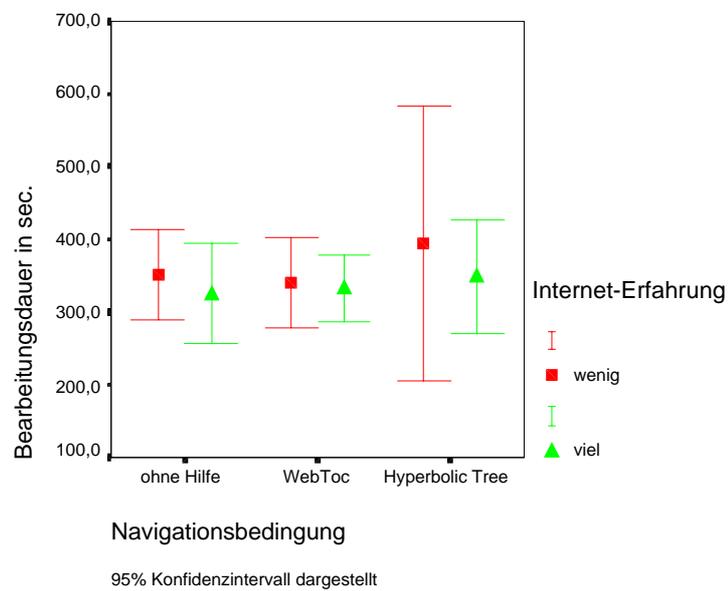


Abbildung 30: Bearbeitungsdauer für Struktur-Fragen in Abhängigkeit der Interneterfahrung und Navigationsbedingung

6.11.2. Analyse des Zusammenhangs mit usability

Gesamtskala

Die Varianzanalyse mit den Faktoren *Navigationshilfe* und „Internetenerfahrung“ zeigt einen signifikanten Effekt für den Faktor Navigationsbedingung, $F(1,175)=3.72$; $p<.03$. Alle anderen involvierten Faktoren werden hingegen nicht signifikant.

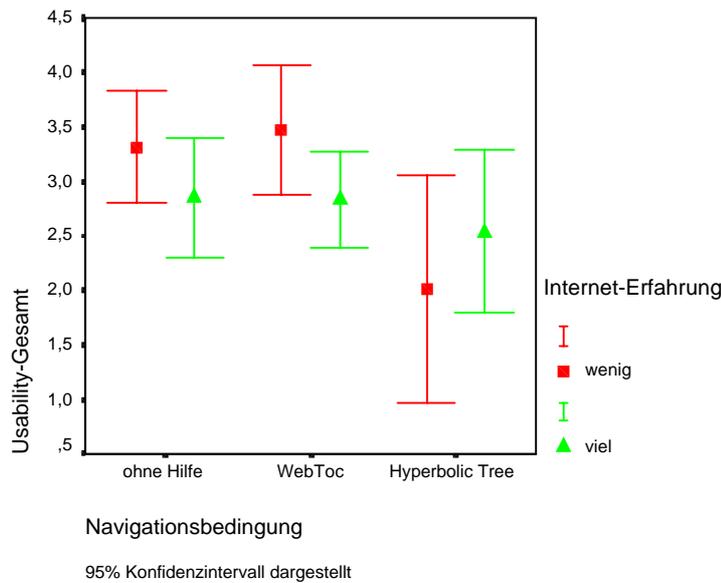


Abbildung 31: „usability (gesamt)“ in Abhängigkeit der Internetenerfahrung und Navigationsbedingung

„System-Usefulness“

Im Zusammenhang mit dem *usability*-Faktor „System-Usefulness“ zeigt sich kein signifikanter Effekt für den Faktor „Internetenerfahrung“, $F(1,175)=0.31$. Im Gegensatz dazu erweist sich der Faktor *Navigationshilfe* hier als signifikant, $F(2,175)=4.28$; $p<.02$. Die Interaktion ist nicht signifikant, $F(2,175)=1.49$.

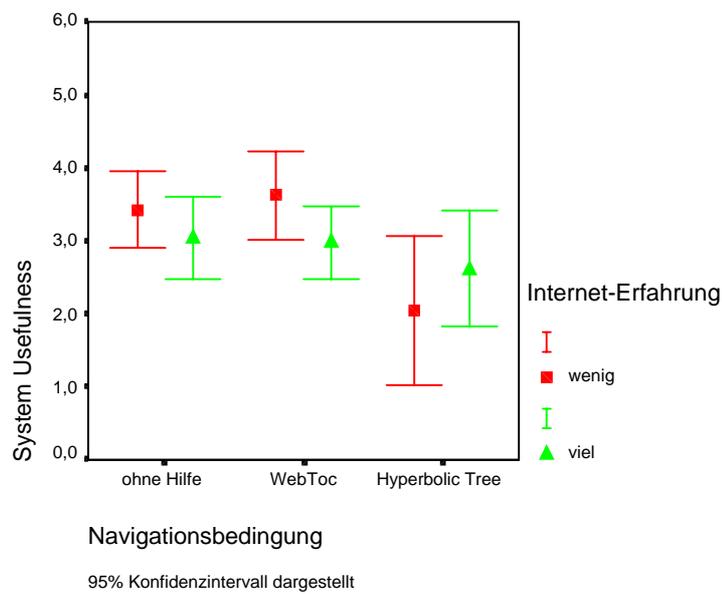


Abbildung 32: „System-Usefulness“ in Abhängigkeit der Interneterfahrung und Navigationsbedingung

„Information-Quality“

Im Hinblick auf den *usability*-Faktor „Information-Quality“ zeigt sich für den Faktor *Navigationshilfe* ein signifikanter Effekt, $F(2,175)=3.29$; $p<.05$. Alle übrigen Faktoren und die Interaktion werden nicht signifikant.

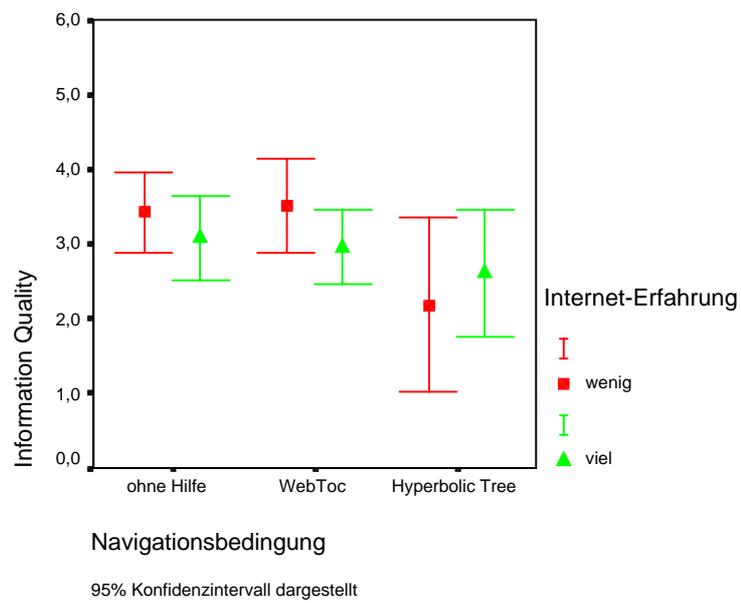


Abbildung 33: „Information Quality“ in Abhängigkeit der Interneterfahrung und Navigationsbedingung

„Interface-Quality“

Auch für den Faktor „Interface-Quality“ wird kein Faktor signifikant, ebenso verhält es sich bezüglich der Interaktion.

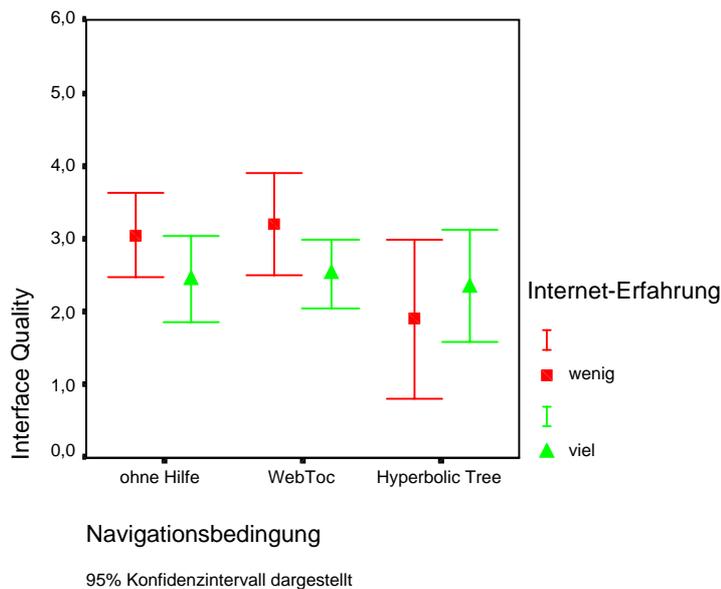


Abbildung 34: „Interface Quality“ in Abhängigkeit der Interneterfahrung und Navigationsbedingung

6.12 Analysen der zusätzlichen Fragen zum Experiment

6.12.1. Schätzung über die Anzahl der Knoten in der Untersuchungs-Site

Die Varianzanalyse zeigt keine signifikanten Unterschiede dieser Variable in Abhängigkeit der Navigationsbedingung. Da die Verteilung der Daten für diese Variable ein ordinales Skalenniveau nahelegt und die Normalverteilungsvoraussetzung auch nicht erfüllt ist, wurde zusätzlich ein nonparametrisches Verfahren für ordinale Daten angewendet (Kruskal-Wallis-Test). Auch mit diesem Verfahren ergibt sich kein signifikanter Unterschied in den mittleren Rängen, $\chi^2(2, N=163)=2.31$.

6.12.2. Weitere Zusatzfragen

Bei der Analyse der drei zusätzlichen Fragen zum Experiment ergab sich für die Frage „Das Experiment hat mir gut gefallen“ und „Ich hatte manchmal das Gefühl, nicht mehr zu wissen wo ich bin“ kein signifikanter Mittelwertsunterschied hinsichtlich des Faktors *Navigationshilfe* (vgl. Tabelle 25).

Bezüglich der Frage „Es fiel mir leicht, die Fragen zu beantworten“ zeigt sich hingegen ein signifikanter Unterschied, $F(2,177)=3.01$; $p<.03$.

In den Post-Hoc-Tests zu dieser Frage erweist sich einzig der Unterschied zwischen der Bedingung *Hyperbolic-Tree* und *WebToc* als signifikant ($p < .03$).

Tabelle 25: Varianzanalysen über zusätzliche Fragen

		Sum of Squares	df	Mean Square	F	Sig.
Gefallen	Between Groups	3,446	2	1,723	1,958	,144
	Within Groups	153,096	174	,880		
	Total	156,542	176			
Leicht	Between Groups	6,033	2	3,016	3,928	,021
	Within Groups	133,628	174	,768		
	Total	139,661	176			
Orientierung	Between Groups	3,928	2	1,964	2,418	,092
	Within Groups	141,337	174	,812		
	Total	145,266	176			

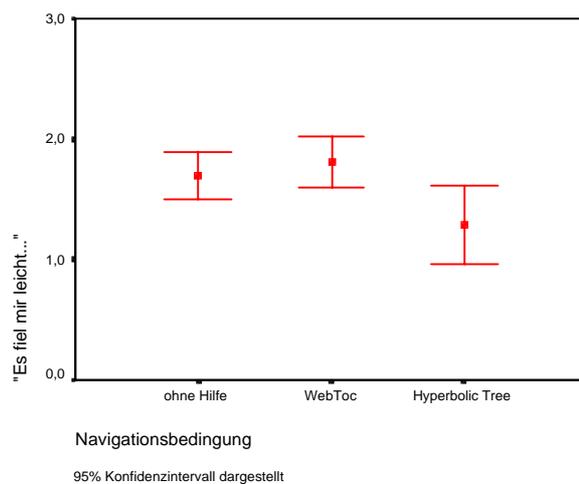


Abbildung 35: Frage: „Es fiel mir leicht die Fragen zu beantworten“ in Abhängigkeit der Navigationsbedingung (hohe Werte = hohe Zustimmung)

Da es aufgrund der Konstruktion dieses Teils der Erhebung denkbar ist, daß allein methodische Mängel (zu wenige Items pro Skala) dazu geführt haben, daß für die verbleibenden beiden Zusatzfragen keine signifikanten Unterschiede festgestellt werden konnten, sollen die Ergebnisse zu explorativen Zwecken graphisch dargestellt werden:

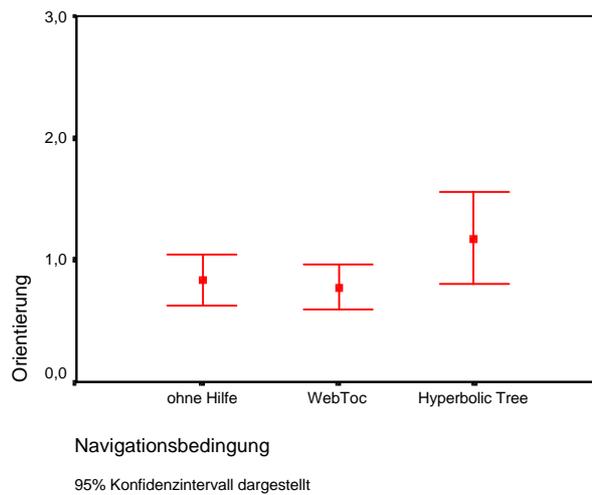


Abbildung 36: Frage: „Ich hatte manchmal das Gefühl, nicht mehr zu wissen, wo ich bin“ (Orientierung) in Abhängigkeit der Navigationsbedingung (hohe Werte = hohe Zustimmung)

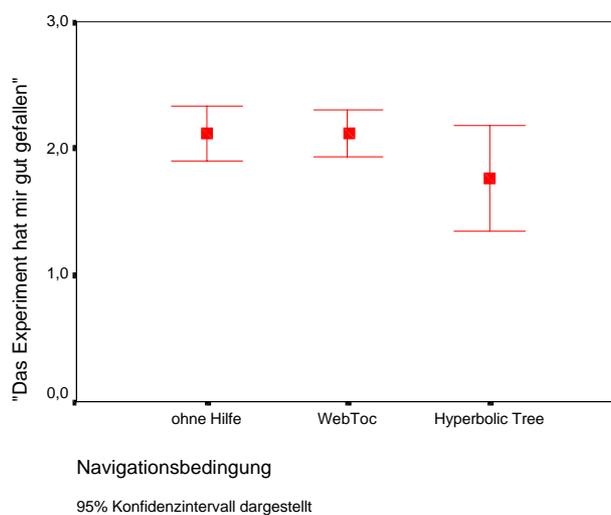


Abbildung 37: Frage: „Das Experiment hat mir gut gefallen“ in Abhängigkeit der Navigationsbedingung (hohe Werte = hohe Zustimmung)

6.12.3. Korrelationen der Antwortausprägungen für die Zusatzfragen mit den abhängigen Variablen

Wiederum zu exploratorischen Zwecken sollen korrelative Beziehungen zwischen den Ausprägungen der Antworten für die drei zuvor analysierten Zusatzfragen und die abhän-

gigen Variablen „Anzahl richtiger Fragen“ und „Bearbeitungsdauer“, sowie die Kovariate *spatial ability* untersucht werden. Hierzu wurden die entsprechenden Korrelationen errechnet. Die Ergebnisse sind aus Tabelle 26 zu ersehen.

Tabelle 26: Korrelationen

Kombination	r (Pearson)	N	p
Gefallen * Leicht	.48	177	sig.
Gefallen * Orientierung	-.11	177	n.sig.
Gefallen * richtige Antworten	.46	177	sig.
Gefallen * Bearbeitungsdauer	.17	159	n.sig.
Gefallen * <i>spatial ability</i>	.26	163	sig.
Leicht * Orientierung	-.19	177	n.sig.
Leicht * richtige Antworten	.57	177	sig.
Leicht * Bearbeitungsdauer	-.15	159	n.sig.
Leicht * <i>spatial ability</i>	.21	163	sig.
Orientierung * richtige Antworten	-.15	177	sig.
Orientierung * Bearbeitungsdauer	.03	159	n.sig.
Orientierung * <i>spatial ability</i>	.00	163	n.sig.

