

Wissenschaftliches Problemlösen lernen:
Dyaden und Einzelpersonen experimentieren
im simulierten Labor

Dissertationsschrift

vorgelegt am 13. April 2005

der Fakultät für Psychologie der Universität Basel

von
Dipl.-Psych. Cornelia Kneser
aus Stuttgart, Deutschland.

Für Philipp

Genehmigt von der Fakultät für Psychologie auf Antrag der Promotionskommission

Prof. Dr. K. Opwis
1. Gutachter

Prof. Dr. M. Boos
2. Gutachterin

Prof. Dr. M. Wänke
Vorsitzende

Prof. Dr. A. Grob
Mitglied

Basel, den 30. Juni 2005

Prof. Dr. K. Opwis,
Dekan der Fakultät für Psychologie
Universität Basel

Veröffentlicht im Juli 2005 auf dem Dokumentenserver *e-Diss@UNI BASEL* der
Universitätsbibliothek Basel.

Bitte zitieren als:

Kneser, C. (2005). *Wissenschaftliches Problemlösen lernen: Dyaden und Einzel-
personen experimentieren im simulierten Labor*. Basel: Universitätsbibliothek.
Verfügbar unter <http://www.unibas.ch/diss/2005/>

Inhaltsverzeichnis

Verzeichnis der Tabellen	ix
Verzeichnis der Abbildungen	xi
Verzeichnis der Transkripte	xiii
Dank	xvii
Zusammenfassung	xix
1 Einleitung	1
2 Wissenschaftliches Problemlösen	7
2.1 Logisches Schlußfolgern	9
2.2 Konzeptbildung	11
2.3 Kognitionswissenschaftliche Analysen	14
2.4 Empirische Untersuchungen	17
2.4.1 Beobachtungen in Forschungslabors	17
2.4.2 Experimentelle Untersuchungen	21
3 Erlernen von Fertigkeiten für das wissenschaftliche Problemlösen	27
3.1 Novizen beim wissenschaftlichen Problemlösen	27
3.2 Entwicklung wissenschaftlichen Denkens	32
3.3 Entdeckendes Lernen	37
3.4 Situiertes Lernen	40
4 Kooperatives Lernen	45
4.1 Pädagogisch-psychologische Theorien	46
4.2 Einflußfaktoren auf das kooperative Lernen	48
4.2.1 Zusammensetzung von Gruppen	49
4.2.2 Gruppenaufgaben	52
4.2.3 Leistungsvergleich von Gruppen und Individuen	54
4.3 Prozesse kooperativen Lernens und ihre didaktische Unterstützung	57
5 Lernen am Computer	65
5.1 Simulationen als offene Lernumgebungen	66
5.2 Kognitive Werkzeuge für das Lernen am Computer	78
6 Hypothesen	85
7 <i>virtue</i>: das simulierte psychologische Labor	91
7.1 Gegenstandsbereich visuelle Suche	91
7.2 Gestaltung der Lernumgebung <i>virtue</i>	96
7.2.1 Hypothesenscratchpad	97
7.2.2 Versuchsplanung	99
7.2.3 Ergebnisse	100
7.3 Technische Implementierung	102
7.4 Wirksamkeit des Hypothesenscratchpads	104

8	8 Methode	107
8.1	Versuchsplan	107
8.2	Material	108
8.2.1	Paralleltests zum bereichsspezifischen und bereichsübergreifenden Lernen beim wissenschaftlichen Problemlösen	108
8.2.2	Leistungsdiagnostik und Fragebögen	110
8.3	Versuchspersonen	114
8.4	Versuchsablauf	115
9	9 Ergebnisse I: Quantitative Daten aus Tests und Logfiles	121
9.1	Vor- und Nachtest	121
9.1.1	Reliabilitätsanalyse	121
9.1.2	Lernerfolg	123
9.2	Logfileanalyse	129
9.2.1	Allgemeines Vorgehen im Hypothesen- und Experimenterraum	129
9.2.2	Formulierung der Hypothesen	130
9.2.3	Systematik der Experimentalreihe	132
9.2.4	Abdeckung des Hypothesen- und Experimenterraums	134
9.2.5	Ergebnisinterpretation und Bewertung von Hypothesen ..	138
9.3	Soziodemographische Daten, Fragebögen und Intelligenzdiagnostik	139
9.3.1	Vergleich der Experimentalgruppen	140
9.3.2	Zusammenhänge von Diagnostik und Lernerfolg	141
9.3.3	Frauen und Männer	141
9.3.4	Nachbefragung der Versuchspersonen	142
10	10 Ergebnisse II: Qualitative Analyse ausgewählter Einzelfälle	145
10.1	Ella: Systematische Exploration des Experimenterraums	148
10.2	Emelie: Lernen trotz aller Fehler	157
10.3	Dana und Delia: Wenn eine gut erklären kann... ..	169
10.4	Dido und Dora: Gemeinsam erarbeitetes Wissen	185
10.5	Metaaussagen	207
10.6	Vergleich von Dyaden und Einzelpersonen	211
11	11 Diskussion	217
11.1	Bewertung der Hypothesen	218
11.2	Testdiagnostische Kennwerte	229
11.3	Einordnung der Studie in den breiteren Forschungskontext	230
11.4	Ausblick	232
12	12 Literatur	237
13	13 Anhang	257
A	Eingangsfragebogen	259
B	Instruktion visuelle Suche	267
C	Vortest	270
D	Nachtest	280
E	Fragebogen für Dyaden nach dem Experiment	291

Verzeichnis der Tabellen

Tabelle 1	Beschreibungsschema für Lernprozesse beim explorativen Lernen mit Simulationen nach Goodyear, Njoo, Hijne und van Berkum (1991).	71
Tabelle 2	Versuchsplan	107
Tabelle 3	Versuchsablauf mit Angaben zur Dauer	117
Tabelle 4	Reliabilitätsanalyse der Untertests des Vortests	122
Tabelle 5	Zusätzliche bereichsspezifische Items im Nachtest	127
Tabelle 6	Kennwerte zu den Versuchsreihen in <i>virtue</i>	129
Tabelle 7	Beispiele für sauber und unsauber formulierte Hypothesen von Einzelperson E114.	131
Tabelle 8	Qualität der formulierten Hypothesen in absoluten Häufigkeiten	131
Tabelle 9	Konstanthaltung des Faktors Suchtyp	133
Tabelle 10	Kennwerte für die systematische Entwicklung einer Versuchsreihe in absoluten Häufigkeiten	134
Tabelle 11	Bewertung von Hypothesen: absolute Häufigkeiten	139
Tabelle 12	Diagnostische Kennwerte der Experimentalgruppen im Vergleich	140
Tabelle 13	Ergebnisse der für die qualitative Analyse ausgewählten Versuchspersonen	146
Tabelle 14	Experimentelle Zyklen von Einzelperson 1: Ella	149
Tabelle 15	Experimentelle Zyklen von Einzelperson 2: Emelie	158
Tabelle 16	Experimentelle Zyklen von Dyade 1: Dana und Delia	171
Tabelle 17	Experimentelle Zyklen von Dyade 2: Dido und Dora	186
Tabelle 18	Absolute und relative Häufigkeiten der Metaaussagen	208
Tabelle 19	Zusammenhänge, in denen Metaaussagen auftreten	209

Verzeichnis der Abbildungen

Abbildung 1	Typisierung von Gruppenaufgaben nach McGrath (1984).	53
Abbildung 2	Klassifikation von Lernzielen beim Lernen mit Simulationen nach van Berkum, Hijne, de Jong, van Joolingen und Njoo (1991).	69
Abbildung 3	Schematische Darstellung der Merkmals-Integrations-Theorie von Treisman (nach Posner & Raichle, 1996).	92
Abbildung 4	Vorhersagen der Merkmals-Integrations-Theorie.	94
Abbildung 5	Visuelle Anordnung zur Merkmalsuche.	94
Abbildung 6	Visuelle Anordnung zur Verbindungssuche.	94
Abbildung 7	Hauptbildschirm von <i>virtue</i>	96
Abbildung 8	Hypothesenarchiv von <i>virtue</i>	97
Abbildung 9	Hypothesenscratchpad von <i>virtue</i>	98
Abbildung 10	Definition des Versuchsplans in <i>virtue</i>	99
Abbildung 11	Festlegung des Stichprobenumfangs in <i>virtue</i>	100
Abbildung 12	Ergebnisfenster mit Hypothesenbewertung in <i>virtue</i>	102
Abbildung 13	Versuchsanordnung.	116
Abbildung 14	Standbild aus einer Videoaufzeichnung.	116
Abbildung 15	Ergebnisse der Untertests Theorie, Empirie und Methodologie.	125
Abbildung 16	Ein- und zweifaktorielle Hypothesen nach Faktoren.	135
Abbildung 17	Ein- und zweifaktorielle Experimente nach Faktoren.	135
Abbildung 18	Korrektheit ein- und zweifaktorieller Vorhersagen nach Faktoren.	137

Verzeichnis der Transkripte

Transkript 1	Ella wundert sich über das Ergebnis ihres ersten Experiments.	148
Transkript 2	Ella erinnert sich an die Auswirkungen von Verbindungs- und Merkmalsuche.	150
Transkript 3	Ella interpretiert die statistische Interaktion des Experiments aus Zyklus 3.	151
Transkript 4	Ella vergleicht die Ergebnisse der Experimente aus den Zyklen 6 und 2.	153
Transkript 5	Ella stellt Überlegungen zur experimentellen Ökonomie an.	154
Transkript 6	Ella interpretiert Haupteffekte und Interaktion des Experiments in Zyklus 9.	154
Transkript 7	Ella überprüft ihr bisher erworbenes Wissen und plant ihr weiteres Vorgehen.	155
Transkript 8	Ella rekapituliert ihre bisherigen Ergebnisse.	156
Transkript 9	Emelie reflektiert die Aufgabenstellung.	159
Transkript 10	Emelie interpretiert die Ergebnisse aus Zyklus 1.	160
Transkript 11	Emelie interpretiert die Ergebnisse aus Zyklus 2.	162
Transkript 12	Emelie plant ihr weiteres Vorgehen.	163
Transkript 13	Emelie plant ihre nächsten Experimente.	164
Transkript 14	Emelie vergleicht die Ergebnisse aus den Zyklen 3 und 4.	165
Transkript 15	Emelie rekapituliert, inwiefern sie die Theorie überprüft hat.	165
Transkript 16	Emelie verwechselt die Faktoren Farbähnlichkeit und Farbkombination.	167
Transkript 17	Emelie interpretiert die Interaktion zwischen Größe der Anordnung und Farbkombination.	167
Transkript 18	Dana und Delia erarbeiten ihre erste Hypothese.	170
Transkript 19	Dana erklärt Delia das Konzept der Konstanthaltung.	172
Transkript 20	Dana und Delia interpretieren den Effekt der Farbkombination.	173
Transkript 21	Dana und Delia planen Experiment 2.	176