Abkürzungen:

- „Leicht“: Frage: „*Mir fiel es leicht, die Fragen zu beantworten*“
- „Orientierung“: Frage: „*Ich hatte manchmal das Gefühl nicht mehr zu wissen wo ich bin*“
- „Gefallen“: Frage: „*Mir hat das Experiment gefallen*“

6.12.4. Aufruf der Hilfefunktion

Die Hilfe zur Bedienung des Navigationstools war nur für *WebToc* und *Hyperbolic-Tree* verfügbar. Die Funktion wurde insgesamt nur fünf mal in Anspruch genommen, und zwar ausschließlich von Vpn der Bedingung *Hyperbolic-Tree*.

6.13 Analyse der Bearbeitungszeit des Einführungstutorials

Für diese Analyse wurde die Zeit ermittelt, die die Vpn vom Start der Untersuchung bis zur Präsentation der ersten Frage benötigten. In diesem Abschnitt des Experiments wurde die Versuchsinstruktion präsentiert, aber auch ein einführendes Tutorial, um die Möglichkeiten der nachfolgend angebotenen Navigationshilfen zu erläutern. Vom Umfang der Einführung sind hinsichtlich der Navigationsbedingung *WebToc* und *Hyperbolic-Tree* nahezu identisch, die Einführung für die Bedingung *ohne Hilfe* war weniger umfangreich.

Tabelle 27: Deskriptive Statistiken: Bearbeitungszeit für die Einführung in Abhängigkeit der Navigationsbedingung

Bearbeitungsdauer Tutorial						
	N	Mean	Std. Deviation	Std. Error	Minimum	Maximum
ohne Hilfe	81	55,4444	37,4606	4,1623	6,00	187,00
WebToc	68	160,1765	108,1400	13,1139	31,00	519,00
Hyperbolic Tree	36	193,8056	104,4695	17,4116	21,00	476,00
Total	185	120,8649	102,1788	7,5123	6,00	519,00

Eine univariate Varianzanalyse zeigt einen signifikanten Mittelwertsunterschied, $F(2,185)=45,81$; $p<.01$.

Die multiplen Mittelwertsvergleiche (Bonferonni) zeigen allerdings, daß dieser Effekt vermutlich nur auf den Unterschied zwischen den Toolbedingungen (*WebToc* und *Hyperbolic-Tree*) zurückzuführen ist.

Tabelle 28: Post-Hoc-Analyse: multiple Mittelwertsvergleiche

Dependent Variable: Bearbeitungsdauer Tutorial
Bonferroni

(I) Navigations bedingung	(J) Navigations bedingung	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
ohne Hilfe	WebToc	-104,7320*	13,7812	,000	-138,0316	-71,4324
	Hyperbolic Tree	-138,3611*	16,7838	,000	-178,9159	-97,8063
WebToc	ohne Hilfe	104,7320*	13,7812	,000	71,4324	138,0316
	Hyperbolic Tree	-33,6291	17,2703	,159	-75,3596	8,1014
Hyperbolic Tree	ohne Hilfe	138,3611*	16,7838	,000	97,8063	178,9159
	WebToc	33,6291	17,2703	,159	-8,1014	75,3596

*. The mean difference is significant at the .05 level.

Da für die hier untersuchte Variable sowohl die Voraussetzung der Varianzhomogenität als auch die der Normalverteilung verletzt werden, wurde zur Absicherung der Ergebnisse zusätzlich ein nonparametrisches Verfahren herangezogen (Kruskal-Wallis-Test). Auch hier zeigt sich ein signifikanter Anteilsunterschied, $\chi^2(2, N=185)=91.72, p<.00$. Interessanterweise wird mit diesem Test aber auch der Unterschied zwischen den Bedingungen *WebToc* und *Hyperbolic-Tree* signifikant, $\chi^2(1, N=185)=4.81, p<.03$.

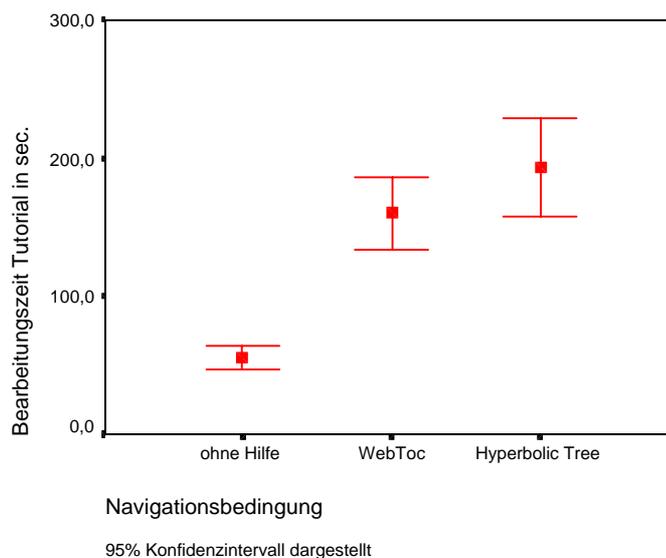


Abbildung 38: Bearbeitungsdauer der Einführung in Abhängigkeit der Navigationsbedingung

Analysen verschiedener bivariater Korrelationen zeigen keine signifikanten Korrelationen mit den abhängigen Variablen „richtige Antworten“ (Retrieval, Struktur, Gesamt). Im Zusammenhang mit den Bearbeitungszeiten für alle drei Skalen hingegen zeigen sich durchgängig signifikante Korrelationen (Struktur-Fragen: $r(N=161)=.264$; $p<.01$; Retrieval-Fragen: $r(N=170)=.30$; $p<.01$; Gesamt: $r(N=161)=.34$; $p<.01$).

6.14 Analyse von möglichen Lerneffekten

Da sowohl hinsichtlich der Vertrautheit mit dem Untersuchungsmaterial als auch im Hinblick auf die Bedienung der Programmoberfläche selbst bei der insgesamt relativ kurzen Bearbeitungszeit in einem Online-Experiment dennoch mit Lerneffekten gerechnet werden kann, soll untersucht werden, ob entsprechende Mittelwertsunterschiede hinsichtlich der Bearbeitungsdauer über die Items der beiden Skalen des Faktors „Fragentyp“ festgestellt werden können.

6.14.1. Lerneffekte bei Retrieval-Fragen

Die multivariate Analyse des Meßwiederholungsfaktors „Zeitpunkt der Bearbeitung“ in Abhängigkeit der Navigationsbedingung ergibt keinen signifikanten Effekt, auch die Interaktion mit der Navigationsbedingung wird nicht signifikant.

Tabelle 29: Deskriptive Statistiken über die drei Meßzeitpunkte für Retrieval-Fragen

	Navigationsbedingung	Mean	Std. Deviation	N
T1	ohne Hilfe	60,2286	56,9287	70
	WebToc	60,7692	48,9728	65
	Hyperbolic Tree	143,9231	75,8672	26
	Total	73,9627	64,8452	161
T2	ohne Hilfe	69,6143	42,3872	70
	WebToc	66,8154	42,0468	65
	Hyperbolic Tree	124,0769	82,6733	26
	Total	77,2795	54,5528	161
T3	ohne Hilfe	105,3857	77,7658	70
	WebToc	82,8462	70,5831	65
	Hyperbolic Tree	124,6923	109,2168	26
	Total	99,4037	81,8262	161

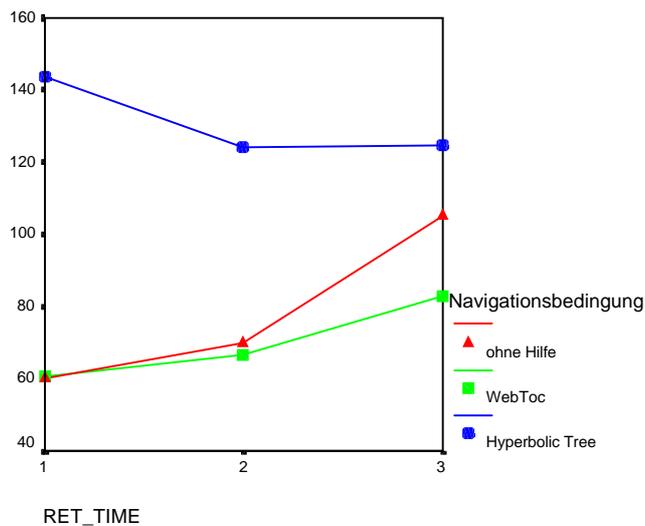


Abbildung 39: Bearbeitungsdauer über den Verlauf der Erhebung bei Retrieval-Fragen

6.14.2. Lerneffekte bei Struktur-Fragen

Die multivariate Analyse des Meßwiederholungsfaktors „Zeitpunkt der Bearbeitung“ in Abhängigkeit der Navigationsbedingung ergibt einen signifikanten Effekt desselben, $F(2, 134)=12.62$; $p<.01$. Auch die Interaktion mit der Navigationsbedingung wird signifikant, $F(4, 134)=4.11$; $p<.01$.

Tabelle 30: Deskriptive Statistiken über die drei Meßzeitpunkte für Struktur-Fragen

	Navigationsbedingung	Mean	Std. Deviation	N
T1	ohne Hilfe	144,9091	102,5550	55
	WebToc	121,7119	63,2038	59
	Hyperbolic Tree	124,8500	77,3830	20
	Total	131,7015	83,6137	134
T2	ohne Hilfe	142,8909	63,1889	55
	WebToc	147,4915	113,0149	59
	Hyperbolic Tree	130,8000	51,7541	20
	Total	143,1119	87,2068	134
T3	ohne Hilfe	86,2545	76,4698	55
	WebToc	71,2034	40,9212	59
	Hyperbolic Tree	137,8500	91,1992	20
	Total	87,3284	69,2265	134

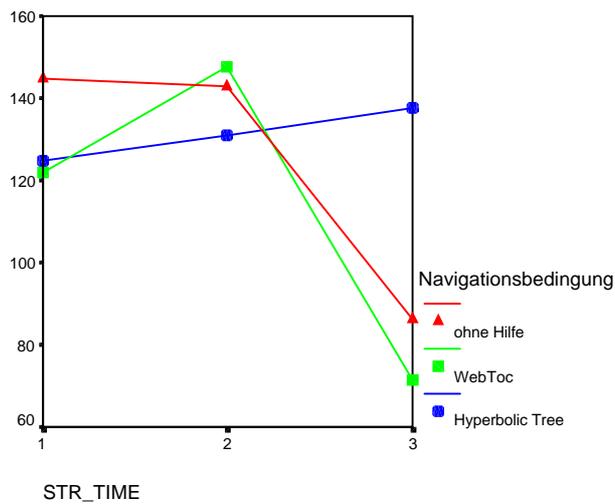


Abbildung 40: Bearbeitungsdauer über den Verlauf der Erhebung bei Struktur-Fragen

6.14.3. Lerneffekte über alle Meßzeitpunkte

Zwar macht es zunächst keinen Sinn, Lerneffekte über alle sechs Meßzeitpunkte zu analysieren, da es sich bei den sechs Items ja um Items zweier unterschiedlicher Skalen handelt, für die unterschiedliche Schwierigkeiten angenommen werden. Dennoch soll hier zumindest die entsprechende Graphik präsentiert werden, da sich ein sehr interessanter Effekt zeigt:

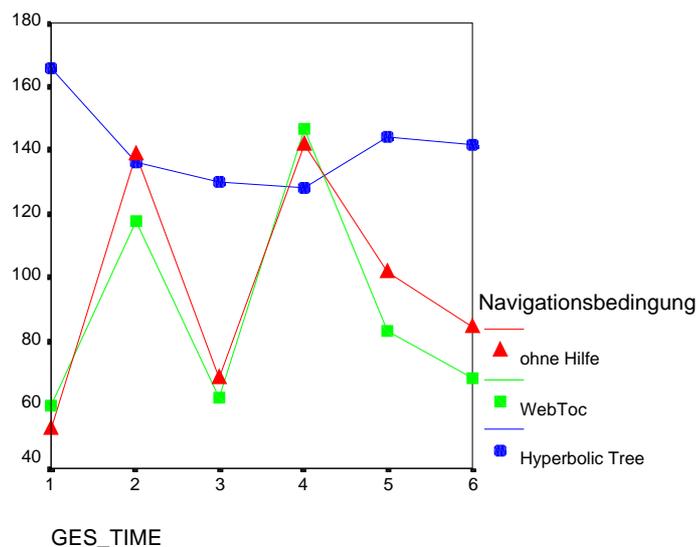


Abbildung 41: Lerneffekte über alle sechs Meßzeitpunkte (=Fragen)

Interessant ist diese Graphik insofern, als daß sich zwar für die Bedingungen *WebToc* und *ohne Hilfe* der erwartete „Zick-Zack-Verlauf“ zeigt, welcher im wesentlichen auf das erwähnte unterschiedliche Schwierigkeitesniveau der Skalen zurückzuführen ist, aber für die Bedingung *Hyperbolic-Tree* kein derartiger Effekt zu sehen ist (näheres siehe Diskussion, 7.3.1).

6.15 Analyse des Dropouts

Wie bereits berichtet (Abschnitt 4.6) beendete ein erheblicher Teil der Vpn, die mit der Bearbeitung des Experimentes begonnen hatten dieses nicht.

Zunächst ist eine Betrachtung des Verlaufs der Dropout-Quote über die gesamte Untersuchung angezeigt. In der zugehörigen Graphik (Abbildung 42) ist deutlich zu erkennen, daß der größte Teil der Abbrecher/innen zu den ersten beiden Meßzeitpunkten abbrechen. Danach weist die Kurve keinen erkennbaren Trend mehr auf, und die Menge der resultierenden Daten wäre für eine weitere inferenzstatistische Untersuchung zu gering.

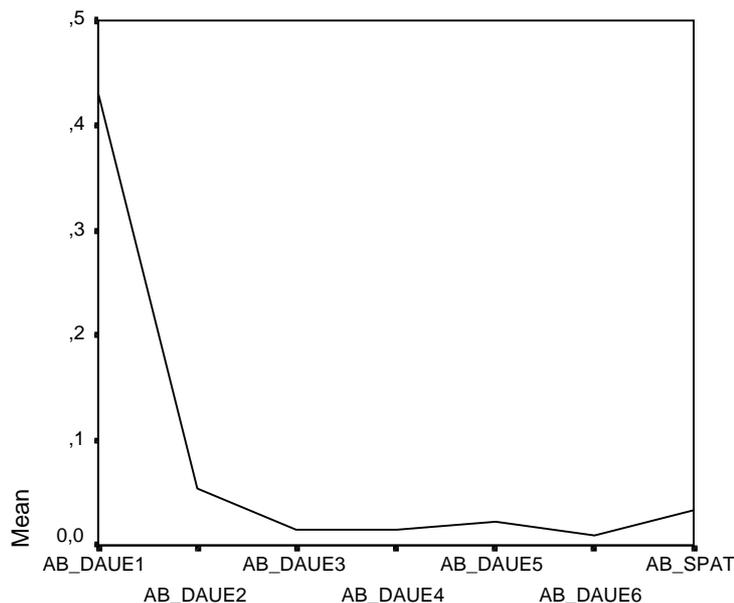


Abbildung 42: Dropout über den Verlauf der Untersuchung

Es liegt in der Natur der Sache, daß bei Abbrechern wenige Daten zur Analyse von Zusammenhängen mit bestimmten Versuchsbedingungen oder Stichprobenmerkmalen vorliegen. Das Design dieser Untersuchung gestattet es allerdings, vier Variablen auszuwerten, die von Beginn der Untersuchung an erhoben wurden und für alle Datensätze zur Verfügung

stehen. Bei diesen beiden Kennwerten handelt es sich um die Bildschirmauflösung am Computer der Vp, die zugewiesene Navigationsbedingung, das eingesetzte Betriebssystem und den eingesetzten Web-Browser.