Transkript 22	Dana und Delia verwerfen Hypothese 2.	177
Transkript 23	Dana entwickelt Hypothese 3 fast alleine.	178
Transkript 24	Dana und Delia stellen eine Hypothese zum kritischen Experiment auf.	179
Transkript 25	Dana und Delia entdecken abfallende Reaktionszeiten bei der Merkmalsuche.	180
Transkript 26	Dana schlägt ein pseudo-dreifaktorielles Experiment vor und überlegt eine Interpretationsmöglichkeit für nicht gefundene Effekte.	181
Transkript 27	Dana und Delia verwerfen Hypothese 6.	183
Transkript 28	Dido und Dora diskutieren ihr Vorwissen und unterhalten sich nicht aufgabenbezogen.	185
Transkript 29	Dido und Dora diskutieren die Komplexität ihrer ersten Hypothese. 187	
Transkript 30	Dido und Dora bestimmen die Faktoren ihrer ersten Hypothese.	187
Transkript 31	Dido und Dora erwägen, mehrere Hypothesen nacheinander aufzustellen.	188
Transkript 32	Dido und Dora wägen die Stichprobengröße für ihr erstes Experiment ab.	189
Transkript 33	Dido und Dora nehmen Bezug auf den Bestätigungsfehler.	190
Transkript 34	Dido und Dora klären ihr Vorwissen zu Theoretikern und Experimentatoren nach Klahr und Dunbar (1988). ...	194
Transkript 35	Dora erklärt Dido die Interpretation eines signifikanten Haupteffekts.	194
Transkript 36	Dido bemerkt die begrenzten Interpretationsmöglichkeiten pseudo-dreifaktorieller Experimente.	195
Transkript 37	Dido und Dora nehmen zwei Anläufe für ihr nächstes Experiment.	196
Transkript 38	Interpretation einer nur teilweise bestätigten Hypothese.	197
Transkript 39	Dido legt die weiteren Experimente fest.	199
Transkript 40	Dido und Dora formulieren mehrere Hypothesen hintereinander weg.	199
Transkript 41	Dido und Dora wechseln die Konstanthaltung des Faktors Suchtyp zur Verbindungssuche.	200
Transkript 42	Dido und Dora beziehen qualitative und quantitative Aspekte in die Interpretation ihrer Hypothese ein.	201

Transkript 43	Dido und Dora entscheiden sich gegen das kritische Experiment.	203
Transkript 44	Dido spricht über die Falsifikation von Hypothesen.	203
Transkript 45	Dido und Dora rekapitulieren die Suche im Hypothesenraum.	205

Dank

Diese Arbeit hat Prof. Dr. Klaus Opwis betreut. Für seine fachlichen Anregungen und seine konstruktive Kritik bedanke ich mich herzlich.

Teile dieser Arbeit konnte ich im Rahmen des Projektes „Kooperatives wissenschaftliches Denken und Problemlösen“ des schweizerischen Nationalfonds realisieren. Dieses Projekt wurde von Prof. Dr. Dieter Wallach mitbeantragt. Ihm danke ich für sein Engagement und viele richtungweisende Impulse.

Im einzelnen möchte ich folgenden Personen danken: Ester Reijnen hat ihren Datensatz zur visuellen Suche für das simulierte psychologische Labor zur Verfügung gestellt. Olivier Rüeeggsegger hat die Programmierung von *virtue* übernommen. Yolanda Métrailler hat die Videos transkribiert. Allen dreien danke ich zudem für zahlreiche spannende aber auch kritische Diskussionen. Dr. Markus Stöcklin war mir für statistische Fragen immer ein kompetenter Ansprechpartner. Dr. Iris-Katharina Penner hat in vielen Gesprächen für alle Belange ein offenes Ohr gehabt. Marret Popp hat das Manuskript redigiert und hilfreiche Verbesserungsvorschläge gemacht. Prof. Dr. Margarete Boos hat mich als Mentorin ein Stück des Weges begleitet und mir den einen oder anderen Blick über den Tellerrand ermöglicht.

Meinem Mann Philipp Bachmann danke ich für sein ungebrochenes Interesse an meiner Arbeit, seine Ermutigung und seine stetige, liebevolle Unterstützung.

Lörrach, im Frühjahr 2005, Cornelia Kneser

Zusammenfassung

Wissenschaftliches Argumentieren und Problemlösen sind zentrale Ziele der universitären Ausbildung. Die zugrundeliegenden Kompetenzen umfassen komplexe Fertigkeiten in den Bereichen Theoriebildung, Formulierung von Hypothesen, Versuchsplanung, statistische Datenauswertung, Ergebnisinterpretation sowie bei der Integration von Inhalten und Methoden. Dabei treffen Studienanfängerinnen und -anfänger immer wieder auf typische Schwierigkeiten: Wie kann aus einer Theorie eine Hypothese abgeleitet werden? Wie kann der empirische Gehalt einer Hypothese experimentell überprüft werden? Wie ist eine Hypothese im Licht empirischer Befunde zu bewerten? Bisher ist noch wenig darüber bekannt, wie sich der Erwerb von Fertigkeiten für das wissenschaftliche Entdecken fördern lässt. Zwei Herangehensweisen bieten sich für diesen Zweck besonders an, das Arbeiten mit Computersimulationen und das kooperative Lernen. Simulationen sind eine Möglichkeit für Studierende, ein Kernstück wissenschaftlichen Problemlösens, die experimentelle Methodologie, aktiv zu erlernen und einzuüben, bevor sie eigene reale Experimente planen. Wo die erwähnten Schwierigkeiten beim Experimentieren auftreten, bieten kognitive Werkzeuge, an der entsprechenden Stelle in die Simulation integriert, Möglichkeiten der didaktischen Unterstützung. Kooperatives Problemlösen hat sich im Hinblick auf das wissenschaftliche Entdecken als günstig erwiesen. Durch das kooperative Setting werden die Teilnehmenden dazu angeregt, Wissen zu explizieren, Aussagen zu hinterfragen, zu reflektieren und ihre Argumente zu begründen.

Um Studierende beim wissenschaftlichen Problemlösen zu beobachten und im Erwerb der dafür notwendigen Kompetenzen zu unterstützen, wurde *virtue* entwickelt, ein internet-basiertes psychologisches Labor. In ihm können Studierende selbständig Experimente zur Überprüfung der Merkmals-Integrations-Theorie über das wahrnehmungspsychologische Phänomen der visuellen Suche planen und durchführen. Die dem System zugrundeliegende Simulation basiert dabei auf realen experimentellen Daten.

An einer experimentellen Studie nahmen Psychologiestudierende in kooperierenden Dyaden oder als Einzelpersonen teil. Sie absolvierten parallele Vor- und Nachtests. Logfiles der simulierten Experimente wurden ebenso aufgezeichnet wie Videos von den Dialogen der Dyaden und den Einzelpersonen beim lauten Denken. Beide Gruppen zeigten vom Vortest zum Nachtest einen deutlichen Wissenszuwachs, sowohl inhaltsübergreifend im Bereich der Versuchsplanung und der experimentellen Methodik als auch für den inhaltlichen Gegenstandsbereich der visuellen Suche. Die Gruppen unterschieden sich in Bezug auf den Wissenszuwachs nicht voneinander. Deutliche Unterschiede zugunsten der Dyaden ergaben sich hinsichtlich eines insgesamt sorgfältigeren Vorgehens, bezogen auf die Versuchsplanung und die theoriegeleitete Abfolge von Experimenten, und im Sinne einer geringeren Tendenz, eine durch die Ergebnisse eines Experiments falsifizierte Hypothese beizubehalten (Bestätigungsfehler). Eine qualitative Analyse der Transkripte von zwei Einzelpersonen und zwei Dyaden gibt Aufschluß über unterschiedliche Vorgehensweisen in beiden Settings. Einzelpersonen zeigten ein eingeschränktes Repertoire an Lernstrategien. Es erstreckt sich darauf, die Theorie zu konsultieren, sich frühere Experimente nochmals zu vergegenwärtigen, überraschende Ergebnisse zu hinterfragen und zu reflektieren, sowie eine gewisse Stringenz bei der Planung der Experimentalreihe zu demonstrieren. Dyaden nutzten im Gegensatz zu Einzelpersonen Analogien für ihre Argumentation und diskutierten ein breiteres Inhaltsspektrum als diese. Zudem reflektierten Dyaden ihr Vorgehen in häufigeren Metaaussagen als Einzelpersonen.

In der Studie erwies sich das simulierte psychologische Labor *virtue* als geeignete Lernumgebung zur Vermittlung von Fertigkeiten für das wissenschaftliche Problemlösen. Während es in den parallelen Vor- und Nachtests nicht möglich war, eine Überlegenheit von Dyaden gegenüber Einzelpersonen zu messen, zeigten Prozeßanalysen, daß Dyaden beim wissenschaftlichen Problemlösen Vorteile gegenüber Einzelpersonen haben. Damit entsprechen die Ergebnisse der vorliegenden Studie dem Stand der Forschung. Die Ergebnisse erlauben Schlußfolgerungen über den weiteren Forschungsbedarf im Bereich des wissenschaftlichen Problemlösens, über Gestaltung von Lehrveranstaltungen für Methodologie sowie über den Einsatz simulierter Labors im universitären Kontext.

Why study scientific discovery?

For its human and humane value, to understand its mythology, to study the processes of human thinking in some of its most creative and complex forms, to gain insight into the developmental course of scientific thinking, and to design artifacts – computer programs and associated instrumentation – that can carry out some of the discovery processes of science and aid human scientists in carrying out others.

Klahr und Simon (1999).

1

Einleitung

Wissenschaft – wie kann man das lernen? An den Universitäten ist wissenschaftliches Problemlösen ein zentrales Lernziel der Ausbildung. Die daran beteiligten kognitiven Prozesse betreffen die Entwicklung von Theorien, die Ableitung von Hypothesen, die Planung von Experimenten zur Überprüfung der Hypothesen und die Interpretation von Daten. Studienanfängerinnen und -anfänger treffen dabei immer wieder auf einige bekannte Schwierigkeiten (Schunn & Anderson, 1999): Wie können aus einer Theorie Hypothesen abgeleitet werden? Wie sieht ein Versuchsplan aus, mit dem sich eine Hypothese angemessen überprüfen läßt? Wie ist eine Hypothese angesichts bestimmter empirischer Ergebnisse zu bewerten? Bisher ist noch wenig darüber bekannt, wie Studierende auf der Grundlage des Zusammenspiels von Hypothesen, empirischen Daten und deren theoriebezogener Interpretation ein Verständnis für wissenschaftliche Theorien und Modelle erwerben und wie sie bei den genannten Schwierigkeiten unterstützt werden können.