Im folgenden soll dementsprechend analysiert werden, ob es systematische Zusammenhänge zwischen einer dieser Variablen und der Quote der Abbrecher/innen gibt. Aufgrund der geringen Fallzahlen für spätere Meßzeitpunkte in der Untersuchung werden nur die Meßzeitpunkte T1 (Präsentation der ersten Frage nachdem das Navigationstool geladen wurde) und T2 (Präsentation der zweiten Frage) sowie die Gesamt-Abbrecher/innenquote analysiert. Die Ausfälle zu den Zeitpunkten T1 und T2 machen 84% der Gesamtausfälle aus. Um zufällige Signifikanzen durch mehrmaliges Durchführen des entsprechenden Chi-Quadrat-Tests im folgenden zu vermeiden, wurde eine Alpha-Adjustierung (Bortz, 1993) vorgenommen.

Analyse für Zeitpunkt T1 (Präsentation der ersten Frage)

Insgesamt beträgt der Dropout zu diesem Zeitpunkt 178 Fälle. Das Histogramm für Zeitpunkt T1 zeigt einen im Vergleich deutlich höheren Wert in der Navigationsbedingung *Hyperbolic-Tree*.

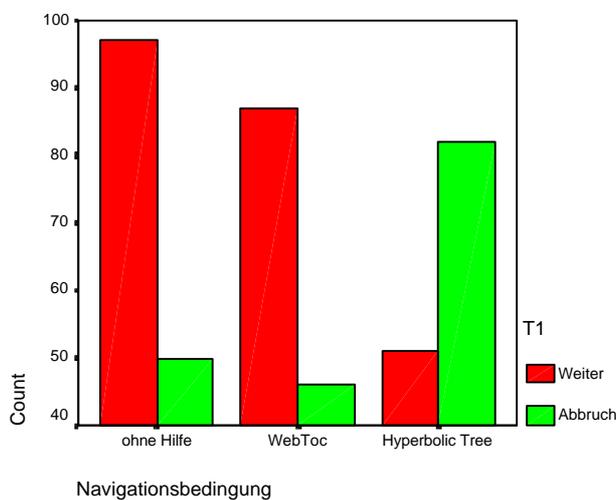


Abbildung 43: Dropout-Fälle ($n=178$) zum ersten Meßzeitpunkt in Abhängigkeit der Navigationsbedingung

Der Chi-Quadrat-Test (nach Pearson) für diesen Zusammenhang (Einfluß der Navigationsbedingung) weist einen signifikanten Anteilunterschied nach, $\chi^2(3, N=413)=27.55, p<.00$.

Zur Analyse des Zusammenhangs von Bildschirmauflösung und Dropout-Quote wurden die gemessenen Werte für die Auflösung in zwei Kategorien zusammengefasst. Bildschirmauflösungen bis zu 800x600 Punkten wurden als „geringe Auflösung“ klassifiziert, Auflösungen über dieser Grenze als „hohe Auflösung“.

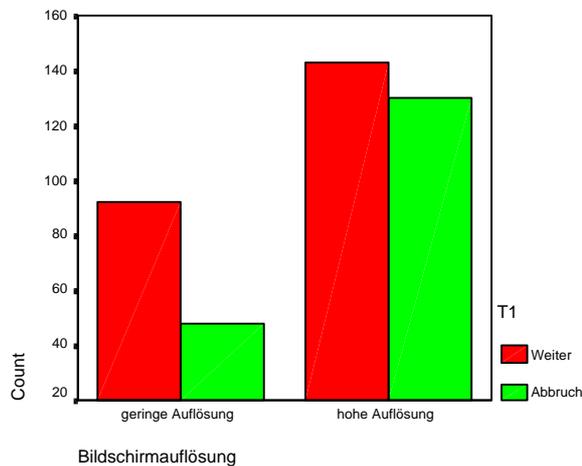


Abbildung 44: Dropout-Fälle ($n=178$) in Abhängigkeit der Bildschirmauflösung

Der Chi-Quadrat-Test (nach Pearson) für diesen Zusammenhang (Einfluß der Bildschirmauflösung) weist ebenfalls einen signifikanten Anteilsunterschied auf, $\chi^2(1, N=413)=6.7$, $p<.05$. Im Vergleich der erwarteten und gemessenen Zellenmittelwerte für diese Bedingungskombination zeigt sich, daß dieser signifikante Unterschied auf eine höhere Dropout-Quote bei Vpn mit hoher Auflösung zurückzuführen ist bzw. auf eine niedrigere Dropout-Quote als erwartet bei Vpn mit geringer Auflösung.

Als weiterer möglicher Faktor für eine Beteiligung am Dropout-Effekt kommt das Betriebssystem auf dem Computer der jeweiligen Vp in Frage. Da sich für die Zusammenhänge dieser Untersuchung relevante Effekte voraussichtlich nur anhand prinzipieller Aspekte der Systemarchitektur zeigen werden, wurden für die folgenden Analyse zusätzlich die Betriebssysteme der Microsoft-Familie in einer Kategorie zusammengefasst und gegen die Gruppe der anderen Betriebssysteme abgehoben. Ein entsprechender Chi-Quadrat-Test (nach Pearson) weist über alle Betriebssysteme einen signifikanten Anteilsunterschied für T1 nach, $\chi^2(4, N=413)=19.2$, $p<.01$. Auch der Vergleich der Werte für Microsoft-Betriebssysteme und Andere ergibt einen signifikanten Unterschied, $\chi^2(1, N=413)=16.56$, $p<.00$.

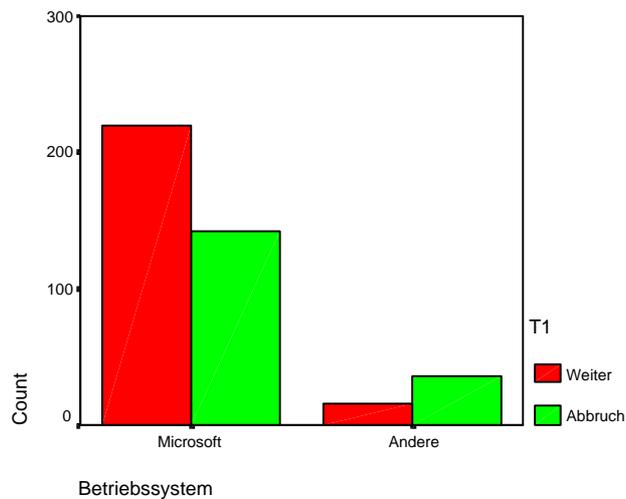


Abbildung 45: Dropout in Abhängigkeit des Betriebssystems (Microsoft-Systeme vs. Andere)

Der Vergleich der gemessenen Werte legt nahe, daß der signifikante Unterschied darauf zurückzuführen ist, daß in der Gruppe der Microsoft-Betriebssysteme weniger Abbrüche stattfanden als bei anderen Betriebssystemen.

Für den Faktor „Browser“ ergaben sich keine signifikanten Unterschiede, $\chi^2(1, N=412)=.00$.

Um den Zusammenhang zwischen den Einflußfaktoren „Auflösung“ und „Betriebssystem“ näher zu beleuchten (bzw. deren kombinierte Wirkung), wurde zusätzlich ein Gesamt-Chi-Quadrat mit beiden Bedingungen in Kombination ausgeführt (vgl. Tabelle 32).

Tabelle 31: kombinierter Chi-Quadrat-Test Auflösung*Betriebssystem*Abbruch

T1				Bildschirmauflösung		Total
				geringe Auflösung	hohe Auflösung	
Weiter	Betriebssystem	Microsoft	Count	92	127	219
			Expected Count	85,7	133,3	219,0
			Residual	6,3	-6,3	
	Andere	Count	0	16	16	
		Expected Count	6,3	9,7	16,0	
		Residual	-6,3	6,3		
	Total		Count	92	143	235
			Expected Count	92,0	143,0	235,0
	Abbruch	Betriebssystem	Microsoft	Count	44	98
Expected Count				38,3	103,7	142,0
Residual				5,7	-5,7	
Andere		Count	4	32	36	
		Expected Count	9,7	26,3	36,0	
		Residual	-5,7	5,7		
Total		Count	48	130	178	
		Expected Count	48,0	130,0	178,0	

Die Anteilsunterschiede sind für beide Stufen der Variablen „Abbruch“ signifikant, Abbruch=0 (nicht-Abbrecher): $\chi^2(1, N=413)=11.04$; $p<.01$; Abbruch=1 (Abbrecher): $\chi^2(1, N=413)=5.76$; $p<.02$.

Analyse für Zeitpunkt T2 (Präsentation der zweiten Frage)

Die Analysen für T2 (Präsentation der zweiten Frage) ergeben in keiner verfügbaren Bedingung einen signifikanten Unterschied.

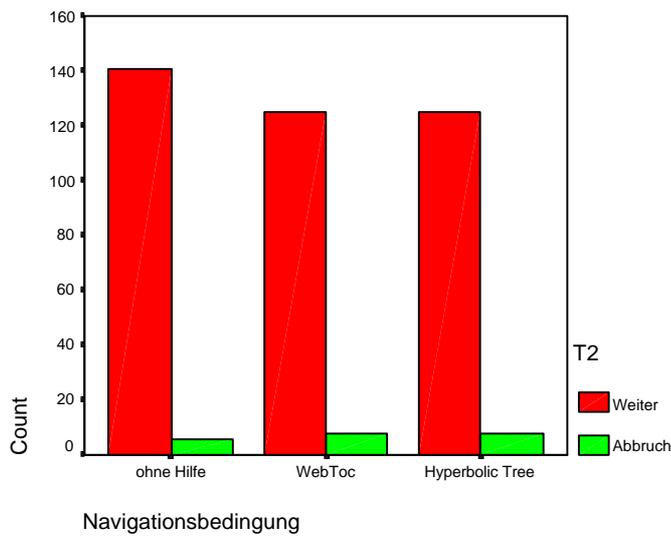


Abbildung 46: Dropout-Fälle ($n=22$) zum Zeitpunkt T2 in Abhängigkeit von der Navigationsbedingung

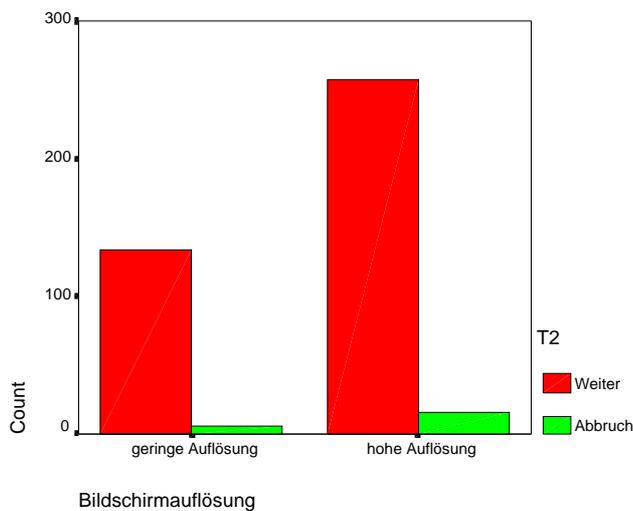


Abbildung 47: Dropout-Fälle ($n=22$) zum Zeitpunkt T2 in Abhängigkeit von der Bildschirmauflösung

Analyse für alle Dropout-Fälle

Zusätzlich zu den Analysen über einzelne Meßzeitpunkte interessiert noch die Betrachtung des Dropouts über den Verlauf der ganzen Untersuchung in Abhängigkeit der verfügbaren Variablen.

Erwartungsgemäß ergibt sich für die Navigationsbedingung ein signifikanter Anteilsunterschied der Abbrecher, $\chi^2(2, N=413)=25.06$, $p<.00$.

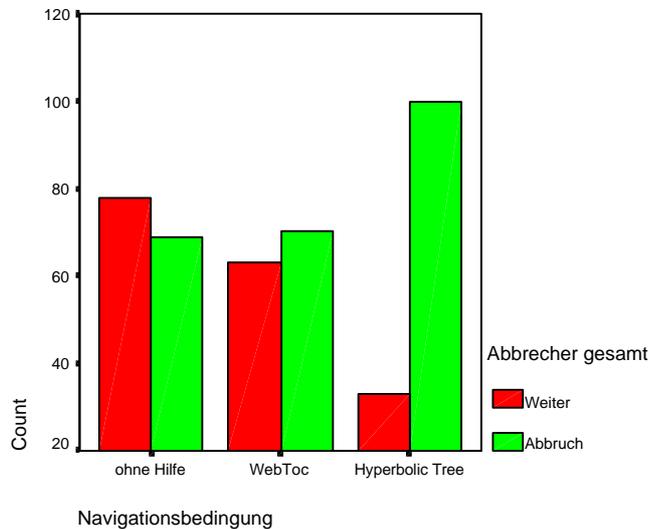


Abbildung 48: Dropout über alle Meßzeitpunkte in Abhängigkeit der Navigationsbedingung

Wie aus Abbildung 48 bereits ersichtlich wird, zeigt auch in diesem Fall die Analyse der erwarteten und beobachteten Mittelwerte einen Wert deutlich über dem erwarteten in der Bedingung *Hyperbolic-Tree* (Erwartet: 77, Beobachtet: 100).

Ein signifikanter Zusammenhang für die Bildschirmauflösung und die Dropout-Quote konnte nicht nachgewiesen werden $\chi^2(1, N=413)=3.6$, ebenso verhält es sich für den Faktor des Browsers, $\chi^2(1, N=412)=.00$.

Bei der Betrachtung des Faktors „Betriebssystem“ ergibt sich folgendes Bild: Eine Analyse über alle erfassten Betriebssysteme ergibt keinen signifikanten Unterschied, $\chi^2(4, N=413)=12.62$, allerdings wird der Anteilsunterschied im Falle der Unterscheidung von Microsoft-Systemen und anderen Betriebssystemen signifikant, $\chi^2(1, N=413)=8.85$, $p<.05$.

Um den Zusammenhang zwischen den Einflußfaktoren „Auflösung“ und „Betriebssystem“ näher zu beleuchten (bzw. deren kombinierte Wirkung) wurde zusätzlich ein Gesamt-Chi-Quadrat mit beiden Bedingungen in Kombination ausgeführt (vgl. Tabelle 32).

Tabelle 32: kombinierter Chi-Quadrat-Test Auflösung*Betriebssystem*Abbruch

AB_GES				Bildschirmauflösung		Total
				geringe Auflösung	hohe Auflösung	
0	Betriebssystem	Microsoft	Count	68	94	162
			Expected Count	63,3	98,7	162,0
			Residual	4,7	-4,7	
	Andere	Count	0	12	12	
		Expected Count	4,7	7,3	12,0	
		Residual	-4,7	4,7		
	Total	Count	68	106	174	
		Expected Count	68,0	106,0	174,0	
	1	Betriebssystem	Microsoft	Count	68	131
Expected Count				59,9	139,1	199,0
Residual				8,1	-8,1	
Andere		Count	4	36	40	
		Expected Count	12,1	27,9	40,0	
		Residual	-8,1	8,1		
Total		Count	72	167	239	
		Expected Count	72,0	167,0	239,0	

Die Anteilsunterschiede sind für beide Stufen der Variablen „Abbruch“ signifikant, Abbruch=0 (nicht-Abbrecher/innen): $\chi^2(1, N=413)=8.26$; $p<.01$; Abbruch=1 (Abbrecher): $\chi^2(1, N=413)=9.24$; $p<.01$.