Beim wissenschaftlichen Problemlösen werden verschiedene komplexe Fertigkeiten eingesetzt. Diese können nicht in Form von Faktenwissen vermittelt werden, so daß sich ein Erlernen allein aus Büchern oder in Vorlesungen als ungeeignet erweisen. Simulationen, in diesem Fall virtuelle Labors, sind eine Möglichkeit für Studierende, ein Kernstück wissenschaftlichen Problemlösens, die experimentelle Methodologie, aktiv zu erlernen und einzuüben, bevor sie eigene reale Experimente planen. Die Lernenden haben im Umgang mit Simulationen

einen direkten Zugang zum Gegenstandsbereich und bekommen intrinsisches Feedback, welches Simulationen durch die Ergebnisse der Lernereingaben zurück-melden.

Auch beim Experimentieren mit Simulationen bleiben die erwähnten Schwierigkeiten der Studierenden jedoch bestehen. Für die didaktische Unterstützung bieten sich verschiedene Möglichkeiten. Sowohl die soziale Situation beim wissenschaftlichen Problemlösen als auch die Simulation als solche bieten Ansatzpunkte für Interventionen.

Günstige Auswirkungen auf den Erfolg beim wissenschaftlichen Problemlösen hat die Kooperation in Dyaden gezeigt (Okada & Simon, 1997; Teasley, 1995). In Dyaden regen sich Teilnehmende gegenseitig dazu an, ihr Wissen zu explizieren, Aussagen zu hinterfragen und ihre Argumente zu begründen. Dies alles regt auch zur Reflexion an. Zum Lernerfolg beitragen können ebenfalls Erklärungen auf einem kognitiv angemessenen Niveau, wie sie Kooperationspartner im selben Lernstadium oft erfolgreich geben können. Dies sind jeweils wertvolle Strategien, wie sie sich auch Forschungsgruppen zunutze machen. Daher bietet kooperatives Problemlösen auch ein realistisches Setting für den Erwerb wissenschaftlicher Fertigkeiten.

Eine andere Möglichkeit, Studierende beim wissenschaftlichen Lernen und Problemlösen zu unterstützen, ist die Ausstattung eines simulierten Labors mit didaktischen Hilfen. Hier bieten sich sogenannte kognitive Werkzeuge (*cognitive tools*) an. Sie unterstützen die Lernenden bei den kognitiven Komponenten einer Aufgabe und aktivieren kognitive und metakognitive Lernstrategien. Indem sie Denkprozesse komplementieren, entlasten sie die Lernenden von bestimmten kognitiven Prozessen und ermöglichen es ihnen so, andere kognitive Prozesse zu intensivieren, die den Lernprozeß fördern. Ein Beispiel dafür ist das von van Joolingen (1999) vorgeschlagene Hypothesenscratchpad, das das Aufstellen einer Hypothese in mehrere Schritte untergliedert und den Lernenden Wahlmöglichkeiten für einzelne Textbausteine zur Formulierung von Hypothesen präsentiert.

In der vorliegenden Arbeit wird einer doppelten Fragestellung nachgegangen. Sie vergleicht Studierende beim wissenschaftlichen Problemlösen als Einzelpersonen und Dyaden. Gleichzeitig soll aufgezeigt werden, welche Lernstrategien sich beim Erwerb wissenschaftlicher Fertigkeiten positiv auswirken beziehungsweise wie sich solche Lernprozesse beim Experimentieren mit einem simulierten Labor unterstützen lassen. Hierbei werden sowohl bereichsspezifisches Wissen als auch methodologische Fertigkeiten betrachtet. Die folgenden Fragen stehen dabei im Vordergrund: Wie gehen Novizen beim wissenschaftlichen Problemlösen in einem simulierten psychologischen Labor vor, welche Chancen bietet dies, welche Schwierigkeiten ergeben sich dabei? Inwiefern unterscheiden sich Dyaden und Einzelpersonen in ihrem Vorgehen beim wissenschaftlichen Problemlösen? Wirkt sich gegebenenfalls unterschiedliches Vorgehen von Dyaden und Einzelpersonen auf den Erwerb der Fertigkeiten zum wissenschaftlichen Problemlösen aus? Lassen sich spezifische erfolgreiche Vorgehensweisen identifizieren? Welche didaktischen Maßnahmen erweisen sich als hilfreich?

Überblick über die Arbeit

Die folgenden vier Kapitel geben eine Übersicht über den theoretischen Hintergrund und empirische Ergebnisse zum wissenschaftlichen Problemlösen, zu dessen Vermittlung, zum kooperativen Lernen sowie zum Lernen mit Simulationen. Es folgt die Beschreibung des für die vorliegende Arbeit entwickelten simulierten psychologischen Labors *virtue*. Anschließend erfolgen die Darstellung einer empirischen Untersuchung zum wissenschaftlichen Problemlösen und Lernen von Einzelpersonen und Dyaden mit *virtue* und deren abschließende Diskussion.

Kapitel 2: Wissenschaftliches Problemlösen. Die am wissenschaftlichen Entdecken beteiligten kognitiven Prozesse wurden in verschiedenen Zusammenhängen erforscht. Relevant sind hier insbesondere das logische Schlußfolgern und die Konzeptbildung. Kognitionswissenschaftliche Analysen zeigen, daß sich wissenschaftliches Entdecken als eine Form des Problemlösens einordnen läßt. Anhand von beobachtenden sowie experimentellen empirischen Untersuchungen werden Charakteristika des wissenschaftlichen Denkens aufgezeigt, deren Kenntnis für dessen Vermittlung von Bedeutung sind.

Kapitel 3: Erlernen von Fertigkeiten für das wissenschaftliche Problemlösen.

Um Studienanfängerinnen und -anfänger beim Erlernen wissenschaftlichen Denkens zu unterstützen, ist es notwendig, ihre Voraussetzungen und Schwierigkeiten genau zu kennen. Daher werden zunächst Arbeiten dargestellt, die sich Novizen beim wissenschaftlichen Problemlösen und der Entwicklung wissenschaftlicher Fertigkeiten bei Kindern und Jugendlichen widmen. Dann werden zwei pädagogische Ansätze analysiert, die sich zur Vermittlung von Fertigkeiten in diesem Bereich vorrangig anbieten: Entdeckendes Lernen und Situiertes Lernen.

Kapitel 4: Kooperatives Lernen. Kooperation wirkt sich positiv sowohl auf komplexe Lernprozesse als auch auf wissenschaftliches Problemlösen aus. Hier werden zunächst pädagogisch-psychologische Theorien zum kooperativen Lernen vorgestellt. Es folgen ein Überblick zu den Einflußfaktoren auf die Qualität der Kooperation aus sozialpsychologischer und pädagogisch-psychologischer Sicht und schließlich die Betrachtung didaktischer Maßnahmen zur Unterstützung kooperativen Lernens.

Kapitel 5: Lernen am Computer. Simulationen bieten eine Möglichkeit, Fertigkeiten des wissenschaftlichen Problemlösens einzuüben. Eigenschaften, die sie dafür besonders geeignet erscheinen lassen, werden ebenso beschrieben wie Aspekte, die eine zusätzliche didaktische Unterstützung nahelegen. Diese bietet sich in Form der kognitiven Werkzeuge.

Kapitel 6: Hypothesen. Aus den vorangegangenen vier Kapiteln werden Hypothesen für eine experimentelle Untersuchung des wissenschaftlichen Problemlösens und Lernens von Dyaden und Einzelpersonen in einem simulierten psychologischen Labor abgeleitet.

Kapitel 7: *virtue* – das simulierte psychologische Labor. Zunächst wird in den Gegenstandsbereich für die simulierten Experimente in *virtue*, die visuelle Suche, eingeführt. Es folgt eine Beschreibung der in *virtue* integrierten kognitiven Werkzeuge zur didaktischen Unterstützung des wissenschaftlichen Pro-

blemlösens und Lernens. Abschließend werden die Implementation von *virtue* und eine Studie zu einem der darin integrierten kognitiven Werkzeuge dargestellt.

Kapitel 8: Methode. Es wurde ein Experiment mit zwei Experimentalgruppen, Dyaden und Einzelpersonen durchgeführt. Neben Versuchsplan, -ablauf und -personen werden hier die eigens für die vorliegende Untersuchung entwickelten Tests des bereichsspezifischen und des bereichsübergreifenden Wissens zum wissenschaftlichen Problemlösen mit *virtue* vorgestellt.

Kapitel 9: Ergebnisse I – quantitative Daten aus Tests und Logfiles. Hier werden Resultate des Experiments zum kooperativen wissenschaftlichen Problemlösen mit *virtue* berichtet, und zwar Messungen über den Wissenszuwachs, Daten über das Vorgehen im Umgang mit *virtue* und schließlich einige Kennwerte zu Eigenschaften der Versuchspersonen, die im Hinblick auf die Untersuchung von Bedeutung sind.

Kapitel 10: Ergebnisse II – Qualitative Analyse ausgewählter Einzelfälle. Die Analyse von Transkripten zweier Einzelpersonen und zweier Dyaden gibt Aufschluß über unterschiedliche Vorgehensweisen in beiden Settings. Schwierige Situationen beim wissenschaftlichen Problemlösen und Lernen lassen sich ebenso identifizieren wie besonders hilfreiche Strategien dafür.

Kapitel 11: Diskussion. Abschließend werden die Befunde der Studie diskutiert und Schlußfolgerungen für weiterführende Forschung sowie für den Einsatz von *virtue* in der universitären Lehre gezogen.

2

Wissenschaftliches Problemlösen

Wissenschaft ist der engagierteste Versuch der Menschen, Wissen zu erwerben (Bechtel, 1988). Sie weckte daher schon in der Antike das Interesse der Philosophen, genannt seien hier Plato und Aristoteles. Auch später gab es stets philosophische Ansätze, die sich mit dem Thema Wissenschaft auseinandersetzten, als eigenständige Disziplin entstand die Wissenschaftstheorie jedoch erst im 20. Jahrhundert. Sie untersucht unter anderem Fragen der wissenschaftlichen Methoden, etwa unter welchen Bedingungen wissenschaftliche Theorien angenommen werden können und inwiefern wissenschaftliche Behauptungen durch empirische Daten falsifiziert werden können (Sklar, 1995). Während frühe wissenschaftstheoretische Ansätze etwa bei Plato, Hegel oder Descartes davon ausgingen, daß eine Einsicht in die Natur der Dinge nur durch Nachdenken allein erreicht werden kann, erkennt die moderne Wissenschaftstheorie die Bedeutung der Beobachtung für die Theoriebildung an (Blachowicz, 1998), Mitroff (1974) forderte sogar ausdrücklich die Untersuchung psychologischer Aspekte wissenschaftlichen Entdeckens. Tatsächliche Entwicklungen in den Einzelwissenschaften, etwa in der Physik, werden in der Wissenschaftstheorie vor allem als Beispiele zur Erläuterung der epistemologischen Thesen herangezogen.