7. Analyse der Ergebnisse und Diskussion

7.1 Zusammenfassung der Ergebnisse

7.1.1. Ergebnisse der Suchaufgaben

Obwohl in der konfirmatorischen Faktorenanalyse die zweifaktorielle Struktur zur Klassifizierung der hier verwendeten Suchaufgaben bestätigt werden konnte und entsprechende Niveauunterschiede für die zugehörige abhängige Variable sich ebenfalls als statistisch signifikant erwiesen (s.u.), soll hier nochmals auf die berichtete (vgl. 6.5.1) schlechte psychometrische Qualität der Skalen eingegangen werden. Die Interpretation der Ergebnisse - insbesondere für die Skala „Struktur-Fragen“ - muß unter Vorbehalt dieser Werte erfolgen. In weiteren Untersuchungen sollte dann versucht werden, in höherem Maß reliable Skalen zur Erfassung von unterschiedlichen Fragentypen zu entwickeln und die vorliegenden Fragestellung anhand derer abermals zu überprüfen. Allerdings ist mit diesem Befund auch ein grundlegendes methodisches Problem von Online-Untersuchungen angesprochen, denn die stärkere zeitliche Beschränkung in derartigen Untersuchungen ist für die Konstruktion reliabler Skalen - insbesondere in komplexen Untersuchungen - ein prinzipielles Hindernis und dürfte nur unter Inkaufnahme noch stärkerer Dropout-Daten zu korrigieren sein.

Die Ergebnisse zeigen im multivariaten Vergleich einen klaren Einfluß der Navigationsbedingung auf die Performanz und die Anzahl richtiger Antworten bei den Suchaufgaben. Auch die hinzugenommene Kovariate *spatial ability* kovariiert mit den genannten Faktoren signifikant.

Betrachtet man die Leistungen über alle Fragen im univariaten Vergleich, so zeigt sich ein signifikanter Effekt für die Bearbeitungszeit in der Weise, daß *User* des *Hyperbolic-Tree* höhere Bearbeitungszeiten zeigen als *User* von *WebToc* und auch die *User* ohne zusätzliche Navigationshilfen. Der Unterschied zwischen *WebToc* und der Gruppe ohne Tool wird dabei nicht signifikant.

Schlüsselt man die Ergebnisse auf hinsichtlich der beiden Fragentypen *Retrieval-Fragen* und *Strukturfragen*, zeigt sich folgendes Bild:

Sowohl im multivariaten Vergleich als auch in den univariaten Einzelvergleichen hat die Navigationsbedingung einen signifikanten Einfluß auf die Beantwortung der *Retrieval-Fragen*, und zwar hinsichtlich Bearbeitungszeit und Anzahl richtiger Antworten. Versuchspersonen, die die Fragen mit der Hilfe von *WebToc* beantworteten, erzielten mehr richtige

Antworten als solche, die mit *Hyperbolic-Tree* arbeiteten oder in der *ohne Hilfe*-Bedingung. Hinsichtlich der Bearbeitungszeit verursachten die User des *Hyperbolic-Tree* die längsten Zeiten, die Unterschiede sind sowohl hinsichtlich *WebToc* als auch in Bezug auf die Bedingung *ohne Hilfe* signifikant. Die Unterstützung durch *WebToc* trug allerdings nicht zu signifikant verkürzten Bearbeitungszeiten im Vergleich mit der *ohne Hilfe*-Bedingung bei. Die Kovariate *spatial ability* zeigt einen Einfluß auf die Anzahl richtiger Antworten, nicht jedoch auf die Bearbeitungszeit für diesen Fragentyp.

Bezüglich der Beantwortung von Fragen aus der Skala „Struktur-Fragen“ kann der Faktor *Navigationshilfe* hingegen nicht ausreichend Varianz aufklären. Allerdings wird für diese Fragen die Kovariate *spatial ability* signifikant, was zumindest die Annahme der Validität dieser Skala zu stützen vermag.

Die Annahme unterschiedlicher Fragentypen wird ebenfalls von der Analyse der AV-Aspekte für „Fragentyp“ gestützt. Hier zeigt sich im multivariaten Vergleich, aber auch in den univariaten Analysen zur Anzahl richtiger Antworten und zur Bearbeitungszeit ein signifikanter Unterschied derart, daß Struktur-Fragen mehr Bearbeitungszeit in Anspruch nehmen und offenbar schwieriger in der Beantwortung sind (d.h. es resultieren - trotz längerer Bearbeitung - weniger richtige Antworten). Hierbei ergibt sich eine interessante Interaktion von Fragentyp und Navigationsbedingung bezüglich der Bearbeitungsdauer, und zwar schneiden Benutzer des *Hyperbolic-Tree* hinsichtlich der Bearbeitungsdauer bei Retrieval-Fragen noch schlechter ab, als bei der Beantwortung von Struktur-Fragen.

Die berichteten Effekte konnten auch bei explorativen Analysen unter Einbeziehung des Gruppenfaktors „Interneterfahrung“ gesichert werden. Zwar hatte dieser Faktor einen (erwartbaren) signifikanten Einfluß auf die Beantwortung der Fragen, allerdings blieb der Einfluß der Navigationsbedingung dennoch signifikant (näheres siehe unter 7.3.1).

Zusammenfassend kann für die Bearbeitung der Suchaufgaben gesagt werden, daß die zusätzliche Visualisierungshilfen im Falle von Retrieval-Fragen offensichtlich dazu beitragen können, daß in gleicher Zeit mehr Fragen richtig beantwortet werden konnten. Allerdings gilt dies nur für die Visualisierung, die das Tool *WebToc* anbietet. Die Benutzer des *Hyperbolic-Tree* hingegen schneiden für diesen Fragentyp hinsichtlich der Bearbeitungsdauer schlechter ab als die Benutzer in beiden Konkurrenzbedingungen, und im Hinblick auf die

Anzahl richtiger Antworten liegen sie nicht über dem Niveau der Gruppe *ohne Hilfe*. *Hyperbolic-Tree* kann demnach für diesen Fragentyp als nicht hilfreich eingestuft werden.

Für die Beantwortung von komplexen Strukturfragen konnte keines der eingesetzten Tools die Benutzer derart unterstützen, daß sich bessere Leistungen gezeigt hätten. Die signifikante Wirkung des Faktors *spatial ability* gibt Anlaß zu der Vermutung, daß diese Art von Fragen (wie auch Retrieval-Fragen) vermutlich dennoch unter Zuhilfenahme räumlicher Operationen gelöst werden und eine Art der Visualisierung, die diesen Problemlösemodus unterstützt, angezeigt wäre. Da insbesondere für den *Hyperbolic-Tree* genau diese Art der Unterstützung reklamiert wird (vgl. 2. 2. 6. 3), ist es von besonderem Interesse, wie sich die Einschätzung dieser Qualitäten anhand weiterer Daten im Folgenden verdichten lässt.

Der festgestellte Effekt ist auch insofern interessant, da die Untersuchung aufgrund der Entscheidung für ein Online-Experiment naturgemäß mit zahlreichen methodischen Unzulänglichkeiten behaftet ist. Allerdings sind viele der zu erwartenden Störeffekte in einer Weise kontrolliert worden, daß sie entweder direkt in die Analyse einbezogen werden konnten (*spatial ability* und „Interneterfahrung“), oder hypothesenkonservativ gewirkt hätten (z. B. der zur Verfügung stehende Platz auf dem Bildschirm). Zusätzlich gibt es Anlaß zur Vermutung, daß der starke ungleichmäßige Dropout (s.u.) sich eher zugunsten der Treatment-Bedingung *Hyperbolic-Tree* ausgewirkt haben dürfte (weil wahrscheinlich eher motivierte Versuchspersonen weitergemacht haben).

In einer aktuellen Studie (Pirolli, Stuart & Van der Wege, 2000) hat eine Gruppe von Wissenschaftlern/Wissenschaftlerinnen der Xerox-Forschungslabors um einen der Entwickler des *Hyperbolic-Tree* (P. Pirolli) einen in vielen Punkten mit der vorliegenden Untersuchung vergleichbaren Versuch unternommen, die Informationssuche mit dem *Hyperbolic-Tree* im Vergleich mit anderen Visualisierungen zu erheben. Der Unterschied in der Versuchsanordnung bestand im wesentlichen darin, daß nur *Hyperbolic-Tree* und ein hierarchischer Browser (Microsoft Explorer) untersucht wurden, und daß es sich um ein Laborexperiment handelte. Die beiden Laborexperimente (es wurde ein zweites Experiment mit leicht modifizierten Bedingungen durchgeführt) wurden jeweils mit Gruppen von acht Psychologie-Studierenden der Universität Stanford durchgeführt. Dabei konnte im ersten Experiment kein signifikanter Unterschied in der Bearbeitungszeit festgestellt werden. Aufgrund der Analysen der deskriptiven Ergebnisse führten die Autoren dann ein zweites Experiment mit einer kleinen Auswahl der ursprünglichen Aufgaben durch, die ausschließlich aus dem Be-

reich einfacher Retrieval-Fragen entnommen wurden. Hier konnte dann ein signifikanter Vorteil des Hyperbolic-Tree-Browsers festgestellt werden, allerdings nur für Suchaufgaben mit hohem „information scent“ (zur Theorie des „information scent“ vgl. 2. 1. 3. 2). Zusätzlich wurde festgestellt, daß die User des *Hyperbolic-Tree* mehr Knoten in der gleichen Zeit besuchten.

Wie aus der oben geführten Diskussion der Ergebnisse der vorliegenden Studie ersichtlich ist, stehen diese in diametralem Gegensatz zu den Ergebnissen der Studie von Pirolli et al.(2000). Ob dieser eklatante Unterschied dem Umstand zuzuschreiben ist, daß in der vorliegenden Untersuchung nicht zwischen Fragen mit hohem und niedrigem „information scent“ unterschieden wurde, müßten weitere Untersuchungen klären.

7.1.2. Usability-Beurteilung

Die vom Autor des *usability*-Instruments (Lewis, 1995) berichteten hervorragenden Werte für Reliabilität des Verfahrens zur Erfassung der (subjektiven) *usability* konnten reproduziert werden, ähnliches gilt auch für die 3-faktorielle Struktur des Instruments.

Hinsichtlich der Ausprägungen dieser Einschätzungen unter den Stufen des Faktors Navigationshilfe ergibt sich folgendes Bild:

Im Hinblick auf den Gesamtwert für *usability* zeigt sich ein signifikanter Mittelwertsunterschied, und zwar beurteilen die Versuchspersonen den *Hyperbolic-Tree* deutlich schlechter als die Bedingung *ohne Hilfe*. Zwischen *WebToc* und den anderen beiden Bedingungen ergeben sich (in Post-Hoc-Analysen) keine derartigen Unterschiede.

Für den Faktor „system-usefulness“ zeigt sich ein ähnliches Bild, allerdings wird hier nicht nur die Bedingung *ohne Hilfe* deutlich besser eingeschätzt, als *Hyperbolic-Tree*, sondern auch der Unterschied zwischen *WebToc* und *Hyperbolic-Tree* wird signifikant. D.h. *Hyperbolic-Tree* wird bezüglich dieses Faktors als deutlich schlechter im Vergleich mit beiden konkurrierenden Bedingungen eingeschätzt.

Hinsichtlich des Faktors „information quality“ zeigt sich wiederum ein signifikanter Unterschied zwischen *Hyperbolic-Tree* und *ohne Hilfe*. Auch hier wird *Hyperbolic-Tree* als signifikant schlechter beurteilt.

Die Analysen für den Faktor „interface quality“ ergeben keinen signifikanten Unterschied.

Zusammenfassend lässt sich für die Beurteilung der *usability* demnach Folgendes sagen: der *Hyperbolic-Tree* wird auch im Hinblick auf *usability* deutlich schlechter beurteilt als die

konkurrierenden Bedingungen. Hierbei ist zwischen den Navigationsbedingungen *WebToc* und *ohne Hilfe* kein Unterschied festzustellen, mit der Ausnahme des Faktors „system-usefulness“. Hier beurteilen die Versuchspersonen den Hyperbolic-Tree auch im Vergleich mit *WebToc* schlechter.

Insgesamt ist es erstaunlich, daß die angebotenen Hilfsmittel durchgängig nicht als solche verstanden wurden, zumindest nicht für die Beurteilung der *usability*. Dies wird vor allem im Zusammenhang mit der Analyse der Ergebnisse der Suchaufgaben interessant, da hier *WebToc* z.T. sich als klar überlegene Bedingung erwiesen hat.

7.1.3. Logfile-Maße

Hinsichtlich der Logfile-Maße wurde deutlich, daß in mehrfacher Hinsicht deutliche Unterschiede zwischen den Versuchspersonen in den verschiedenen Navigationsbedingungen bestehen.

So zeigte sich, daß in der Bedingung *Hyperbolic-Tree* deutlich weniger Knoten besucht wurden als in beiden anderen Bedingungen. Auch wenn es denkbar ist, daß diese Werte einer Verzerrung aufgrund des Designs der Untersuchung (Diskussion siehe unter 4. 4. 3. 1) unterliegen, so überrascht dieser Befund dennoch insbesondere wiederum im Zusammenhang mit den diesbezüglichen Ergebnissen, die Pirolli et al. (2000) in der bereits erwähnten Untersuchung berichten. Auch hier kommen sie zu einem entgegengesetzten Ergebnis, indem sie feststellen „participants visited more nodes in the tree with the Hyperbolic browser“ (Pirolli et al., 2000, S. 13).

Zur Interpretierbarkeit dieses Effektes trägt (neben der hohen Signifikanz; $p < .01$ - auch in den Alpha-adjustierten Post-Hoc-Vergleichen) allerdings eine weitere Analyse der Logfiledaten bei, bei der gezeigt werden konnte, daß kein signifikanter Unterschied zwischen den beiden Tool-Bedingungen hinsichtlich der Anzahl an Knoten, die aus dem Tool heraus aufgerufen wurden, bestand. Dies legt den Schluß nahe, daß diese beiden Bedingungen auch hinsichtlich des Umfangs der Navigation innerhalb des Tools vergleichbar sein könnten, welche aus technischen Gründen nicht aufgezeichnet werden konnte. Dies wiederum würde den methodischen Einwand gegen die Interpretation des Unterschiedes in der Anzahl insgesamt besuchter Knoten erheblich abschwächen.

Auch bezüglich der graphentheoretischen Maße Compactness und Stratum ergaben sich signifikante Unterschiede zwischen den Navigationsbedingungen. So wird deutlich, daß hin-

sichtlich der Werte für Stratum in der Bedingung *ohne Hilfe* höhere Werte erzielt wurden als in der Bedingung *WebToc*, die ihrerseits höhere Werte zeitigte als die Bedingung *Hyperbolic-Tree*. In Bezug auf die Werte für Compactness ergab sich der gegenläufige Befund, was z.T. im Konzept dieser Größe begründet liegen dürfte. Allerdings weisen Botafogo et al. (1992) und McEneaney (1999) explizit auf die dennoch gegebene Unabhängigkeit der zugrundeliegenden Konstrukte hin. So zeigt sich für Compactness auch nur ein signifikanter Unterschied zwischen *Hyperbolic-Tree* und den anderen beiden Bedingungen. Diese Ergebnisse machen deutlich, daß User in der Bedingung *ohne Hilfe* in ihren Navigationen durch den Hypertext offenbar einen stärker vernetzten Pfad „herstellen“. Auch die Gruppen *Hyperbolic-Tree* und *WebToc* unterscheiden sich diesbezüglich.

Explorative Analysen, insbesondere innerhalb der nicht von möglichen methodischen Artefakten belasteten Bedingung *ohne Hilfe*, ergaben einen Zusammenhang zwischen den Maßen für Compactness und Stratum und der Anzahl richtiger Antworten, und zwar zeigt sich ein Zusammenhang zwischen hohen Ausprägungen für Compactness und niedrigen Ausprägungen für Stratum mit der Anzahl richtiger Antworten. Dies würde wiederum einen Vorteil des vernetzteren Pfades nahelegen. In den Bedingungen *WebToc* und *Hyperbolic-Tree* ließen sich diese Zusammenhänge nicht zeigen, hier wurde hingegen eine Korrelation zwischen niedrigen Werten für Stratum und hohen Werten in der Bearbeitungsdauer deutlich.

Unter Umständen lassen sich die Zusammenhänge mit der Anzahl richtiger Antworten so deuten, daß ein vernetzterer Pfad mit weniger linearen Sequenzen ein Zeichen für einen effektiveren Umgang mit dem Hypertext anzeigt, da lineare Sequenzen in einer gezielten Suchaufgabe als Hinweis auf eine uneffektive Suchstrategie gedeutet werden können.

Gestützt wird diese Vermutung durch eine weitere explorative Analyse des Aufrufs bestimmter Seitentypen, der aus den Logfiledaten ermittelt werden konnte. Hier zeigt sich, daß in der Bedingung *ohne Hilfe* häufiger Index- und Suchseiten aufgerufen wurden. Dieser Unterschied ergibt sich auch für den Vergleich von *WebToc* und *Hyperbolic-Tree*. Unter Umständen war es den Benutzern des *Hyperbolic-Tree* nicht möglich, eine ähnlich effektive Suchstrategie zu verwenden, was sich in allen Kennwerten der Logfilemaße (und auch in anderen Analysen) zeigt.

7.2 Dropout-Analyse

Das Problem selbstselegierender Stichproben wird als eines der zentralen Probleme der Online-Forschung überhaupt betrachtet (vgl. Reips, 2000; Bandilla 1999).

Tatsächlich konnte auch in dem hier berichteten Online-Experiment ein erheblicher Dropout festgestellt werden (über 50%), der sich zudem nicht zufällig auf die zugewiesenen experimentellen Bedingungen zu verteilen schien.

Glücklicherweise konnten für alle Versuchspersonen, die das Experiment nach dem ersten Betätigen des „Start-Knopfes“ verließen, dennoch eine Reihe von Daten erhoben werden, die eine detaillierte Analyse zulassen.

Demnach ergeben sich für diesen Stichprobeneffekt folgende Zusammenhänge:

Zunächst ist auffällig, daß deutlich mehr Versuchspersonen das Experiment vorzeitig beendeten, die zu Beginn als Navigationsbedingung den *Hyperbolic-Tree* zugewiesen bekommen hatten.

Als weitere Einflußfaktoren konnten die Bildschirmauflösung und das verwendete Betriebssystem gesichert werden - der verwendete Browsertyp spielte offenbar keine Rolle.

Der zunächst wenig einleuchtende Zusammenhang, daß offenbar mehr Vpn das Experiment verließen, die mit einer höheren Bildschirmauflösung arbeiteten, konnte durch eine kombinierte Analyse auf den eher plausiblen Zusammenhang mit dem eingesetzten Betriebssystem zurückgeführt werden.

Demnach sieht es so aus, daß offenbar Versuchspersonen mit Betriebssystemen, die nicht aus der Microsoft-Familie stammen, deutlich häufiger abbrechen, als solche mit einem Microsoft-Betriebssystem.

Diesbezüglich gibt es Anlaß zu der Vermutung, daß sich hinter diesen Analysen eine Unterscheidung von Experten/innen und „Alltags-User“ verbirgt, denn sowohl hohe Bildschirmauflösungen, als auch „ungewöhnliche“ Betriebssysteme werden häufiger von Experten/innen eingesetzt. Leider stehen keine weiteren Daten zur Verfügung, um diese Hypothese zu stützen.

Zusammenfassend läßt sich sagen, daß ein erheblicher Teil des Dropouts auf den *Hyperbolic-Tree* zurückzuführen ist, und zwar vor allem zu Beginn der Untersuchung. Es liegt deshalb nahe, einen Zusammenhang mit der hohen Ladezeit anzunehmen, wobei dies sicherlich nicht der einzige Faktor sein kann, da auch *WebToc* ähnlich hohe Ladezeiten zu Beginn verursacht, und der Effekt auch deskriptiv zum Zeitpunkt T2 zu beobachten ist.

7.3 Zusätzliche Fragen zum Experiment

Eine Annahme bezüglich der unterschiedlichen Akquisition von Überblicks- und Strukturwissen bezüglich des Untersuchungsmaterials lautete, daß bessere mentale Repräsentation zu einer besseren Abschätzung des Gesamtumfanges der Web-Site des deutschen Bundestages führen würde. Die Ergebnisse der dazu eingeführten Schätzaufgabe können diese Annahme nicht bestätigen - der Unterschied in den Schätzungen ist nicht signifikant.

Hinsichtlich der weiteren drei Zusatzfragen zum Experiment ergab sich lediglich für die Fragen „Es fiel mir leicht, die Fragen zu beantworten“ ein signifikanter Unterschied, der sich in der Post-Hoc-Analyse als ein Unterschied zwischen den Bedingungen *WebToc* und *Hyperbolic-Tree* herausstellte. Benutzer des *Hyperbolic-Tree* verneinten diese Frage im Vergleich zu solchen in der *WebToc*-Bedingung deutlich häufiger.

Im Hinblick auf die weiteren Fragen ergab sich zwar kein signifikanter Unterschied, allerdings zeigt die deskriptive Analyse dennoch interessante Tendenzen.

So tritt das Phänomen des Orientierungsverlusts offenbar am stärksten im Zusammenhang mit dem *Hyperbolic-Tree* auf, was selbst anhand dieser inferenzstatistisch nicht gesicherten Tendenz in der Stichprobe angesichts der diesbezüglichen Hypothesen erstaunlich ist.

Ähnlich erstaunlich ist die deskriptiv zu beobachtende Tendenz für die Frage „Das Experiment hat mir gefallen“. Auch hier bestand die Erwartung, daß *Hyperbolic-Tree* aufgrund der aufwendigen graphischen Gestaltung eine bessere Bewertung erzielen würde als die anderen Bedingungen. Die Daten zeigen aber ein anderes Bild, denn die Werte sind für die Bedingung *ohne Hilfe* am besten, die schlechtesten Werte erzielt *Hyperbolic-Tree*.

Die Frage nach dem aufgabenspezifischen politischen Vorwissen, die zur Kontrolle eventueller Störeffekte eingefügt worden war, zeigt keine signifikanten Unterschiede über die Stufen des Faktors *Navigationshilfe*.

7.3.1. Weitere exploratorische Analysen

Hinsichtlich der durchschnittlichen Bearbeitungsdauer zeigen exploratorische Analysen zwei weitere interessante Effekte, die auf Probleme insbesondere des *Hyperbolic-Tree*-Browsers hindeuten. Zum einen kann festgestellt werden, daß die Versuchspersonen mit dem Einführungstutorial für *Hyperbolic-Tree* signifikant mehr Zeit benötigen, als Personen denen *WebToc* präsentiert wurde. Dies wirkt sich aber nicht auf die Qualität und Performanz bei der Beantwortung der Suchfragen mit dem jeweiligen Tool aus.

Der zweite interessante Befund ergibt sich im Zusammenhang mit den Analysen zu möglichen Lerneffekten in den unterschiedlichen Stufen des Faktors *Navigationshilfe*. Hier zeigt sich bei der Beantwortung von Struktur-Fragen in den Bedingungen *WebToc* und *ohne Hilfe* über den Verlauf der Erhebung ein deutlicher Lerneffekt, insbesondere vor dem Hintergrund des nicht-ansteigenden Schwierigkeitsniveaus der zu dieser Skala gehörenden Items (vg. Tabelle 6). Für den *Hyperbolic-Tree* hingegen scheint dieser auszubleiben. Dieses Bild wird auch durch die Veranschaulichung des entsprechenden Effekts über alle sechs Fragen bestätigt. Offenbar kann der Hyperbolic-Tree-Browser also selbst bei längerer Bearbeitung für die Beantwortung von Suchfragen keinen Vorteil bieten. Eventuell wären deutlich längere Einarbeitungszeiten nötig (eine Vermutung, die übrigens Pirolli auch bezüglich der nicht reproduzierbaren Ergebnisse des CHI '97-Browse-off äußert, bei dem der Hyperbolic-Tree-Browser sich als deutlich überlegen erwies - allerdings wurde er von einem der Entwickler selbst bedient; Pirolli et al, 2000). Im Zusammenhang mit Suchsituationen im Internet wären derartig lange Einarbeitungszeiten allerdings sicherlich intolerabel.

Weitere exploratorische Analysen für den Post-Hoc-Faktor „Internetenerfahrung“ zeigten hinsichtlich der inferenzstatistischen Ergebnisse zunächst nur die Stabilität der zuvor ermittelten Ergebnisse bezüglich der Wirksamkeit des Faktors *Navigationshilfe*.

Deskriptiv zeigten sich allerdings einige interessante Tendenzen für die Bearbeitung von Strukturfragen. Es deutet sich dabei an, daß erfahrene Internet-User mit dem *Hyperbolic-Tree* in etwa gleich viele richtige Antworten erzielen wie in den anderen Bedingungen. Dies steht im klaren Gegensatz zur Gruppe der unerfahrenen User (vgl. Abbildung 29). Andererseits erzielen die erfahrenen User bei der Bearbeitung von Strukturfragen gleichzeitig mit dem *Hyperbolic-Tree* die höchsten Bearbeitungszeiten, was in der Tendenz den Werten der unerfahrenen User diametral entgegengesetzt ist. Im Ergebnis muß natürlich gesagt werden, daß dieser deskriptiv erkennbare, leichte Vorteil der erfahrenen User in der Verwendung des *Hyperbolic-Tree* durch höhere Bearbeitungszeiten „erkauft“ wurde. Immerhin scheint es aber einen auffälligen Unterschied in der Bereitschaft zu geben, sich mit einem derartigen Tool auseinanderzusetzen. Sollte es eine derartige Tendenz tatsächlich geben, so spiegelt sie sich allerdings nicht in den *usability*-Bewertungen der User wieder, denn hier beurteilen die erfahrenen User die Umgebungen ähnlich wie die Unerfahrenen.

Die Korrelationen der Antwortausprägungen für die zusätzlichen Fragen zum Experiment, also z.B. „Das Experiment hat mir gut gefallen“ mit den abhängigen Variablen „Anzahl richtiger Fragen“ und „Bearbeitungsdauer“ zeigt in mehrfacher Hinsicht interessante Zusammenhänge. So scheint es einen starken Zusammenhang zwischen dem subjektiven Eindruck der Schwierigkeit des Experimentes und der Zufriedenheit mit demselben zu geben - was zunächst trivial ist- allerdings ist dieser Zusammenhang auch im Hinblick auf die „Anzahl richtiger Antworten“ deutlich zu erkennen. Da es sich um korrelative Zusammenhänge handelt, sind zwei Interpretationen dieses Zusammenhanges denkbar: Entweder ein motivationaler Faktor spielt bei der Bearbeitung der Fragen eine große Rolle (was gut mit Annahmen bezüglich der Charakteristik selbstselektierender Stichproben in Online-Untersuchungen in Einklang zu bringen wäre; Reips, 2000), d.h. die Zufriedenheit/positive Grundstimmung oder eine charakteristische Leistungsattribution bewirkt eine effektivere Problemlösestrategie für die Beantwortung der Fragen. Oder die das Gefühl den bearbeiteten Aufgaben gewachsen gewesen zu sein führt generell zu einer positiveren Beurteilung des Experiments und einer mildereren Einschätzung des Schweregrades der bearbeiteten Fragen. Die Tatsache, daß für die genannten Fragen zur Variablen „Bearbeitungsdauer“ keine substantiellen Korrelationen bestehen, stützt die zweite Interpretation, denn die Vermutung liegt nahe, daß sich eine höhere Motivation zur Bearbeitung der Fragen auch in einer höheren Bearbeitungsdauer niederschlagen müsste. Die diesen Überlegungen zugrundeliegenden Daten sind aber bei weitem nicht ausreichend, um entsprechende Schlüsse ziehen zu können. Allerdings könnte der Frage eines möglichen motivationalen Faktors in weiteren Untersuchungen größere Aufmerksamkeit gewidmet werden.

Im Rahmen der genannten Korrelationen fällt noch ein weiterer Befund auf: Es konnte keine Korrelation zwischen der Frage nach dem subjektiv empfundenen Orientierungsverlust und den Ausprägungen für *spatial ability* festgestellt werden. Da in anderen Analysen für den Faktor *spatial ability* eine vergleichsweise gute Varianzaufklärung nachgewiesen werden konnte, dürfte dieser (mit Einschränkungen aufgrund der dürftigen Daten) Befund eher als weiteres Argument gegen die Brauchbarkeit des Konstrukts „Orientierungsverlust“ angebracht werden. Auch hier wären weitere Erhebungen mit Instrumenten besserer psychometrischer Qualität aufschlußreich.

7.4 Diskussion

Insgesamt kann gesagt werden, daß für die Lösung von Suchaufgaben in komplexen Web-Sites eine entsprechende Visualisierung einen klaren Einfluß hat, der allerdings auch leicht ins Negative ausschlagen kann.

Zentral für die Zusammenfassung der Ergebnisse scheint eine getrennte Betrachtung der Visualisierungen des *Hyperbolic-Tree* und von *WebToc* zu sein.

Der *Hyperbolic-Tree* stellt die konsequenteste und technisch aufwendigste Implementierung dar. Leider zeigen die Ergebnisse, daß in keiner Bedingung mit diesem Tool bessere Werte erzielt werden konnten, als wenn überhaupt keine spezielle Unterstützung angeboten wurde. Erschwerend für diesen Befund kommen die klaren Abwertungen der User hinsichtlich der *usability* des Tools hinzu, aber auch die indirekt erfaßten *usability*-Probleme durch den starken Dropout und die nicht vorhandene Lernkurve (trotz längerer Einarbeitungszeit). In der Analyse der möglichen Ursachen für das schlechte Abschneiden des *Hyperbolic-Tree*-Browsers sollte zwischen *usability*-Problemen und prinzipiellen Einwänden gegen die Art der Visualisierung unterschieden werden, wenngleich in vielen Punkten Überschneidungen vorhanden sein dürften.

Für den Bereich *usability* lassen sich zunächst technische Kritikpunkte finden. Die Anwendung verursacht enorme Ladezeiten und ist auch im Betrieb auf eine gute Rechnerausstattung angewiesen. Zusätzlich zeigte sich bei der Planung der vorliegenden Untersuchung, daß *Hyperbolic-Tree* bei Zahlen von 4000 Knoten und mehr nur noch schwer bedienbar ist und zu häufigen Abstürzen des Computers führt (die Web-Site des deutschen Bundestages beinhaltet in der offiziellen Fassung mehr als 13 000 Seiten). Ein weiteres Problem dürfte in einem Grundprinzip der Darstellung des *Hyperbolic-Tree* begründet liegen. Schon die Autoren/innen des *Hyperbolic-Tree* weisen in ihrem einführenden Artikel zur ersten Vorstellung darauf hin, daß ein Problem der verzerrten Darstellung großer Informationsmengen darin besteht, daß die gesamte Darstellung ständig neu berechnet wird, selbst bei kleineren „Manövern“ in der Baumstruktur (Lamping, Rao & Pirolli, 1995). Dies hat zur Folge, daß zentrale Punkte und ganze Segmente einer Struktur praktisch nie an der gleichen Stelle innerhalb der Gesamtdarstellung auftauchen, ja es kann sogar vorkommen, daß durch einen einzigen Mausklick die gesamte Darstellung sich fundamental ändert und der gerade aktive Bereich als nahezu unsichtbarer Punkt irgendwo am Rand des Bildschirms dargestellt wird. Erschwerend für diesen Effekt kommt hinzu, daß die Anwendung über keine Funktion verfügt, um einen solchen Schritt rückgängig zu machen, vielleicht sogar über mehrere Stufen,

um einen Suchprozeß nachvollziehen zu können. Aus diesem Grund ist es vermutlich auch nicht gelungen, den Usern mit diesem Tool den Eindruck des „Lost in hyperspace“ zu ersparen, ja das Phänomen trat sogar mit dem *Hyperbolic-Tree* noch häufiger auf als in den anderen Bedingungen. Über weitere *usability*-Probleme mit dieser Anwendung kann im Prinzip nur spekuliert werden, aber es zeigte sich z. B. in der Voruntersuchung bereits ein Problem mit dem Icon-basierten Ansatz in der Menüführung des *Hyperbolic-Tree* (d.h. Befehle werden nur über kleine Icons angeboten). Die ständig verfügbare Online-Hilfe, die eigens für dieses Experiment verfaßt wurde, wurde allerdings praktisch nicht genutzt. Daß die Programmschnittstelle des *Hyperbolic-Tree* insgesamt als wenig intuitiv gelten kann, zeigen zusätzlich die erhöhte Bearbeitungsdauer für das einführende Tutorial und die schlechte Lernkurve über den Verlauf der Untersuchung.