Die Psychologie stützt einerseits ihre eigenen wissenschaftlichen Methoden auf die Wissenschaftstheorie (Breuer, 1977; Kriz, Lück & Heidbrink, 1990; Westermann, 2000) und untersucht andererseits wissenschaftliches Vorgehen selbst als Gegenstand ihrer empirischen Forschung. In der experimentellen Psychologie besteht ein weitgehender Konsens über wissenschaftstheoretische Grundlagen. Diese schlagen sich in methodischen Standards nieder: Experimente müssen unter kontrollierten Bedingungen stattfinden; aus Theorien werden empirisch überprüfbare Null- und Alternativhypothesen deduziert; diese werden experimentell operationalisiert; anhand der im Experiment gewonnenen Daten wird mit Hilfe der

α -Fehlerwahrscheinlichkeit statistisch überprüft, ob sie die aufgestellten Hypothesen falsifizieren; anschließend können Hypothesen neu überdacht oder gar Theorien überarbeitet werden (Bortz & Döring, 1995; Suter & Lindgren, 1989). Die Psychologie kennt auch induktive, explorative und deskriptive Herangehensweisen, wie sie für den Fortschritt der Wissenschaft unerlässlich sind (Hayes, 1997). Die deduktive Überprüfung wissenschaftlicher Theorien ist jedoch eine zentrale Methode und daher auch ein Hauptbestandteil der universitären Ausbildung im Fach Psychologie. Als Psychologin formuliert Deanna Kuhn (1989) die Fähigkeiten eines Wissenschaftlers – wie sie auch in der methodischen Ausbildung in Psychologie vermittelt werden sollen – folgendermaßen: (1) eine Theorie bewußt artikulieren; (2) wissen, welche Evidenz eine Theorie unterstützt bzw. widerlegt; (3) begründen können, warum eine Theorie aufgrund der Evidenz angenommen und eine andere Theorie aufgrund derselben Evidenz verworfen wird.

Die empirische Wissenschaftsforschung der Psychologie und auch der Soziologie untersucht, wie einzelne Forscherinnen und Forscher und ganze Forschungsgruppen tatsächlich vorgehen. Sie beschäftigt sich mit dem Forschungsalltag, Mikroprozessen sowie kognitiven und sozialen Aspekten wissenschaftlichen Problemlösens. Dabei kann die empirische Wissenschaftsforschung oft nur Ausschnitte dieses Forschungsgegenstandes betrachten. Einen breiten Überblick dazu geben Feist und Gorman (1998). Wichtige theoretische Impulse erhält die psychologische Forschung auf diesem Gebiet auch immer wieder aus der Kognitionswissenschaft (z.B. Simon, Langley & Bradshaw, 1981).

Wissenschaftliches Denken wird in modernen Arbeiten als komplexer Vorgang betrachtet, an dem viele Teilprozesse beteiligt sind, zum Beispiel Hypothesentesten, Theoriebildung oder Kooperation (Dunbar, 2002). Kognitive Leistungen wissenschaftlichen Denkens treten auch in anderen Zusammenhängen auf und wurden daher in anderen Kontexten untersucht, so zum Beispiel beim logischen Schlußfolgern und bei der Konzeptbildung. Diese werden im Verlauf dieses Kapitels zuerst vorgestellt. Es folgen kognitionswissenschaftliche Analysen des wissenschaftlichen Problemlösens. Den Abschluß bilden empirische Untersuchungen, die auf den kognitionswissenschaftlichen Analysen aufbauen.

2.1 Logisches Schlußfolgern

Die Formulierung und das Testen von Hypothesen beim logischen Schlußfolgern zu untersuchen, hat eine lange Forschungstradition in der Psychologie. Anregung für eine bis heute nicht abgeschlossene Diskussion zum logischen Denken gaben die Experimente von Wason (1966). Typische Experimente aus dieser Tradition bestehen aus Aufgaben, deren korrekte Antwort die Versuchspersonen anhand eines normativen Systems, beispielsweise der propositionalen Logik, ableiten können. Variiert werden dabei sowohl die logische Struktur der vorgegebenen Aufgaben als auch nicht logische Faktoren wie die inhaltliche Einkleidung der Aufgabe oder die Instruktion. Drei zentrale Forschungsergebnisse lassen sich herauskristallisieren (Manktelow, 1999): (1) Intelligente Erwachsene ohne spezielle Ausbildung in propositionaler Logik machen bei diesen Aufgaben viele Fehler. (2) Diese Fehler unterliegen verschiedenen systematischen Gesetzmäßigkeiten. (3) Logisches Schlußfolgern ist stark kontextabhängig.

Der am häufigsten zitierte Fehler ist der Bestätigungsfehler (*confirmation bias*). Menschen tendieren dazu, Experimente so zu planen, daß sie ihre Hypothesen bestätigen statt widerlegen (Tschirgi, 1980); den Hypothesen entgegenlaufende Ergebnisse werden ignoriert; Daten werden verzerrt interpretiert, um an der aufgestellten Hypothese festhalten zu können (Mynatt, Doherty & Tweney, 1977). Auch für Arbeitsgruppen konnte Laughlin und seine Mitarbeiter ähnliche systematische Fehler finden (Laughlin, Bonner & Altermatt, 1998). Gruppen sind Einzelpersonen beim Testen von Hypothesen jedoch im Durchschnitt überlegen, da sich ihre Leistung auf dem Niveau des zweitbesten Gruppenmitglieds einpendelt (Laughlin, VanderStoep & Hollingshead, 1991).

Eine andere Perspektive auf die genannten Befunde hat Poletiek (2001). Sie kann nach einer Analyse theoretischer und empirischer Befunde keinen Bestätigungsfehler erkennen. Stattdessen postuliert sie, daß der Versuch, eine Hypothese empirisch zu bestätigen, zugleich immer auch das Streben nach Falsifikation in sich trägt und umgekehrt. Sie stellt fest, daß die Falsifikation vor allem dann gelingt, wenn die Aufmerksamkeit darauf fokussiert ist. Dies kann auf verschiedene Arten geschehen. Um die Befunde zu erklären, favorisiert Poletiek ein statistisches Modell, das zwei Dilemmas gegeneinander abwägt: Welche Evidenz soll

zu welchen Kosten erreicht werden? Wieviele Entscheidungsfehler der einen Art sind auf Kosten wievieler Entscheidungsfehler der anderen Art akzeptabel? Um ihr Modell weiter zu überprüfen, fordert Poletiek vor allem die weitere Erforschung des Hypothesentestens im Alltag.

Evans (2002) wählte eine andere Annäherung an die Befunde zu Fehlern beim logischen Schlußfolgern. Er führt die These an, daß dem logischen Schließen ein dualer Denkprozeß zugrunde liege: Einerseits laufen implizit mächtige, kontextabhängige, von der Kapazität des Arbeitsgedächtnisses jedoch unabhängige Denkprozesse ab. Andererseits gibt es explizite Denkprozesse, die allgemeines Schließen erlauben, die aber langsam, seriell und durch die Kapazität des Arbeitsgedächtnisses beschränkt sind. Um normativ korrekte Schlußfolgerungen zu ziehen, muß auf die expliziten Denkprozesse zurückgegriffen werden. Sie korrelieren, im Gegensatz zu den impliziten Denkprozessen, mit Intelligenz. Für persönliche Schlußfolgerungen im Alltagsleben sind oftmals die Ergebnisse des impliziten Systems ausreichend. Die impliziten Schlußfolgerungsprozesse sind universell und müssen unterdrückt werden, damit die expliziten Denkprozesse zum Zug kommen können.

Beim schlußfolgernden Denken wird laut Evans (2002) meist nur eine Hypothese gleichzeitig betrachtet. Diese Hypothese ist angesichts des aktuellen Hintergrundwissens die wahrscheinlichste oder plausibelste und wird solange beibehalten, wie sie die Kriterien zufriedenstellend erfüllt. Ist dies aufgrund empirischer Evidenz nicht mehr der Fall, wird die Glaubwürdigkeit der Hypothese graduell reduziert. Entsprechen die Ergebnisse den auf dem Vorwissen basierenden Erwartungen, werden sie unkritisch akzeptiert, ohne daß ihre methodische Qualität überprüft würde. Erst wenn sie den Erwartungen widersprechen, werden sie methodisch angezweifelt. Dies ist aus Sicht der Wissenschaftstheorie allerdings keine korrekte Vorgehensweise.

Aus seiner Analyse folgert Evans (2002), daß Wissenschaft nicht so sehr aus der Popperschen Perspektive als vielmehr nach Bayesschen Grundsätzen betrieben wird: Wissenschaftler verfügen über ein umfangreiches Vorwissen über ihren Fachbereich. Auf diese Weise basieren ihre impliziten Schlußfolgerungen auf korrekten Vorannahmen, weshalb sie meist näher an der logischen Norm argumentieren, als

experimentelle Untersuchungen zum abstrakten logischen Schlußfolgern vermuten ließen. Andererseits kann das Einbeziehen von Vorwissen zu einem unkritischen Umgang mit hypothesenkonformen Ergebnissen führen. Hypothesen werden positiv getestet – eine normalerweise erfolgreiche Strategie in der Wissenschaft. In der Widerlegung von Hypothesen sieht Evans dennoch kein Problem, da er Wissenschaft im Sinne Kuhns (1976) als ein gemeinschaftliches Unternehmen betrachtet und die *scientific community* für eine kritische Bewertung wissenschaftlicher Behauptungen sorgt.

Für das Erlernen von Fertigkeiten des wissenschaftlichen Problemlösens ist Evans' (2002) Analyse als Einschränkung zu betrachten: Da die impliziten Prozesse sehr tief verankert sind und automatisch ablaufen, lassen sie sich durch explizite Lernprozesse nur schwer beeinflussen. Stattdessen kann Lernen auf der Ebene des Vorwissens (und anderer Prozesse des wissenschaftlichen Denkens) stattfinden. Dieses Vorwissen kann die impliziten Denkprozesse beeinflussen und auf diese Weise wieder die Qualität des hypothetischen Schlußfolgerns verbessern.

Während im vorherigen Abschnitt Hypothesen als Schritte zu einer formal korrekten deduktiven Schlußfolgerung betrachtet wurden, geht es bei den im Folgenden vorgestellten Ansätzen zur Konzeptbildung um induktive Hypothesenbildung.