Ein anderes Problem ist im Zusammenhang mit der hier intendierten Erleichterung der Bildung eines adäquaten mentalen Modells von großer Bedeutung. So muß vor dem Hintergrund der guten Ergebnisse der Versuchspersonen *ohne Hilfe*, aber auch der mit dem Tool *WebToc* die Frage gestellt werden, ob es sich beim *Hyperbolic-Tree* überhaupt um eine Visualisierung handelt, die die Aneignung von Informationen über Strukturen der angebotenen Informationen und das Auffinden von Informationen erleichtern kann. Ein wesentlicher Nachteil könnte darin begründet liegen, daß mit der Darstellung einer Web-Site als Baumstruktur in Kauf genommen wird, daß eine stark dekontextualisierte Umgebung präsentiert wird. So naheliegend die Vorteile einer solchen Visualisierung aus technischer Sicht sein mögen, insbesondere wenn es um die Lösung des Problems, möglichst viel Informationen auf einen Bildschirm „zu packen“ geht, so deutlich sind doch auch bei näherer Betrachtung die Nachteile für den Prozeß der Informationsverarbeitung, der ja für die Rezeption und die Arbeit mit einer solchen Seite zentral sein muß. Hier geben die vorliegenden empirischen Ergebnisse nun Anlaß zu der Vermutung, daß die oben diskutierten Einwände gegen eine Thematisierung der Verarbeitung von Hypertexten unter Zuhilfenahme solcher Konstrukte wie „Navigation“, „Orientierungsverlust“ usw. für das schlechte Abschneiden der Benutzer des *Hyperbolic-Tree* verantwortlich sein könnten.

Eine weitere mögliche Ursache für das schlechte Abschneiden des *Hyperbolic-Tree*-Browsers könnte in der Umsetzung des Programmes hinsichtlich der gewählten Metapher liegen. Wie oben bereits ausgeführt (Abschnitt 2. 1. 5. 2), hat Ortony (1979) eine elaborierte Theorie entwickelt, die geeignet ist, die Qualität von Metaphern zu beurteilen. Da es sich um ei-

nen wesentlichen Punkt in der Analyse der Ergebnisse handelt, soll diese Analyse kurz für die Metapher des *Hyperbolic-Tree* durchgeführt werden. Dabei müssen im wesentlichen Antworten für drei Fragen gefunden werden:

1. Handelt es sich wirklich um eine Metapher?
2. Knüpft die Metapher tatsächlich an vorhandenes Wissen an?
3. Wie stark sind die Überlappungen des Quell- mit dem Zielbereich (Nützlichkeit)?

Nach der Klassifikation von Ortony (1979) muß zur Beantwortung der ersten Frage geklärt werden, ob es Überlappungen zwischen den Attributen der beiden Objekte (als Elemente der Metapher) gibt und welcher Art diese sind.

Tabelle 33: Attributvergleich für die Baummetapher

Wichtigkeit	Hypertext	Verknüpfung	Baum
1.	besteht aus (Knoten)		ist (Pflanze)
2.	besteht aus (Verknüpfungen)		besteht aus (Stamm)
3.	ist (Text)		besteht aus (Ästen)
4.	ist (vernetzt)		besteht aus (Blättern)
5.	ist (hierarchisch)		besteht aus (Wurzeln)
6.	ist (computerbasiert)		verändert (sich)
7.	beinhaltet (Indexseiten)		
Domäne	Information		Biologie

Wie aus Tabelle 33 ersichtlich ist, besteht zunächst keine offensichtliche Überlappung von Attributen aus dem Quell- und Zielbereich. Außerdem liegt eine Domäneninkompatibilität vor. Wenn nun dennoch versucht werden soll, das Statement „Hypertexte sind wie Bäume“ zu analysieren, können anhand einer zusätzlichen Abstraktionsebene die Attribute „besteht aus (Knoten, Verknüpfungen“ und „besteht aus (Ästen, Blättern, Stamm, Wurzeln)“ in ein Meta-Attribut „besteht aus Elementen, die untereinander verknüpft sind“ überführt werden. Um die Domäneninkompatibilität zu überwinden wäre eine sehr abstrakte Meta-Domäne zu formulieren, etwa in der Art „Ein Bereich, in welchem Einheiten existieren, die durch die Relation ihrer Elemente untereinander konstituiert werden“. Somit wäre eine leichte Unbalanciertheit in den verglichenen Attributen festzustellen, die durch die Domäneninkompa-

tibilität und die Notwendigkeit einer Abstraktionsebene verstärkt wird. Problematisch ist allerdings, daß diese Unbalanciertheit in der Form von „Low A/High B“ konstatiert werden muß. Nach der Definition von Ortony würde für diesen Fall keine Metapher wahrgenommen werden, sondern der Vergleich stellt mit dieser Relation ein „anomalous statement“ dar (Ortony führt als Beispiel für diese Art von Statements folgendes Beispiel an: „sleeping pills are like sermons“, Ortony 1979, S. 165). Allerdings könnten ebenso gute Gründe für eine andere Sortierung der Attribute für das Objekt „Baum“ angeführt werden, denn auch Ortony selbst weist auf mögliche Verschiebungen durch Kontexteffekte hin. Ein solcher Effekt könnte hier darin bestehen, daß bei der Verwendung der Baum-Metapher sicherlich nicht auf die biologische Klassifikation („Pflanze“) abgehoben wird, sondern explizit auf die Eigenschaft der Vernetztheit seiner Bestandteile (Äste und Blätter und Wurzeln). Folgt man diesem Einwand, läge keine Unbalanciertheit mehr vor, und somit auch kein „anomalous statement“, sondern eine Analogie. Wie oben bereits ausgeführt, wird für diesen Fall - also eine Analogie-Beziehung - die Konstruktion eines beide Objekte verbindenden übergreifenden Schemas bei der Perzeption des Statements angenommen.

Für die Brauchbarkeit der Baum-Metapher im Zusammenhang mit Hypertexten kann also festgestellt werden, daß vermutlich nicht von einer Tauglichkeit im Sinne einer metapherischen Beziehung ausgegangen werden kann. Gegen die Brauchbarkeit im Falle der Analogiebeziehung spricht technisch nichts, allerdings dürfte es fraglich sein, ob Personen ohne weiteres schematisches Wissen zur Verfügung haben, das einen Vergleich der Strukturen aus dem Quell- und Zielbereich nahelegen würde (zumal die entlastende und verständnisfördernde Funktion der Analogie gerade darauf beruht, daß dieser Zusammenhang ohne Schwierigkeiten hergestellt werden kann).

Was aus der Tabelle aber zusätzlich abzulesen ist, ist eine Antwort auf die Frage nach der Nützlichkeit, denn offensichtlich ist der Überlappungsbereich der Attribute nur gering. Problematischer als dieser Teilbefund ist jedoch die Frage, ob im Falle einer Analogiebeziehung in Bezug auf die von Gentner und Gentner (1983) bzw. Gentner (1983) formulierte structure-mapping-Theorie für die Wirkweise von Analogien im Falle der vorliegenden Analogie tatsächlich von einer Strukturübertragung ausgegangen werden kann. Fokal ist dabei die Frage, ob das Verhältnis der Äste eines Baumes zu den Blättern (und vielleicht noch zum Stamm bzw. den Wurzeln) tatsächlich auf das Verhältnis der Knoten eines Hypertexts zueinander, festgelegt durch die verbindenden Links, übertragen werden kann. Denn wenn tatsächlich die Blätter des Baumes für die Knoten des Hypertextes stehen sol-

len, wird ein zentrales Merkmal von Hypertexten nicht abgebildet, und zwar die netzartige, häufig wenig hierarchische und hochkomplexe Verbindung der Knoten untereinander. Anders gesagt - bei einem Baum verbinden sich die Äste untereinander zu einem „verzweigten“ Netzwerk, an dessen Randgebieten Blätter ausgebildet werden. Diese Struktur ist auf Hypertexte bei näherer Betrachtung nicht gut zu übertragen. Wenn die Metapher allerdings so gemeint sein soll, daß die Knoten des Hypertextes durch Verzweigungen der Äste eines Baumes symbolisiert sein sollen, dann müsste im Vergleich der Attribute das Attribut „hat (Äste)“ im Vergleich sowohl für „besteht aus (Knoten)“ und „besteht aus (Verknüpfungen)“ stehen und es bestünde keine weitere Überlappung der Attribute.

Die abschließend zu klärende Frage, inwiefern von einer Vertrautheit von Personen mit der gewählten Metapher/Analogie ausgegangen werden kann ist zwar theoretisch nicht gut zu klären, allerdings geben die eben ausgeführten theoretischen Einwände gegen die prinzipielle Qualität dieser Metapher Anlaß zu der Befürchtung, daß eine solche Vertrautheit gar nicht in ausreichendem Maß vorhanden sein kann.

Gestärkt wird dieser Einwand vom Befund der Bedingung *WebToc*.

WebToc wurde hinsichtlich der *usability* signifikant besser beurteilt, als der *Hyperbolic-Tree-Browser*. Mit der Hilfe von *WebToc* wurden schließlich auch die besten Ergebnisse für die Suchaufgaben (Retrieval-Fragen) erzielt, was noch bedeutsamer wird wenn man bedenkt, daß ein starker (hypothesen-konservativer) Effekt die Benutzer dieses Tools im Vergleich zu solchen *ohne Hilfe* „behindert“ hat, denn sie hatten durch die Präsentation nur ca. 2/3 des Bildschirms zur Verfügung. Sicherlich lassen sich auch für dieses Tool zahlreiche Mängel hinsichtlich der *usability* finden, so leidet *WebToc* z. B. bei umfangreichen Hierarchien unter dem Problem des Scrollings, d.h. es können nicht wie beim *Hyperbolic-Tree* immer alle Elemente auf einem Bildschirm dargestellt werden. Auch eine beidseitige Synchronisation des Inhaltsverzeichnisses mit dem dargestellten Inhalt wäre wünschenswert (wie es z. B. mit der Visualisierung des Tools „The Brain“, <http://www.thebrain.com> realisiert wurde).

Schließlich war es auch mit dem Tool *WebToc* nicht möglich, User in der Beantwortung komplexer Struktur-Fragen zu unterstützen. Daß es derartige Unterstützungsmöglichkeiten geben müßte, wird anhand der signifikanten Auswirkung des Faktors *spatial ability* für die Beantwortung dieses Fragentyps deutlich (dieser Faktor hat im übrigen auch für Retrieval-Fragen deutliche Auswirkungen). Offensichtlich spielen räumliche Operationen (mit einem

mentalen Modell ?) bei der Bearbeitung derartiger Fragen durchaus eine Rolle - es gelingt mit den hier eingesetzten Tools nur nicht, diese Operationen angemessen zu unterstützen. Um die Ergebnisse im Zusammenhang mit *WebToc* besser verstehen zu können, soll auch hier kurz versucht werden, die Qualität der dem Tool zugrundeliegenden Metapher zu beleuchten.

In diesem Fall lautet das zu untersuchende Statement „Hypertexte sind wie Bücher“ - mit der Einschränkung, daß *WebToc* sozusagen nur das Inhaltsverzeichnis realisiert hat und die jeweiligen Seiten nur in einem separaten Fenster eingeblendet werden.

Tabelle 34: Attributvergleich für die Buchmetapher von *WebToc*

Wichtigkeit	Hypertext	Verknüpfung	Buch
1.	besteht aus (Knoten)		ist (Text)
2.	besteht aus (Verknüpfungen)		besteht aus (Seiten)
3.	ist (Text)		besteht aus (Inhaltsverzeichnis)
4.	ist (vernetzt)		ist (linear)
5.	ist (hierarchisch)		beinhaltet (Indexseiten)
6.	ist (computerbasiert)		
7.	beinhaltet (Indexseiten)		
Domäne	Information		Information

Nach dem Analyseraster von Ortony (1979) handelt es sich bei dem genannten Statement weniger um eine Metapher (es besteht eine relativ geringe Unbalanciertheit in der Wichtigkeit der verglichenen Attribute (zumal „ist (Text)“ in beiden Fällen mit gutem Grund auch an Position eins stehen könnte) und es besteht keine Domäneninkompatibilität), als vielmehr um eine Analogie. So kommen die Prinzipien der structure-mapping Theorie (Gentner & Gentner, 1983; Gentner, 1983) zur Anwendung, und es kann festgestellt werden, daß die Strukturübertragung relativ gut gelingt - vor allem wenn es sich um eher hierarchisch organisierte Hypertexte handelt. Theoretisch bedeutsam ist dabei, daß mit dieser Analogie auch zentrale Merkmale von Hypertexten, nämlich deren Nicht-Linearität, nicht abgebildet werden.

Die Qualität der Analogie, die der Darstellung von *WebToc* zugrundeliegt, ist also auch theoretisch der in der Visualisierung des *Hyperbolic-Tree* überlegen. Zusätzlich muß hier-

bei in Erwägung gezogen werden, daß die Buchmetapher einen weiteren Vorteil auf Ihrer Seite haben dürfte, der an sich wenig mit der Qualität oder der Abbildungsleistung zu tun hat: Es ist eine sehr vertraute Metapher, die die meisten der User vermutlich mehrmals täglich in anderen („Offline-“) Zusammenhängen anwenden werden.

Wenn es nun tatsächlich zutreffen sollte, was Mani und Johnson-Laird (1982) als Folge einer Reihe von Untersuchungen festhielten, nämlich daß „the recovery of a propositional representation is a necessary precursor to the construction of a mental model, and the encoding of a spatial description in the form of a model necessarily requires more processing than merely encoding it propositionally“ (Mani & Johnson-Laird, 1982, S. 185), wird ein wesentlicher Punkt deutlich, an welchem die Idee hinter der Darstellung des *Hyperbolic-Tree* scheitern mußte: Durch die starke Dekontextualisierung wurde den Versuchspersonen die Möglichkeit genommen, eine Verarbeitung der Informationen auf propositionaler Ebene vorzunehmen oder sie wurde zumindest eher behindert. Trifft diese Analyse zu, entsteht allerdings wiederum das Dilemma, daß durch kontextualisierte Darstellung große Informationsmengen praktisch nicht darstellbar sind. Die positiven Ergebnisse von *WebToc* im Zusammenhang mit Retrieval-Fragen deuten dabei in eine Richtung, die vielleicht noch weitaus bessere Ergebnisse erzielen könnte, denn *WebToc* gelingt es zumindest rudimentär, eine stärker komprimierte und gut navigierbare Übersicht mit direkter Einsicht in den Kontext bis auf Seitenebene hinunter zu vereinen. Für die Bearbeitung von Retrieval-Fragen ist dieser Ansatz bereits nach den hier vorliegenden Ergebnissen deutlich im Vorteil gegenüber den anderen Bedingungen. Eine zentrale Frage für die Weiterentwicklung derartiger Visualisierung dürfte allerdings dahingehend lauten, ob mit dieser Art der Darstellung durch entsprechende Optimierungen auch die Bearbeitung komplexer Strukturfragen erleichtert werden könnte. Einige deskriptive Details legen hier eher einen tendenziellen Vorteil der Visualisierung des Hyperbolic-Tree-Browsers - zumindest für erfahrene User - nahe. Eine Konsequenz aus dieser Einsicht könnte für den Hyperbolic-Tree-Browser eine zielgerichtete Weiterentwicklung in Richtung eines Spezialisten-Tools zur Bearbeitung komplexer Fragestellungen sein. Die dabei zu behebenden *usability*-Probleme wiegen allerdings so schwer und die tatsächliche Effektivität des Werkzeuges unter solchen Bedingungen ist noch immer mit so vielen Fragezeichen behaftet, daß der Sinn einer entsprechenden Weiterentwicklung eher sehr fraglich erscheinen muß.