2.2 Konzeptbildung

Frühe Arbeiten zur Konzeptbildung und Formulierung von Hypothesen stammen von Hovland (1952) und Bruner, Goodnow und Austin (1956). Konzeptbildung ist für die Wissenschaft von großer Bedeutung bei der Definition von Begriffen und Kategorien, der Vorhersage aufgrund der Zuordnung zu Kategorien und trägt damit wesentlich zur Theoriebildung bei. In den klassischen Experimenten dazu sollten die Versuchspersonen Regeln für die Klassifikation von Objekten herausfinden. Der Versuchsleiter hatte eine Klasse von Objekten ausgewählt, beispielsweise visuelle Stimuli verschiedener Farben, Formen und Schraffuren. Einige dieser Objekte gehörten zu der durch die Klassifikationsregel definierten Zielkategorie, etwa alle quadratischen, gestreiften Objekte. Die Versuchsperson konnte die Klassifikationsregel entweder anhand von Stimuli erschließen, die ihr der Versuchsleiter vorlegte (Rezeptionsparadigma), oder sie konnte selbst Stimuli auswählen (Selektionspara-

digma). Für jedes Objekt mußte sie eine Vorhersage machen, ob das vorliegende Objekt zur Zielkategorie gehört oder nicht, und bekam dann vom Versuchsleiter eine Rückmeldung über die Korrektheit ihrer Vorhersage.

Dieses Paradigma regte viele Arbeiten zur Induktion von Hypothesen und zur Konzeptbildung an, ist jedoch in seiner Aussagekraft für das wissenschaftliche Denken begrenzt. So kritisiert Reimann (1991), daß Konzepte in diesem Paradigma viel zu eng definiert sind. Die Merkmale interagieren nicht miteinander. Natürlich auftretende Probleme haben im Gegensatz zu den beschriebenen Konzepten keine klar definierten Grenzen, sondern werden anhand ihrer Typikalität definiert; untersucht man sie wissenschaftlich, so sind die Ergebnisse oft nicht eindeutig. Zudem wird bei diesem Paradigma das Lernen eines Konzepts von jeglichem darüber hinausgehenden Problemkontext isoliert, Vorwissen hat keine Bedeutung. Auch die Kosten für die Identifikation eines relevanten Merkmals spielen keine Rolle, weil alle Merkmale bereits im ersten Beispiel enthalten sind.

Trotz der Kritik griff Dunbar (1993) den Ansatz der Konzeptbildung für eine Laboruntersuchung zum wissenschaftlichen Denken auf. Als Gegenstandsbereich wählte Dunbar ein näher an der wissenschaftlichen Realität orientiertes Problem aus der Genetik: Mechanismen gegenseitiger Aktivierung und Hemmung einzelner Gene. In zwei Untersuchungen wurden insgesamt 40 Studierende im Grundstudium der Biologie dabei beobachtet, wie sie Experimente in einem simulierten genetischen Labor durchführten. Neben den Daten aus der Simulation wurden die verbalen Protokolle ausgewertet, welche die Versuchspersonen beim lauten Denken produziert hatten. In der ersten Untersuchung zeigte sich, daß nur wenige der 20 teilnehmenden Studierenden den der Simulation zugrundeliegenden genetischen Mechanismus entdeckten. Alle Versuchspersonen begannen ihre Experimente mit einer Aktivierungshypothese, fanden jedoch keine empirische Bestätigung für ihre anfängliche Hypothese. Statt Hypothesen über andere Mechanismen zur Erklärung der genetischen Daten zu formulieren, begannen 13 Versuchspersonen, die einmal aufgestellte Hypothese zu verzerren und ihr widersprechende empirische Befunde zu ignorieren. Sie versuchten nicht, die

Ursache für die unerwarteten Ergebnisse herauszufinden. Dieses Ziel verfolgten nur diejenigen Studierenden, die letzten Endes den zugrundeliegenden Hemmungsmechanismus auch entdeckten.

Diese Beobachtung regte Dunbar (1993) zu seiner zweiten Untersuchung an. Hier ging er der Frage nach, ob das Ziel der Versuchspersonen, Daten in Übereinstimmung mit ihrer Anfangshypothese zu finden, sie daran hindert, Alternativhypothesen aufzustellen. Zu diesem Zweck wurde der Gegenstandsbereich leicht modifiziert: Ein bisher wirkungsloses Gen wurde mit einem Aktivierungsmechanismus versehen. Auch in diesem Experiment begannen alle 20 Versuchspersonen mit einer Aktivierungshypothese. Nach deren Bestätigung für das modifizierte Gen entdeckten 14 Versuchspersonen den Hemmungsmechanismus der anderen Gene des Gegenstandsbereichs und kamen so zu einer korrekten Lösung.

Dunbar (1993) wirft mit diesen Untersuchungen ein anderes Licht auf die Forschung zum Bestätigungsfehler (etwa Mynatt, Doherty & Tweney, 1977). Mynatt et al. beobachteten anhand eines künstlichen Gegenstandsbereichs, daß Personen bestätigende statt falsifizierende Experimente durchführten, Ergebnisse, die ihre Hypothesen widerlegten, ignorierten oder verzerrt wahrnahmen und keine alternativen Hypothesen in Betracht zogen (für einen qualitativen Ansatz mit ähnlichen Ergebnissen siehe Kuhn, 1991). Dunbar führte seine Untersuchungen mit einem realistischen Gegenstandsbereich durch und suchte nach Gründen für den Bestätigungsfehler. Ein Grund für das Festhalten an einer Hypothese könnten Zweifel an der Qualität der erhobenen Daten sein. Aufgrund mangelnden Vorwissens war es für die Studierenden zudem schwierig, überraschende Ergebnisse als solche zu erkennen und für ihre weiteren Experimente zu nutzen. Gleichzeitig stellte Dunbar fest, daß seine Versuchspersonen Daten, die ihren Hypothesen widersprachen, nicht vollständig ignorierten, sondern ihre Hypothesen in Abhängigkeit von ihren Zielen sehr wohl, wenn auch meist nur geringfügig, modifizierten. Erst wenn gehäuft widersprüchliche Daten zu ihren Hypothesen auftraten, setzten sich die Versuchspersonen ein neues Ziel und formulierten eine alternative Hypothese. Mit dieser Argumentation befindet Dunbar sich in Übereinstimmung

mit Evans (2002), der ebenfalls einerseits die Bedeutung des Bestätigungsfehlers relativiert und andererseits die Rolle des Vorwissens für das wissenschaftliche Denken hervorhebt.

Allerdings hat auch Dunbars Ansatz seine Schwächen, werden die Gründe für das Verhalten der Studierenden im Umgang mit dem genetischen simulierten Labor doch aus den Daten erschlossen, aber nicht experimentell untersucht. Zudem wird in dem zur Untersuchung genutzten simulierten Labor zwar ein realistischer Gegenstandsbereich genutzt, Meßfehler, wie sie bei realen Laborexperimenten doch regelmäßig auftreten, sind in der Simulation aber nicht enthalten.

Dunbar (1993) geht mit seiner Untersuchung weit über den theoretischen Rahmen der Konzeptbildung hinaus. Er diskutiert vor allem die Bedeutung von Zielen für die Lösung eines Forschungsproblems und betrachtet die Formulierung von Hypothesen selbst als Zielsetzung. Dieser Ansatz aus der Kognitionswissenschaft wirkte sehr befruchtend auf die Forschung zum wissenschaftlichen Denken. Im Folgenden werden zunächst grundlegende Aspekte des wissenschaftlichen Problemlösens dargestellt. Daran schließen sich verschiedene Weiterentwicklungen und abgeleitete empirische Untersuchungen an.

2.3 Kognitionswissenschaftliche Analysen

In seinem Aufsatz „scientific discovery and the psychology of problem solving“ formulierte Herbert Simon (1966) erstmals die Idee, wissenschaftliches Entdecken als Problemlöseprozeß aufzufassen im Sinne einer Suche in einem Problemraum. Er stellte fest, daß wissenschaftliche Entdeckungen seltene, aber keine plötzlichen Ereignisse sind, sondern vielmehr die Früchte einer langfristigen, intensiven Suche. Bereits in diesem Aufsatz beschreibt Simon das wissenschaftliche Problemlösen als Anwendung von Heuristiken.

Ein umfassendes Modell des Problemlösens stellten Newell und Simon (1972) auf. Ein Problem besteht demnach aus einem Anfangszustand und einem Zielzustand sowie einer Menge erlaubter Transformationen von einem Zustand zum anderen, sogenannten Operatoren. Werden die Operatoren in der richtigen Reihenfolge angewendet, so ergibt sich ein Lösungsweg, der über einige Zwischenzustände und Unterziele vom Anfangs- zum Zielzustand führt. Die Menge der Zustände, Ziele, Operatoren und Regeln wird als Problemraum bezeichnet.

Problemlösen wird betrachtet als Suche in diesem Problemraum nach einem Lösungsweg, der Anfangs- und Zielzustand miteinander verbindet. Ist die Sequenz von Operatoren und Zuständen, die zur Lösung eines Problems führt, unbekannt, so können verschiedene Heuristiken zum Einsatz kommen. Je mehr Vorwissen zu einem Problem vorhanden ist, desto mächtigere Heuristiken können zur Suche im Problemraum eingesetzt werden.

Diese Definition des Problemlösens läßt sich auf verschiedenste Situationen anwenden. Das Spektrum reicht von Alltagsproblemen (Newell & Simon, 1972) und Experimentalaufgaben wie dem Turm von Hanoi (Klix, 1971) bis zu komplexer medizinischer Diagnostik (Boshuizen & Schmidt, 1992; Elstein, Shulman, & Sprafka, 1978, Patel & Groen, 1986). Übertragen auf wissenschaftliches Entdecken als Problemlösen bedeutet dies, daß auch Wissenschaft als Suche mit Zielen, Unterzielen, Operatoren und der Anwendung von Heuristiken betrachtet werden kann. Simon (1966) formuliert folgende drei Annahmen darüber, was einen erfolgreichen Wissenschaftler von einem weniger erfolgreichen abhebt: (1) Er hatte Glück; (2) er hat länger und hartnäckiger gesucht als seine Kollegen; oder (3) er wandte die mächtigeren Heuristiken an. Laut Simon sind wahrscheinlich alle drei Aspekte bei einer Entdeckung im Spiel.