Neben der Überprüfung der genannten Hypothesen zur Informationssuche im World-Wide-Web war ein weiteres Ziel dieser Untersuchung die Beantwortung der Frage, ob es sinnvoll ist, derartige Untersuchungen als Online-Experiment im Internet durchzuführen.

Ein wichtiger Aspekt für die Beantwortung dieser Frage ist die ökologische bzw. externe Validität einer Untersuchung. Diesbezüglich konnte durch das Design als Online-Experiment ein im Vergleich zur Laborsituation relativ natürliches Anwendungsumfeld hergestellt werden, und zwar sowohl hinsichtlich Bearbeitungszeit, Bearbeitungsort, als auch im Hinblick auf vielfältige Faktoren in einem realen Umfeld (z. B. die Computerausstattung). Dennoch ist es möglich, diese Studie von einer Felduntersuchung deutlich abzuheben, da eindeutig isolierte experimentelle Faktoren variiert und z.T. randomisiert zugeteilt werden konnten.

Hinsichtlich der Zusammensetzung und der Größe der Stichprobe ist das Ergebnis ebenfalls zufriedenstellend. Mit einem Laborexperiment hätte im gleichen Zeitraum sicherlich nur ein Bruchteil der hier untersuchten Personen erreicht werden können.

Die guten bis sehr guten psychometrischen Kennwerte der eingesetzten standardisierten Verfahren und die Reproduzierbarkeit der faktoriellen Struktur des Usability-Instruments schließlich geben weitere Hinweise auf die Validität der gewonnenen Daten.

Kritisch ist hingegen festzustellen, daß aufgrund der zeitlichen Beschränkungen durch die Rahmenbedingungen eines Online-Experiments an einigen Stellen Zugeständnisse gemacht werden mußten, die zu schlechten psychometrischen Kennwerten für einige Skalen und Fragedimensionen führten, da nicht genügend Items zur Bearbeitung angeboten werden konnten. Auch die geringeren Möglichkeiten der Kontrolle machten sich insbesondere für die Zeitmessungen bzw. den unkontrollierbaren Einfluß der Bearbeitungszeit (bei der Bearbeitung der Items für *spatial ability*) bemerkbar, allerdings nicht in dem Maße, daß keine statistisch verwertbaren Ergebnisse resultiert hätten.

Insgesamt entstand bezüglich dieser Frage der Eindruck, daß die Vorteile eines Online-Experiments für die vorliegende Fragestellung überwogen.

7.5 Ausblick

Offensichtlich ist es möglich, Tools zur Visualisierung komplexer Informationen im World-Wide-Web mit Online-Experimenten hinsichtlich Performanz, Ergebnisqualität und *usability* zu evaluieren.

Um auch für komplexe Informationsverarbeitungsaufgaben effektive Unterstützung anbieten zu können, sollte unter Umständen zunächst auf einer wesentlich grundlegenderen Ebene überlegt werden, wie derartige Aufgaben gelöst werden, und welche Art der Visualisierung für solche Lösungsprozesse hilfreich sein könnte. Dabei gäbe es zahlreiche Anknüpfungsmöglichkeiten in der Psychologie des komplexen Problemlösens. Die Brauchbarkeit einer eher technisch bedingten Unterscheidung von Relationen der Knoten untereinander im Gegensatz zu Relationen innerhalb von Knoten ist im Zusammenhang mit der Rezeption von Hypertexten prinzipiell, vor allem unter Einbeziehung sprachpsychologischer Befunde zu überprüfen. Vermutlich würden aus derartigen Überlegungen Ansätze zur Visualisierungen gänzlich anderer Art resultieren, vielleicht in der Form selbstorganisierender Karten (SOM, z. B. Zizi & Beaudouin-Lafon, 1994; Lee & Yang, 1999), die versuchen, anhand vorgeschalteter linguistischer Analysen Visualisierungen zu generieren, die über die (künstliche) Grenze von Knoten hinausgehen, oder vielleicht in der Form von adaptiven Hypertexten, die (unter Wahrung der kontextualisierten Darbietungsform) sich dynamisch den jeweiligen Bedürfnissen und Kenntnissen des Users anpassen können (z. B. Kobsa & Stephanidis, 1998; Höök, 1997).

8. Literaturverzeichnis

- Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Anderson, J. R. (1995). *Cognitive Psychology and its Implications*. New York: W.H. Freeman and Company.
- Bandilla, W. (1999). WWW-Umfragen - Eine alternative Datenerhebungstechnik für die empirische Sozialforschung? In B. Batinic, A. Werner, L. Gräf & W. Bandilla (Eds.), *Online Research* (pp. 9-20). Göttingen: Hogrefe.
- Bortz, J. (1993). *Statistik für Sozialwissenschaftler*. Berlin: Springer.
- Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation*. Berlin: Springer.
- Botafogo, R, Rivlin, E. & Shneiderman, B. (1992). Structural Analysis of Hypertexts: Identifying Hierarchies and Useful Metrics. *ACM Transactions on Information Systems*, 10 (2), 142-180.
- Bringelson, L. S. & Eberts, R. (1990). The Effect of Cognitive Workload, Job Enrichment Environment, and Spatial Ability on Goal Structures and Strategies. *Proceedings of the Human Factors Society 34th Annual Meeting*, 2, 1514-1518.
- Card, S., Moran, T. P. & Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale: Lawrence Erlbaum Associates.
- Catledge, L. D. & Pitkow, J. E. (1995). Characterizing Browsing Strategies in the World Wide Web. [WWW-Document] Available at: <http://www.igd.fhg.de/www/www95/proceedings/papers/80/userpatterns/UserPatterns.Paper4.formatted.html> [Stand 18.8.99].
- Catledge, L. & Pitkow, J. (1995). Characterizing browsing behaviors on the World Wide Web. *Computer Networks and ISDN Systems*, 27 (6).
- Chen, C. & Czerwinski, M. (1997). Spatial ability and visual navigation: an empirical study. *New Review of Hypermedia and Multimedia*, 3, 67-90.
- Chen, C., & Rada, R. (1996). Interacting with hypertext: A meta-analysis of experimental studies. *Human Computer Interaction*, 11 (2), 125-156.
- Chun, D. & Plass, J. L. (1997). Research on Text Comprehension in Multimedia Environments. In *Language, Learning and Technology*. [WWW-Document] Online Available: <http://www.unm.edu/~jan/.LLT/Chun&Plass.html>.
- Cockburn, A., & Jones, S. (1996). Which way now? Analysing and easing inadequacies in WWW navigation. *International Journal of Human Computer Studies*, 45 (1), 105-129.

- Conklin, J. (1987). Hypertext: An Introduction and Survey. *IEEE Computer*, 17-41.
- Cooley, R. Bamshad, M. & Srivastava, J. (1999). Data Preparation for Mining World Wide Web Browsing Patterns. *Knowledge and Information Systems*, 1, 5-32.
- de Bra, P., Houben, G. J., De Vocht, Y. & Kornatzky, Y. (1994). Retrieval of Hypertext Structures. *Stinfony-94 Conference Proceedings*, 103-118.
- de Bra, P. & Houben, G.-Jan. (1994). Hypertext Metrics Revisited: Navigational Metrics for Static and Adaptive Link Structures. [WWW-Dokument]. Online Available: <http://www.wis.win.tue.nl/~houben/pub/>. Eindhoven:
- de Vocht, J. (1994). Experiments for the Characterization of Hypertext-Structures. [WWW-Dokument] Online Available: <http://www.wis.win.tue.nl/~debra/joep/>.
- Dillon, A. & Vaughan, M. (1997). "It's the journey and the destination": Shape and the emergent property of genre in evaluating digital documents. *New Review of Multimedia and Hypermedia*, 3, 91-106.
- Dutke, S. (1994). *Mentale Modelle: Konstrukte des Wissens und Verstehens*. Göttingen, Stuttgart: Verlag für Angewandte Psychologie.
- Fowler, R. & Bradley, A. W. & Fowler, W. (1992). Information Navigator: An information system using associative networks for display and retrieval. [WWW-Dokument] Online Available: http://bahia.cs.panam.edu/info_vis/inf_nav/info_nav_tr_92.html [Stand: 18.8.1999].
- Furnas, G. W. (1986). Generalized fisheye views. *Proceedings of ACM CHI 86 Conference on Human Factors in Computing Systems*, 16-23
- Gentner, D. (1983). Structure Mapping: A Theoretical Framework for Analogy. *Cognitive Science*, 7, 155-170.
- Gentner, D. & Gentner, R. G. (1983). Flowing Waters or Teeming Crowds: Mental Models of Electricity. In D. Gentner & A. L. Stevens (Eds.), *Mental Models* (pp. 99-127). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gentner, D. & Stevens, A. L. (1983). *Mental Models*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Gerdes, H. (1997a). Hypertext. In B., Batinic (Hrsg.), *Internet für Psychologen* (S. 137-154). Göttingen: Hogrefe.
- Gerdes, H. (1997b). *Lernen mit Text und Hypertext*. Lengerich: Pabst.

- GFK. (2000). Online-Monitor - 5. Erhebungswelle. [WWW-Document] Online Available: <http://www.gfk.de>.
- Green, D. W. (1996). How we solve Problems. In D. W. Green (Eds.), *Cognitive Science* (pp. 310-340). Cambridge: Blackwell Publishers Ltd.
- Groeben, N. (1982). *Leserpsychologie: Textverständnis - Textverständlichkeit*. Münster: Aschendorff.
- Groeben, N. (1986). *Handeln, Tun, Verhalten als Einheiten einer verstehend-erklärenden Psychologie*. Tübingen: Francke.
- Groeben, N., Wahl, D., Schlee, J. & Scheele, B. (1988). *Forschungsprogramm Subjektive Theorien*. Tübingen: Francke.
- Hussy, W. (1993). *Denken und Problemlösen*. Stuttgart: Kohlhammer.
- Hutchins, E. (1983). Understanding Micronesian Navigation. In D. Gentner & A. L. Stevens (Eds.), *Mental Models* (pp. 191-225). Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Jaeger, A. O. & Althoff, K. (1983). *Wilde-Intelligenz-Test - Handanweisung*. Göttingen: Deutsche Gesellschaft fuer Personalwesen e.V.
- Jih, H. J., & Reeves, T. C. (1992). Mental models: A research focus for interactive learning systems. *Educational Technology Research and Development*, 40 (3), 39-53.
- Johnson, S. (1999). *Interface Culture*. Stuttgart: Klett-Cotta.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inferences, and consciousness*. Cambridge: Cambridge University Press.
- Johnson-Laird, P. N., Girotto, V. & Legrenzi, P. (1998). Mental models: a gentle guide for outsiders. [WWW-Document] Available URL: <http://www.si.umich.edu/ICOS/gentleintro.html>.
- Jonassen, D. H. & Wang, S. (1992). Acquiring Structural Knowledge from Semantically Structured Hypertext. *Proceedings of Selected Research and Development Presentations at the Convention of the Association for Educational Communications and Technology*, 1-13.
- Kelloway, E. K. (1998). *Using Lisrel for Structural Equation Modeling*. Thousand Oaks: SAGE Publications.
- Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, N.J.: Erlbaum.

- Kobsa, A. & Stephanidis, C. (1998). Adaptable and adaptive information access for all users, including disabled and elderly people. *Proceedings of the 2nd Workshop on Adaptive Hypertext and Hypermedia*. [WWW-Document] Available URL: <http://www.wis.win.tue.nl/~ah98/Kobsa.html>.
- Koenemann, J. & Stolze, M. (1999). User Interfaces for Electronic Product Catalogs. *ACM SIGCHI Bulletin*, 31 (4), 47-48.
- Krantz, J. H. & Dalal, R. (2000). Validity of Web-Based Psychological Research. In M. H. Birnbaum (Eds.), *Psychological Experiments on the Internet* (pp. 35-57). San Diego: Academic Press.
- Kueppers, S. & Schilingno, M. (1999). Getting Our Act Together: Human and Technological Factors in Establishing an Online Knowledge Base. *Proceedings of the 27th Annual ACM SIGUCCS User Services Conference*, 135-149.
- Kuhlen, R. (1991). *Hypertext. Ein nicht-lineares Medium zwischen Buch und Wissenschaft*. Berlin: Springer.
- Lamping, J. & Rao, R. (1996). Visualizing large trees using the hyperbolic browser. *Proceedings of the CHI '96 conference companion on Human factors in computing systems*, 388-389.
- Lamping, J., Rao, R., & Pirolli, P. (1995). A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. *Conference proceedings on Human factors in computing systems*, 401-408.
- Larkin, J. H. (1983). The Role of Problem Representation in Physics. In D. Gentner & A. Stevens (Eds.), *Mental Models* (pp.75-97). Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Lee, C. & Yang, H. (1999). A Web Text Mining Approach Based on Self Organizing Map. *Proceedings of the second international workshop on on Web information and data management*, 59-62.
- Lewis, R. J. (1995). IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *International Journal of Human-Computer Interaction*, 7, 57-78.
- Lienert, G. A. & Ratz, U. (1998). *Testaufbau und Testanalyse*. Weinheim: Psychologie Verlags Union.

- Litz, H. P. (2000). *Multivariate statistische Methoden*. München: R. Oldenbourg.
- Mani, K. & Johnson-Laird, P. N. (1982). The mental representations of spatial descriptions. *Memory & Cognition*, 10 (2), 181-187.
- McDonald, S., & Stevenson, R. J. (1996). Disorientation in hypertext: The effects of three text structures on navigation performance. *Applied Ergonomics*, 27 (1), 61-68.
- McDonald, S., & Stevenson, R. J. (1998). Effects of text structure and prior knowledge of the learner on navigation in hypertext. *Human Factors*, 40 (1), 18-27.
- McDonald, S., & Stevenson, R. J. (1998). Navigation in hyperspace: An evaluation of the effects of navigational tools and subject matter expertise on browsing and information retrieval in hypertext. *Interacting with Computers*, 10 (2), 129-142.
- McEneaney, J. E. (1999). Visualizing and assessing navigation in hypertext. *Proceedings of the tenth ACM Conference on Hypertext and hypermedia*, 61-70.
- Mukherjea, S. & Hara, Y. (1997). Focus+Context Views of World-Wide Web Nodes. *Proceedings of the Eighth ACM Conference on Hypertext*, 187-196.
- Musch, J. (1997). Die Geschichte des Netzes: ein historischer Abriß. In Batinic, B. (Hrsg.), *Internet für Psychologen* (S. 27-48). Göttingen: Hogrefe.
- Nation, D. (1998). WebTOC: A Tool to Visualize and Quantify Web Sites using a Hierarchical Table of Contents Browser. *Proceedings of the 1998 ACM CHI-Conference on Human Computer Interaction*, 185-186.
- Nation, D. A., Plaisant, C., Marchionini, G. & Komlodi, A. (1997). Visualizing websites using a hierarchical table of contents browser: WebTOC. *Proceedings of 3rd Conference on Human Factors and the Web*.
- Nation, D., Plaisant, C., Marchionini, G. & Komlodi, A. (1997). Visualizing websites using a hierarchical table of contents browser: WebTOC. *UM Technical Report CS-TR-3791*.
- Naumann, J., Richter, T. & Noller, S. (2000). On the equivalence of the English and German online-version of the questionnaire for the content-differentiated assessment of attitudes toward the computer (QCAAC). *CiP2000 Conference Proceedings* (Bd. 1, S. 62-63). York: [WWW-Document] Available: http://cti-psy.york.ac.uk/cip2000-online/papers/fra_paper47.htm.
- Newell, A. (1990). *Unified Theories of Cognition*. Cambridge: Harvard University Press.
- Newell, A. & Simon, H. (1987). *Computer Science as Empirical Inquiry: Symbols and*

- Search. In J. Haugeland (Eds.), *Mind Design* (pp.35-66). Cambridge: MIT Press.
- Nielsen, J. (1990). The art of navigating through hypertext. *Communications of the ACM*, 33 (3), 296-310.
- Nielsen, J. (1993). *Usability Engineering*. San Diego: Academic Press.
- Noller, S. (2000). *Log-Pat. Ein web-basiertes Tool zur Analyse von Logfile-Daten*. Unveröffentl. Software-Programm, Universität zu Köln
- Norman, D. A. (1983). Some observations on mental models. In D. Gentner & A. L. Stevens (Eds.), *Mental Models* (pp. 7-14). Hillsdale, NJ: Erlbaum.
- Norman, K. (1991). *The Psychology of Menu Selection: Designing Cognitive Control at the Human/Computer Interface*. New York: Ablex Publishing.
- Ohler, P., & Nieding, G. (1997). Kognitive Modellierung der Textverarbeitung und der Informationssuche im World Wide Web. In B. Batinic (Hrsg.), *Internet für Psychologen* (S. 161-180). Göttingen:
- Ortony, A. (1979). Beyond Literal Similarity. *Psychological Review*, 86 (3), 161-180.
- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.
- Parunak, H. V. D. (1989). Hypermedia topologies and user navigation. In *Proceedings of the second annual ACM conference on Hypertext* (S. 43-50). Pittsburgh: ACM Press.
- Pirolli, P. & Card, S. K. (in press). (1999). Information foraging. *Psychological review*.
- Pirolli, P., Card, S. K. & van der Wege, M. M. (in press). The Effect of Information Scent on Searching Information Visualizations of Large Tree Structures. Conference Proceedings of the 5th Working Conference on Advanced Visual Interfaces.
- Pirolli, P, Pitkow, J & Rao, R. (1996). Silk from a sow' ear: Extracting usable structures from the Web. Proceedings of 1996 Conference on Human Factors in Computing Systems.
- Pitkow, J. (1997). In search of reliable usage data on the WWW. *Proceedings of the Sixth International World Wide Web Conference*, 451-463.
- Press, L. (1997). Technology in Bloom: Implications for the next 50 years. *Communications of the ACM*, 40 (2), 11-17.
- Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, 80, 1-24.