Simon, Langley und Bradshaw (1981) differenzieren weiter und unterscheiden im Sinne Kuhns (1976) Normalwissenschaft von revolutionärer Wissenschaft. In der Normalwissenschaft sind die Probleme und Forschungsziele hinlänglich bekannt, ebenso wie die Untersuchungsmethoden. Experten, in diesem Fall ausgewiesene Wissenschaftler des entsprechenden Fachgebiets, sind in der Normalwissenschaft in der Lage, starke Problemlösestrategien, also vorwissensreiche Heuristiken oder gar Algorithmen einzusetzen. Die revolutionäre Wissenschaft findet hingegen unter anderen Voraussetzungen statt. Neue Fragen werden bei den wissenschaftlichen Revolutionen aufgeworfen, wenn neue Repräsentationen für die empirischen Ergebnisse erforderlich werden. Die Probleme sind noch nicht klar definiert, und es existiert nur geringes Vorwissen. In diesem Fall sind die Wissenschaftler nach Simon, Langley und Bradshaw (1981) auf die Anwendung schwacher Heuristiken wie beispielsweise die Mittel-Ziel-Analyse angewiesen, ihr Vorgehen ähnelt damit demjenigen von Novizen. Die Autoren sehen durchaus

Unterschiede zwischen dem allgemeinen Problemlösen, wie es Newell und Simon (1972) beschreiben, und wissenschaftlichem Entdecken. So wurde Problemlösen vorrangig bei Einzelpersonen untersucht, wohingegen wissenschaftliches Entdecken ein sozialer Prozeß ist. Daß die Probleme beim wissenschaftlichen Entdecken nicht so klar definiert sind, wurde bereits erwähnt. Dennoch kommen Simon et al. nach gründlicher Analyse zu dem Schluß, daß wissenschaftliches Entdecken als Sonderfall menschlichen Problemlösens betrachtet werden kann.

Ausgehend vom theoretischen Ansatz des wissenschaftlichen Entdeckens als Problemlösen wurde dieses Thema in der Kognitionswissenschaft mit unterschiedlichen methodischen Herangehensweisen weiterverfolgt, einerseits mit verschiedenen empirischen Methoden wie teilnehmender Beobachtung und experimentellen Laboruntersuchungen, andererseits durch kognitive Modellierung. Während eine Richtung der kognitiven Modellierung es sich zum Ziel gesetzt hat, wissenschaftliche Entdeckungen mit Hilfe von Techniken aus dem Bereich der künstlichen Intelligenz zu unterstützen (für einen Überblick siehe Darden, 1997; Langley, 2000; Oppwis, 1992), werden mit Hilfe anderer kognitiver Modelle die kognitiven Prozesse beim Menschen während des wissenschaftlichen Entdeckens untersucht. Hierbei macht man sich die Eigenschaft der Computermodelle zu nutze, daß sie eine Präzisierung der Annahmen über das in Frage stehende Phänomen erzwingen, und zwar in einem Ausmaß, welches über dasjenige verbal formulierter Theorien hinausgeht.

Basierend auf einer detaillierten Analyse historischer, wissenschaftlicher Entdeckungen wurden kognitive Modelle wissenschaftlichen Entdeckens entwickelt. So konnte BACON von Simon, Langley und Bradshaw (1981) die Entdeckung des 3. Keplerschen Gesetzes sowie weiterer wissenschaftlicher Gesetze aus verschiedenen Gegenstandsbereichen, etwa denen von Ohm und Archimedes, als Produktionensystem simulieren. Ein Produktionensystem zur Entdeckung des Harnstoffzyklus entwickelten Kulkarni und Simon (1988) und identifizierten fast die Hälfte der angewandten Heuristiken als eher allgemeine, also schwache Heuristiken (siehe auch Grasshoff & May, 1995).

Einen Überblick über Arbeiten zur kognitiven Modellierung wissenschaftlichen Entdeckens geben Shrager und Langley (1990). Die Autoren unterscheiden Ansätze zur Repräsentation des durch wissenschaftliches Entdecken zutage geförderten Wissens und Ansätze zu wissenschaftlichen Handlungen. Frühe Arbeiten zur Repräsentation beschäftigten sich mit der Entdeckung empirischer Regelmäßigkeiten und der Ableitung von Taxonomien und wissenschaftlichen Gesetzen. Dabei kamen allgemeine Heuristiken zum Einsatz. Andere Modelle konzentrieren sich auf den Ablauf einer Sequenz qualitativer Zustände beim wissenschaftlichen Entdecken, auf das Schlußfolgern mit Analogien oder die Durchführung wissenschaftlicher Experimente. Zusammenfassend ermöglichen es diese kognitiven Modelle, allgemeine und bereichsspezifische Mechanismen des wissenschaftlichen Problemlösens zu identifizieren sowie das unterschiedliche zugrundeliegende Wissen und dessen Struktur zu definieren.

2.4 Empirische Untersuchungen

Konnten die vorgestellten kognitiven Modelle formale Aspekte des wissenschaftlichen Entdeckens erfassen, so wurden doch auch die Grenzen dieser Methode deutlich. Einerseits bieten die Grundlagen der Modelle historischer Entdeckungen, die sich auf die Analyse von Veröffentlichungen, Laborbüchern und Briefen stützen, keinen Einblick in die Mikroprozesse wissenschaftlichen Denkens, andererseits konnten einige Aspekte wie die Rolle externer Repräsentationen und die Rolle von Kooperation innerhalb und zwischen Forschungseinrichtungen in den kognitions-wissenschaftlichen Modellen nicht berücksichtigt werden (Shrager & Langley, 1990). Zur Untersuchung dieser Aspekte eignen sich sowohl die teilnehmende Beobachtung als auch experimentelle Laboruntersuchungen. Hier sollen zunächst einige Ergebnisse aus Beobachtungsstudien vorgestellt werden. Anschließend werden neuere experimentelle Studien dargestellt.

2.4.1 Beobachtungen in Forschungslabors

Die Beobachtung von Wissenschaftlern im Labor ist eine sehr anspruchsvolle Untersuchungsmethode (Klahr, 2000). Die Beobachter müssen über vertiefte Kenntnisse des Forschungsbereiches verfügen, in dem die beobachteten Wissenschaftler forschen, sie müssen angesichts der zugänglich gemachten Informationen

das volle Vertrauen der Beobachteten genießen. Zudem ist diese Art der Beobachtung sehr zeitaufwendig. Der Erfolg einer Forschergruppe wird in solchen Untersuchungen üblicherweise anhand von Publikationen und eingeworbenen Drittmitteln gemessen. Frühe Beobachtungsstudien in wissenschaftlichen Labors stammen aus dem Bereich der Soziologie (Latour & Woolgar, 1979; Knorr-Cetina, 1984). Beide Arbeiten fragen nach der Bedeutung sozialer Faktoren für wissenschaftliche Entdeckungen und kommen zu dem Schluß, daß ihre Bedeutung unter Betonung der kognitiven Aspekte vernachlässigt wurde. Latour und Woolgar erklären dies damit, daß der Entstehungsprozeß dieser neuen wissenschaftlichen Erkenntnisse mit seinen vielfältigen sozialen Aspekten ausgeblendet wird, sobald wissenschaftliche Fakten etabliert und logisch begründet worden sind. Auch Kraut, Egido und Galegher (1990) untersuchten die Zusammenarbeit zwischen Wissenschaftlern, insbesondere den Einfluß von Faktoren wie thematischer Verwandtschaft und räumlicher Nähe, wobei letztere für die Entstehung fächerübergreifender Kooperation eine bedeutende Rolle spielte. Ein weiterer Themenkreis wird von Latour (1986) angesprochen: die Bedeutung von Artefakten, etwa Meßgeräten und visuellen Repräsentationen für die Wissenschaft. Von der Kognitionswissenschaft wurden diese Fragen durch Shrager und Langley (1990), Dunbar (1995) und Thagard (1997, 1998b) aufgegriffen.

Dunbar führte intensive Beobachtungsstudien in vier führenden molekular-genetischen Labors durch (Dunbar, 1995, 1997, 2000a). Im Mittelpunkt seiner Untersuchungen stand die Beobachtung von Forschungskolloquien. Vor und nach den Kolloquien wurden Interviews mit den entsprechenden Referenten geführt. Die Transkripte der Kolloquien und Interviews wurden diskurs- und protokollanalytisch ausgewertet.

Übergeordnete Fragestellung von Dunbars Untersuchungen war, welche Mechanismen konzeptuellen Veränderungen und Einsicht zugrundeliegen (Dunbar, 1995, 1997, 2000a). Er identifizierte drei Mechanismen: (1) Umgang mit nicht hypothesenkonformen Daten, (2) Heranziehen von Analogien, (3) verteilte Diskussion (*distributed reasoning*). Zum Umgang mit inkonsistenten Ergebnissen stellte Dunbar fest, daß bei ihrem Auftreten zunächst nur kleine Abänderungen an den Hypothesen, etwa Generalisierungen oder Spezifizierungen vorgenommen werden.

Wurde eine Inkonsistenz mit einer ganzen Klasse von Hypothesen manifest, so beteiligten sich an der Formulierung einer völlig neuen Hypothese meist mehrere Mitglieder der Arbeitsgruppe. Erfahrene Wissenschaftler neigten bei inkonsistenten Ergebnissen eher dazu, eine Hypothese sofort zu verwerfen. Jungwissenschaftler hielten länger an einmal aufgestellten Hypothesen fest. Beides könnte laut Dunbar jeweils als *bias* in verschiedene Richtungen betrachtet werden.

Analogien stuft Dunbar (1995, 1997, 2000a, 2001) nach ihrer Nähe zur aktuellen Fragestellung ein. Lokale Analogien zu früheren Experimenten traten häufig auf und wurden zur Lösung technischer experimenteller Probleme herangezogen. Regionale Analogien, etwa die Abbildung eines ganzen Systems von Relationen aus einem verwandten Gegenstandsbereich auf den in Frage stehenden Gegenstandsbereich, kamen wesentlich seltener zum Tragen, und zwar dann wenn Theorien erarbeitet oder eine neue Reihe von Experimenten geplant wurden. Analogien zu gänzlich fernen Gegenstandsbereichen spielten für den Fortschritt der Forschung in den Labors keine Rolle. Sie kamen zum Einsatz, um Ergebnisse im Nachhinein zu verdeutlichen oder wenn neuen Labormitgliedern oder Laien etwas erklärt werden sollte. In demjenigen Labor, das im Beobachtungszeitraum den wenigsten Erfolg hatte, wurden in den Besprechungen keine Analogien benutzt. Als Grund dafür vermutete Dunbar, daß alle Mitglieder der Arbeitsgruppe eine sehr ähnliche Ausbildung hatten, weshalb sie keine fruchtbaren regionalen Analogien ziehen konnten. Unterschiedliches Vorwissen scheint also für den Einsatz von Analogien und die Entwicklung wissenschaftlicher Ideen von Bedeutung zu sein. Experten benutzten zudem insgesamt mehr Analogien und machten wahrscheinlich auch deshalb mehr Entdeckungen.

Als verteilte Diskussion bezeichnet Dunbar (1997) ein Gespräch zwischen mehreren Mitgliedern einer Arbeitsgruppe über Hypothesen, Experimente, Methoden oder die Interpretation von Ergebnissen, wobei verschiedene Personen neue Aspekte zum Diskussionsthema beitragen. Bei Einzelpersonen wurden häufig Schwierigkeiten beobachtet, alternative Interpretationen erhobener Daten zu generieren. In verteilten Diskussionen bringen andere Teilnehmende neue Repräsentationen des Diskussionsthemas ein. Darauf können wieder andere Teilnehmende weitere Erklärungen, Begründungen und Schlußfolgerungen aufbauen. Aus densel-

ben Prämissen zogen Wissenschaftler häufig ganz unterschiedliche Schlüsse, weil sie die Prämissen sehr verschieden interpretierten. Dunbar setzt sich deutlich ab von Ergebnissen aus der Kleingruppenforschung (siehe etwa Laughlin, Vander-Stoep, Hollingshead, 1991), wonach Gruppen gegenüber Einzelpersonen nicht überlegen seien. Er erklärt die Vorteile von Diskussionen in Forschungsgruppen damit, daß dort im Gegensatz zu den Laborgruppen der Kleingruppenforschung eine langfristige Arbeitsgruppe gemeinsame Ziele verfolgt und eine gemeinsame Wissensbasis mit leicht unterschiedlichem Vorwissen vorhanden ist.

Dunbar (1995, 2000a, 2001) wägt Beobachtungsstudien und Laboruntersuchungen gegeneinander ab. Er nennt die beiden Forschungsansätze auch „in vivo“ beziehungsweise „in vitro“ Untersuchungen. Für in vivo Forschung spricht ihre ökologische Validität. Nur so lassen sich Forscher mit ausgewiesener Expertise in ihrer realen Umgebung mit tatsächlichen wissenschaftlichen Problemen untersuchen. Bei der in vivo Forschung gewonnene Erkenntnisse geben dann Anregungen für Laboruntersuchungen in vitro. Dadurch lassen sich Probleme traditioneller Laborexperimente, etwa artifizielle Experimentalaufgaben vermeiden. Klahr (2000) greift die Diskussion auf und formuliert die Vorteile von in vitro Untersuchungen wie folgt: (1) Die Versuchspersonen können genau bestimmt werden, Experten, Doktoranden oder Studierende; (2) das Vorwissen der Versuchspersonen kann kontrolliert werden; (3) eine detaillierte Beobachtung dynamischer Vorgänge beim wissenschaftlichen Entdecken ist leicht möglich; (4) der Gegenstandsbereich für die Experimente der Versuchspersonen kann genau definiert werden; (5) in Laboruntersuchungen können sowohl Erfolge als auch Fehler beim wissenschaftlichen Entdecken beobachtet werden. Klahr und Simon (1999) betonen darüber hinaus die Bedeutung explorativer Experimente, die nicht durch klare Hypothesen geleitet sind und teilweise ohne Kontrollbedingungen durchgeführt werden, sondern nur durch eine vage und allgemeine Richtung von Fragestellungen bestimmt sind und nennen historische Entdeckungen aufgrund solcher Untersuchungen. Diese Experimente fördern Phänomene zu Tage, die dazu anregen, neue Repräsentationsformen und neue Hypothesen zu generieren.

2.4.2 Experimentelle Untersuchungen

Im Rahmen einer gemeinsamen experimentellen Untersuchung entwickelten Klahr und Dunbar (1988) basierend auf dem Konzept des Problemlösens von Newell und Simon (1972) das Modell des *scientific discovery as dual space search* (SDDS, wissenschaftliches Entdecken als Suche in zwei Problemräumen). Eine Formalisierung des SDDS-Modells definierte Klahr (2000). Der erste Problemraum ist der sogenannte Hypothesenraum. Ausgehend vom Vorwissen über den Gegenstandsbereich besteht das Ziel darin, eine Hypothese aufzustellen, die das Vorwissen in eine präzisere oder umfassendere Form bringt. Anschließend werden die aufgestellten Hypothesen bezüglich ihrer anfänglichen Plausibilität bewertet. Je plausibler die Hypothese ist, desto früher wird sie getestet (Klayman & Ha, 1987).

Der zweite Problemraum ist der Experimenterraum. Hier sind mögliche Experimente für den in Frage stehenden Gegenstandsbereich repräsentiert. Ziel der Suche im Problemraum ist es, ein Experiment zu generieren, das den aktuellen Fokus im Gegenstandsbereich untersucht. Dieser Fokus kann eine aufgestellte Hypothese sein. Eine Suche im Experimenterraum ist aber auch ohne Hypothese möglich.

Klahr (2000) unterscheidet im SDDS-Modell drei übergeordnete Prozesse: Das Aufstellen einer Hypothese (*hypothesis generation*), die Versuchsplanung (*test hypothesis*) und die Ergebnisinterpretation (*evidence evaluation*). Diese Prozesse können sich gegenseitig beeinflussen. So kann beispielsweise das durch die Interpretation der Ergebnisse neu dazugewonnene Wissen die Suche im Hypothesenraum lenken. Ebenso kann die Suche im Experimenterraum Anhaltspunkte für die Suche im Hypothesenraum geben. Eine Suche im Experimenterraum findet daher nicht erst während der Versuchsplanung, sondern möglicherweise bereits beim Aufstellen der Hypothesen statt. Es können sogar gezielt Experimente durchgeführt werden, um Parameter einer Hypothese zu spezifizieren.

Der Prozeß der Ergebnisinterpretation basiert auf dem Vergleich der Hypothesen und der Ergebnisse. Dabei geht Klahr (2000) von kumulierter Evidenz aus, das heißt, daß mehrere Experimente zur aktuellen Hypothese durchgeführt wurden. Im Evaluationsprozeß wird entschieden, ob die Datengrundlage hinreichend ist,

um über die Gültigkeit einer Hypothese zu entscheiden. Zusätzliche Kriterien für die Ergebnisinterpretation können beispielsweise Plausibilität und Funktionalität sein.

Empirisch untersuchten Klahr und Dunbar (1988) das SDDS-Modell in einer Untersuchung mit der künstlichen Experimentalumgebung Big Trak. Dabei handelt es sich um einen computergesteuerten Roboter. Die Bewegungen des Roboters können mit Hilfe verschiedener Zahl- und Funktionstasten programmiert und anschließend abgespielt werden. Aufgabe der Versuchspersonen ist es, die genaue Funktion der sogenannten Repeat-Taste entdecken.

An der ersten Untersuchung nahmen 20 Studierende im Grundstudium teil. Nach einer Instruktion für das laute Denken und einer Einführung in die Funktionen von Big Trak wurden die Versuchspersonen aufgefordert, eine erste Hypothese über die Funktion der Repeat-Taste zu formulieren und anschließend die Funktion der Repeat-Taste herauszufinden, indem sie Programme für Big Trak schrieben und ausführen ließen. Innerhalb von 45 Minuten entdeckten 19 der 20 Studierenden die Funktion der Repeat-Taste. Bezüglich der in den Protokollen kodierten Hypothesen zeigten die Versuchspersonen Schwierigkeiten bei der Interpretation ihrer empirischen Befunde. Von den Hypothesen, die durch die empirische Evidenz bestätigt worden waren, wurden trotzdem 25% verworfen und von den Hypothesen, die durch die empirischen Befunde widerlegt worden waren, wurden über 50% beibehalten (sogenannter *confirmation bias*, Bestätigungsfehler).

Bei den Versuchspersonen kristallisierten sich zwei Strategien heraus, diejenige der Theoretiker und diejenige der Experimentatoren. Theoretiker stellten zunächst eine oder mehrere Hypothesen über die Funktion der Repeat-Taste auf, führten also eine Suche im Hypothesenraum durch, und definierten dann im Experimenterraum entsprechende Programme, um diese Hypothesen zu testen. Sie sagten die korrekte Funktion der Repeat-Taste in einer Hypothese vorher. Experimentatoren vernachlässigten ihre Anfangshypothese hingegen sehr schnell und begannen mit einer Suche im Experimenterraum, bei der sie eine ganze Reihe von Experimenten durchführten, um erst viel später wieder zu ihren Hypothesen zurückzukehren. Sie erkannten die richtige Funktion der Repeat-Taste aufgrund von Experimenten im kritischen Bereich des Experimenterraums. Von den Ver-

suchspersonen ließen sich 7 als Theoretiker, 13 als Experimentatoren identifizieren. Die Datenanalyse ergab, daß Theoretiker in ihrem Vorgehen effizienter waren als Experimentatoren. Theoretiker fanden die Funktion der Repeat-Taste verglichen mit den Experimentatoren in der Hälfte der Zeit heraus und führten auch nur halb so viele Experimente durch.

Um den Effekt der Formulierung von Hypothesen zu überprüfen, wurden in einem zweiten Experiment die Versuchspersonen aufgefordert, noch vor der ersten Programmierung von Big Trak so viele plausible Hypothesen wie möglich aufzustellen. Gleich bei der Formulierung dieser Hypothesen gelang es fünf von insgesamt zehn Versuchspersonen, auch die richtige Hypothese über die Repeat-Funktion zu generieren. Alle Versuchspersonen entdeckten die Funktion der Repeat-Taste. Sie brauchten dafür nur etwas mehr als halb so lange wie die Theoretiker aus Experiment 1 und führten auch nur knapp über die Hälfte an Experimenten durch wie sie. Auch hier ließen sich Theoretiker und Experimentatoren in ähnlicher Weise unterscheiden wie in Experiment 1, auch die Unterschiede bezüglich der Effizienz sind ähnlich. Alle Theoretiker, aber nur einer der Experimentatoren stellten unter ihren anfänglichen Hypothesen die korrekte auf. In der Untersuchung zeigte sich, daß die Zuordnung zu den beiden Strategien stark mit dem Vorwissen zusammenhängt. Alle Theoretiker, aber nur einer der Experimentatoren verfügten über Vorerfahrungen beim Programmieren. Für ähnliche Ergebnisse in einer Computersimulationsumgebung sei auf die Untersuchung von van Joolingen und de Jong (1993) verwiesen.

In einem weiteren Experiment zum SDDS-Modell untersuchten Schunn und Klahr (2000) wissenschaftliches Entdecken in einem Gegenstandsbereich, der komplexer als Big Trak angelegt ist. Hier zeigten sich auch neue Heuristiken für die Suche im Experimenterraum. Als ausgeprägtes Muster fand sich, daß erfolgreiche Versuchspersonen zunächst mit einfachen Experimenten begannen und die Komplexität ihrer Experimente schrittweise steigerten. Auch Hypothesen entwickelten die Versuchspersonen in mehreren Schritten.