- Rao, R. & Card, S. K. (1994). Table Lens: Merging Graphical and Symbolic Representations in an Interactive Focus+Context Visualization for Tabular Information. *Proceedings of ACM CHI'94 Conference on Human Factors in Computing Systems, 1*, 318-322.
- Rauterberg, M. (1992). An empirical comparison of menu-selection (CUI) and desktop (GUI) computer programs carried out by beginners and experts. *Behaviour and Information Technology, 11* (4), 227-236.
- Rauterberg, M. (1992). An empirical comparison of menu-selection (CUI) and desktop (GUI) computer programs carried out by beginners and experts. *Behaviour and Information Technology, ehaviour and Information Technology, 11* (4), 227-236.
- Rauterberg, M. (1995). About faults, errors, and other dangerous things. In H. Stassen & P. Wieringa (Eds.) *Proceedings of 'XIV European Annual Conference on Human Decision Making and Manual Control*, 1-7
- Rauterberg, M. (1995). From novice to expert decision behaviour: A qualitative modelling approach with Petri nets. In Y. Anzai, K. Ogawa & H. Mori (Eds.), *Symbiosis of Human and Artifact: Human and Social Aspects of Human-Computer Interaction* (pp. 449-454)
- Rauterberg, M., & Fjeld, M. (1998). Task analysis in human-computer interaction - supporting action regulation theory by simulation. *Zeitschrift fuer Arbeitswissenschaft, (3)*, 152-161.
- Recker, M. M., & Pirolli, P. (1995). Modeling individual differences in students' learning strategies. *Journal of the Learning Sciences, 4*, 1-38.
- Reips, U. (2000). The Web Experiment Method: Advantages, Disadvantages, and Solutions. In M. H. Birnbaum (Eds.), *Psychological Experiments on the Internet* (pp. 89-120). San Diego: Academic Press.
- Richter, T., Naumann, J. & Noller, S. (1999). Computer Literacy und computerbezogene Einstellungen: Zur Vergleichbarkeit von Online- und Paper-Pencil-Erhebungen. In U.-D. Reips, B. Batinic, W. Bandilla, M. Bosnjak, L. Gräf, K. Moser & A. Werner (Hrsg.), *Current internet science - trends techniques, results*. Zürich: Online Press [WWW-Dokument] Available URL: <http://dgof.de/tband99/>.
- Rouet, J. F. (1992). Cognitive processing of hyperdocuments: When does non-linearity help? In D. Lucarella, J. Nanard, M. Nanard & P. Paolini (Eds.), *Proceedings of the 4th ACM Conference on Hypertext* (pp. 131-140). New York: Academic Press.

- Schulmeister, R. (1997). *Grundlagen hypermedialer Lernsysteme*. München: Oldenbourg.
- Shneiderman, B. (1998). *Designing the User Interface*. Reading: Addison-Wesley.
- Shneiderman, B. (2000). Universal Usability. *Communications of the ACM*, 43 (5), 84-92.
- Smelcer, J. B. (1994). Do geographic information systems improve decision making?. Washington: [WWW-Document] Online Available: <http://lattanze.loyola.edu/frames/research/wp1094.023.html>.
- Smith, P. A., Newman, I. A. & Parks, L. M. (1997). Virtual hierarchies and virtual networks: some lessons from hypermedia usability research applied to the World Wide Web. *International Journal of Human-Computer Studies*. Academic Press. [WWW-Document] Available URL: <http://www.hbuk.co.uk/ap/ijhcs/webusability/smith/smith.html> [Stand: 10.8.2000].
- Stachowiak, H. (1973). *Allgemeine Modelltheorie*. Wien: Springer.
- Sullivan, T. C. (1997). Reading Reader Reaction: A Proposal for Inferential Analysis of Web Server Log Files. Online in Internet: <http://www.uswest.com/web-conference/proceedings/index.html> [Stand: 10.6.1999].
- Tasche, K. (1999). "Uses and Gratifications" und Stimmungsregulationstheorie: tragfähige Konzepte zur Erklärung der Nutzung von Online-Medien? In W. Wirth & W. Schweiger (Hrsg.), *Selektion im Internet* (S. 75-96). Opladen: Westdeutscher Verlag.
- Taylor, I. (1990). *Psycholinguistics. Learning and using language*. Englewood Cliffs, N.J.: Prentice Hall.
- Tergan, S. O. (1997). Misleading theoretical assumptions in hypertext / hypermedia research. *Journal of Educational Multimedia and Hypermedia* 6(3/4), 257-283.
- Thüring, M., Hannemann, J. & Haake, J. M. (1995). Hypermedia and cognition: designing for comprehension. *Communications of the ACM*, 38 (8), 57-66.
- Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University Press.
- Verjeij, J., Stoutjesdijk, E. & Beishuizen, J. (1996). Search and Study Strategies in Hypertext. *Computers in Human Behaviour*, 12, 1-15.
- Vicente, K. J. & Williges, R. C. (1988). Accomodating individual differences in searching a hierarchical file system. *International Journal of Man-Machine Studies* 29, 647-668.
- Vinson, N. G. (1999). Design Guidelines for Landmarks to Support Navigation in Virtual Environments. *Proceedings of ACM CHI 99 Conference on Human Factors in Comput-*

ing Systems, 278-285.

- Walker, N., Meyer, D. E. & Smelcer, J. B. (1993). Spatial and Temporal Characteristics of Rapid CursorPositioning Movements with Electromechanical Mice in Human-Computer Interaction. *Human Factors*, 35 (3), 431-458.
- Weber, G. & Specht, M. (1997). *User Modeling and Adaptive Navigation Support in WWW-based Tutoring Systems*. Online in Internet: <http://www.psychologie.uni-trier.de:8000/projects/ELM/Papers/UM97-WEBER.html> [Stand: 18.8.1999].
- Whalley, P. (1990). Models of hypertext structure and learning. In D. H. Jonassen & H. Mandl (Eds.), *Designing hypermedia for learning* (pp. 61-67). Berlin: Springer.
- Wirth, W. & Schweiger, W. (1999). *Selektion im Internet*. Opladen: Westdeutscher Verlag.
- Wittgenstein, L. (1999). *Philosophische Untersuchungen*. Frankfurt a. M.: Suhrkamp.
- Wright, P. & Lickorish, A. (1990). An empirical comparison of two navigation systems for two hypertexts. In Ray McAleese & Catherin Green (Eds.), *Hypertext: state of the art* (pp. 84-93). Oxford: Intellect Books Ltd.
- Young, R. M. (1983). Surrogates and Mappings: Two Kinds of Conceptual Models for Interactive Devices. In D. Gentner & A. L. Stevens (Eds.), *Mental Models* (pp. 35-52). Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Zizi, M. & Beaudoin-Lafon, M. (1994). Accessing hyperdocuments through interactive dynamic maps. *Proceedings of the 1994 ACM European conference on Hypermedia technology*, 126-135.

9. Abbildungsverzeichnis

12	Prozessmodell der Mensch-Computer-Interaktion nach Norman (1991)	18
13	Inspiration für die Visualisierung des Hyperbolic-Tree-Browsers: Graphik von M.C. Escher	54
14	Startseite des Experiments in der Faktorstufe „ohne Hilfe“	67
15	Startseite des Experiments in der Faktorstufe WebToc	68
16	Startseite des Experiments in der Faktorstufe Hyperbolic-Tree	69
17	2-faktorielles Modell	87
18	richtige Antworten in Abhängigkeit der Navigationshilfe.	89
19	Bearbeitungsdauer	90
20	Verteilung der Werte für spatial ability	91
21	Bearbeitungsdauer für den Untertest spatial ability	92
22	Zusammenhang von spatial ability und der Bearbeitungsdauer für den entsprechenden Test	93
23	richtige Antworten - Retrieval	95
24	Bearbeitungsdauer - Retrieval	96
25	richtige Antworten - Struktur	98
26	Bearbeitungsdauer - Struktur	98
27	richtige Antworten in Abhängigkeit von Fragentyp und Navigationsbedingung	99
28	Bearbeitungsdauer in Abhängigkeit von Fragentyp und Navigationsbedingung	101
29	Einfaktorielle Lösung in der konfirmatorischen Faktorenanalyse	102
30	Meßmodell der 3-faktoriellen Struktur	103
31	usability in Abhängigkeit der Navigationsbedingung	105
32	„System-Usefulness“ in Abhängigkeit der Navigationsbedingung	106
33	„Information-Quality“ in Abhängigkeit der Navigationsbedingung	107
34	„Interface-Quality“ in Abhängigkeit der Navigationsbedingung	107
35	Compactness in Abhängigkeit der Navigationsbedingung	110
36	Stratum in Abhängigkeit der Navigationsbedingung	110
37	Anzahl besuchter Knoten in Abhängigkeit der Navigationsbedingung	111
38	Anzahl richtiger Retrieval-Fragen in Abhängigkeit von Interneterfahrung und Navigationsbedingung	118
39	Bearbeitungsdauer für Retrieval-Fragen in Abhängigkeit von Interneterfahrung und Navigationsbedingung	119

40	Anzahl richtiger Struktur-Fragen in Abhängigkeit von Interneterfahrung und Navigationsbedingung	120
41	Bearbeitungsdauer für Struktur-Fragen in Abhängigkeit der Interneterfahrung und Navigationsbedingung	120
42	„usability (gesamt)“ in Abhängigkeit der Interneterfahrung und Navigationsbedingung	121
43	„System-Usefulness“ in Abhängigkeit der Interneterfahrung und Navigationsbedingung	122
44	„Information Quality“ in Abhängigkeit der Interneterfahrung und Navigationsbedingung	123
45	„Interface Quality“ in Abhängigkeit der Interneterfahrung und Navigationsbedingung	124
46	Frage: „Es fiel mir leicht die Fragen zu beantworten“ in Abhängigkeit der Navigationsbedingung (hohe Werte = hohe Zustimmung)	125
47	Frage: „Ich hatte manchmal das Gefühl, nicht mehr zu wissen, wo ich bin“ (Orientierung) in Abhängigkeit der Navigationsbedingung	126
48	Frage: „Das Experiment hat mir gut gefallen“ in Abhängigkeit der Navigationsbedingung	126
49	Bearbeitungsdauer der Einführung in Abhängigkeit der Navigationsbedingung	129
50	Bearbeitungsdauer über den Verlauf der Erhebung bei Retrieval-Fragen	131
51	Bearbeitungsdauer über den Verlauf der Erhebung bei Struktur-Fragen	132
52	Lerneffekte über alle sechs Meßzeitpunkte (=Fragen)	132
53	Dropout über den Verlauf der Untersuchung	133
54	Dropout-Fälle (n=178) zum ersten Meßzeitpunkt in Abhängigkeit der Navigationsbedingung	134
55	Dropout-Fälle (n=178) in Abhängigkeit der Bildschirmauflösung	135
56	Dropout in Abhängigkeit des Betriebssystems (Microsoft-Systeme vs. Andere)	136
57	Dropout-Fälle (n=22) zum Zeitpunkt T2 in Abhängigkeit von der Navigationsbedingung	138
58	Dropout-Fälle (n=22) zum Zeitpunkt T2 in Abhängigkeit von der Bildschirmauflösung	138
59	Dropout über alle Meßzeitpunkte in Abhängigkeit der Navigationsbedingung	139

10. Tabellenverzeichnis

1	Versuchsplan	75
2	Vergleich der Stichprobenkennwerte mit anderen Studien	77
3	Test der Varianzhomogenität	83
4	Test der Normalverteilungsannahme: Bearbeitungsdauer/Anzahl richtiger Fragen	84
5	Test der Normalverteilungsannahme für usability-Faktoren	85
6	Itemanalysen	86
7	Indizes der Modellgüte für die konkurrierenden Meßmodelle	87
8	Deskriptive Statistiken	94
9	Multivariate Vergleiche - Retrieval	94
10	Deskriptive Statistiken	97
11	Multivariate Vergleiche - Gesamt	97
12	Deskriptive Statistiken richtige Antworten in Abhängigkeit der Navigationsbedingung und des Fragentyps	99
13	Deskriptive Statistiken für die AV Bearbeitungsdauer in Abhängigkeit der	
	Navigationsbedingung und des Fragentyps	100
14	Indizes der Modellgüte für das 1-faktorielle Meßmodell	102
15	Indizes der Modellgüte für das 3-faktorielle Meßmodell	104
16	usability-Dimensionen in Abhängigkeit der Navigationsbedingung	105
17	Deskriptive Statistiken zu Compactness und Stratum	108
18	Univariate Varianzanalysen in Abhängigkeit der Navigationsbedingung	109
19	multiple Mittelwertsvergleiche (Post-Hoc-Analyse)	109
20	Deskriptive Statistiken zur Anzahl der besuchten Knoten	111
21	Deskriptive Statistiken für Seitenaufrufe „Index“ und „Suchfunktion“	112
22	Univariate Varianzanalyse über Seitenaufrufe	112
23	Korrelationen	114
24	Aufteilung der Gruppen für „Interneterfahrung“	117
25	Varianzanalysen über zusätzliche Fragen	125
26	Korrelationen	127
27	Deskriptive Statistiken: Bearbeitungszeit für die Einführung in Abhängigkeit der Navigationsbedingung	128
28	Post-Hoc-Analyse: multiple Mittelwertsvergleiche	129
29	Deskriptive Statistiken über die drei Meßzeitpunkte für Retrieval-Fragen	130

30	Deskriptive Statistiken über die drei Meßzeitpunkte für Struktur-Fragen	131
31	kombinierter Chi-Quadrat-Test Auflösung*Betriebssystem*Abbruch	137
32	kombinierter Chi-Quadrat-Test Auflösung*Betriebssystem*Abbruch	140
33	Attributvergleich für die Baummetapher	153
34	Attributvergleich für die Buchmetapher von WebToc	156

11. Anhang