Thagard (1998a) kritisierte das Modell SDDS (Klahr & Dunbar, 1988; Klahr, 2000) dahingehend, daß der sehr klar strukturierte Gegenstandsbereich Big Trak wenig mit der Komplexität wissenschaftlicher Fragestellungen gemein hat. Dieses

Modell vom wissenschaftlichen Entdecken als Problemlösen mit den beiden Problemräumen Hypothesenraum und Experimenterraum funktioniert laut Thagard nur, wo Repräsentationen für die Hypothesen und Paradigmen für die Experimente bereits gegeben sind. Das Modell erklärt nicht, wie neue Repräsentationen und Experimente entwickelt werden können oder neue wissenschaftliche Probleme aufgeworfen werden, was jeweils von großer Bedeutung für den wissenschaftlichen Fortschritt ist. Thagard schlägt eine Erweiterung des Modells vor, in der Wissenschaft auch als Fragen, blinde Variation oder glücklicher Zufall betrachtet werden kann.

Auch Schunn und Klahr (2000) erkannten, daß das SDDS-Modell wissenschaftliches Entdecken nicht vollständig erklären kann (siehe auch van Joolingen und de Jong, 1993) und führten einen weiteren Problemraum ein, den Datenrepräsentationsraum. Dort findet eine Suche statt, wenn Entscheidungen etwa über die zu messenden abhängigen Variablen oder eine Darstellung der Daten als Tabelle oder Graphik anstehen. Klahr (2000) wirft die Frage auf, wieviele Problemräume für das wissenschaftliche Problemlösen theoretisch zu rechtfertigen wären und kommt zu dem Schluß, daß dies nur im konkreten Fall über die Interaktion des Problemlösers mit dem Gegenstandsbereich, dem Entdeckungskontext, zu entscheiden sei. Als Kriterien für die Einführung neuer Problemräume führt er Logik, Empirie und Implementationsmöglichkeiten an. Diese müßten gewährleisten, daß die Problemräume zweifelsfrei voneinander unterschieden werden können. Auf die Vorschläge von Thagard gehen die Autoren jedoch nicht ein. Sie würden auch Fragen und blinde Variation unter dem Begriff der Suche subsumieren.

Grundsätzliche Kritik aus einer anderen Richtung übt Neuweg (1999) an der kognitionswissenschaftlichen Herangehensweise. Er diskutiert die Explizitheit der dargestellten Modelle: Beim wissenschaftlichen Denken handelt es sich seiner Ansicht nach zu großen Teilen auch um implizites Wissen. Der Begriff implizites Wissen ist allerdings keineswegs klar definiert. Neuweg beschreibt implizites Wissen als unbewußte Verhaltenssteuerung, implizites Gedächtnis (z.B. Sequenzlernen), implizites Regelwissen, das nicht verbalisierbar, nicht formalisierbar, aber erfahrungsgebunden ist. Generell lassen sich mentale Prozesse seiner Ansicht nach nur erschließen und sind nicht introspektiv erforschbar. Darüber hinaus gebe es

sehr viel Forschung darüber, wie Wissenschaftlerinnen und Wissenschaftler Hypothesen testen können, aber es sei nur wenig über die impliziten Prozesse bekannt, wie sie zu ihren Hypothesen kommen. Damit wendet sich Neuweg gegen die vorgestellten Forschungsmethoden wie das laute Denken und die Modelle, die wissenschaftliches Entdecken vor dem theoretischen Hintergrund des Problemlösens erklären wollen.

Gleichzeitig betrachtet Neuweg aber auch die Forschung zum impliziten Wissen kritisch und merkt an, daß die Ergebnisse in diesem Bereich bisher ausgesprochen heterogen sind und einem theoriebildenden Anspruch nicht genügen. Weder der Anteil impliziten Wissens am komplexen Problemlösen in künstlichen Laborumgebungen ist geklärt, noch ist bisher bekannt, ob in realen Problemlösenszenarien nicht noch mehr implizites Wissen angewendet wird. Neuweg zeigt eine Reihe prinzipieller Probleme bei der Erforschung impliziten Wissens auf, Lösungsvorschläge und neue Impulse für dieses Forschungsgebiet bleibt er jedoch schuldig. Das in diesem Zusammenhang auftretende Meßproblem, nämlich wie der Einsatz und auch der Zuwachs impliziten Wissens beim wissenschaftlichen Problemlösen nachgewiesen werden kann, wird in einem späteren Kapitel nochmals aufgegriffen.

Betrachtet man die berichteten Forschungsansätze im Überblick, so scheinen folgende Aspekte für das wissenschaftliche Denken von besonderer Bedeutung zu sein: Vorwissen bestimmt sowohl die Qualität von Hypothesenraum und Experimenterraum und damit mögliche Heuristiken für das wissenschaftliche Entdecken, wie das Heranziehen von Analogien, als auch implizite Prozesse bei der Bewertung von Hypothesen und beim allgemeinen experimentellen Vorgehen. Es besteht eine grundsätzliche Tendenz, Hypothesen beizubehalten, indem einerseits Experimente so angelegt werden, daß sie diese Hypothesen gar nicht widerlegen können, und andererseits trotz widerlegender empirischer Evidenz an einmal aufgestellten Hypothesen festgehalten wird. Dennoch ist hypothesengeleitetes Experimentieren effizienter als eine atheoretische Suche im Experimenterraum. Bedeutenden Einfluß auf wissenschaftliches Denken hat auch das Setting, in dem sich die Forschenden bewegen. Dies gilt insbesondere für Artefakte, die bei der Forschung eingesetzt werden, und für soziale Aspekte, denn kaum ein Projekt kann von einer

einzigsten Person betrieben werden. Ein weiterer Aspekt, der sich auf dem theoretischen Hintergrund des Problemlösens abbilden läßt, ist derjenige des Erlernens von Fertigkeiten für das wissenschaftliche Problemlösen. Dies soll Gegenstand des nächsten Kapitels sein.

3

Erlernen von Fertigkeiten für das wissenschaftliche Problemlösen

Sollen Fertigkeiten vermittelt werden, so ist dafür deren genaue Analyse eine bedeutende Voraussetzung. Diese kann auf verschiedenen Wegen erfolgen. Eine Möglichkeit besteht in einer theoretischen Analyse, beispielsweise in Form der oben dargestellten wissensbasierten Systeme (etwa Kulkarni & Simon, 1988). Bei einem wenig formalen Gegenstandsbereich, wie es das Vorgehen beim wissenschaftlichen Problemlösen teilweise ist, bringt aber auch die Beobachtung von Experten auf diesem Gebiet im Vergleich mit Novizen Aufschluß über die Qualität der dazu benötigten Fertigkeiten und über die bei den Novizen vorhandenen Schwierigkeiten oder Wissenslücken.

In diesem Kapitel geht zuerst darum, die Ausgangslage der Lernenden zu klären, ohne daß diese didaktisch unterstützt wurden. Zunächst erfolgt eine Betrachtung zur Unterscheidung von Personen, die bereits Fertigkeiten erworben haben, also Experten, und solchen Personen, die noch kaum die erwähnten Fertigkeiten erwerben konnten, sogenannten Novizen. Eine Darstellung verschiedener Untersuchungen zum wissenschaftlichen Denken bei Kindern schließt sich an. Daraus lassen sich Schlußfolgerungen ableiten, wie Fertigkeiten für das wissenschaftliche Problemlösen am besten zu vermitteln wären. Hier bieten sich zwei Ansätze an: situiertes Lernen und entdeckendes Lernen, die im weiteren Verlauf des nächsten Kapitels beschrieben werden.

3.1 Novizen beim wissenschaftlichen Problemlösen

Die Expertiseforschung zu anderen Gegenstandsbereichen konnte einige grundsätzliche Charakteristika von Expertenwissen identifizieren. Deutlich werden die Qualitäten des Expertenwissens häufig durch den Vergleich mit Anfängern in einem Gebiet, sogenannten Novizen. Einen Überblick über verschiedene Ansätze

geben Ericsson und Smith (1991). Als Experten werden Personen bezeichnet, die für ihren Gegenstandsbereich dauerhaft Herausragendes leisten, und zwar aufgrund stabiler ererbter oder erworbener Persönlichkeitseigenschaften. Für die Expertiseforschung sind die erworbenen Fertigkeiten von besonderem Interesse. Zur Erlangung von Expertise ist eine mindestens zehnjährige Lern- und Trainingsphase notwendig (Ericsson & Smith, 1991). Sowohl was die Wissensrepräsentation als auch was die Herangehensweise an Probleme anbelangt, unterscheiden sich Novizen und Experten (etwa Boshuizen & Schmidt, 1992; Chase & Simon, 1973; Chi, Feltovich & Glaser, 1981; Hmelo-Silver & Pfeffer, 2004). Gestützt auf verschiedene Untersuchungen, faßt VanLehn (1989) zusammen, daß Experten im Vergleich zu Novizen beim Problemlösen schneller und genauer sind, die Schwierigkeit einer Aufgabe besser einschätzen können und ihr Vorgehen häufiger reflektieren. Beide Gruppierungen nutzen abhängig vom Gegenstandsbereich verschiedene Strategien: Beim Programmieren gehen Experten zunächst in die Breite und planen ihre Problemlösung, während Novizen gleich anfangen zu programmieren, also erst in die Tiefe gehen. In der Physik zeigen Experten vorwärtsverkettendes, Novizen rückwärtsverkettendes Vorgehen. Experten haben effizientere Wissensstrukturen ausgebildet. Einerseits ist das Wissen zu größeren Einheiten zusammengefaßt, andererseits sind die Klassifikation von und die Verbindungen zwischen den einzelnen Wissens-elementen für das Lösen von Problemen im entsprechenden Gegenstandsbereich optimiert. Dies ermöglicht auch einen effizienteren Abruf des Wissens.

Da sich doch bedeutende Unterschiede zwischen den verschiedenen Gegenstandsbereichen zeigen, stellt sich die Frage, inwiefern sich diese Befunde auf das wissenschaftliche Problemlösen und Entdecken übertragen lassen. Voss, Tyler und Yengo (1983) untersuchten Novizen und Experten beim Lösen eines sozialwissenschaftlichen Problems und fanden, daß Experten viel Zeit darauf verwendeten, eine Problemrepräsentation zu entwickeln, während Novizen direkt nach Lösungen für einzelne Ursachen des Problems suchten. Experten erarbeiteten wenige, dafür abstrakte Lösungsansätze, während Novizen viele konkrete Lösungen vorschlugen.

Einen experimentellen Vergleich von Experten und Novizen stellte Schraagen (1993) an und legte den Fokus seiner Untersuchung nicht so sehr auf die Struktur des Wissens, das beim Problemlösen zur Anwendung kommt, sondern vor allem auf die Strategien, die seine Versuchspersonen beim Problemlösen anwendeten. Im Zentrum der Untersuchung standen Experten für experimentelles Design, die sich jedoch nicht auskannten mit dem in Frage stehenden Gegenstandsbereich, der Sinnespsychologie. Ihnen wurden drei Gruppen gegenübergestellt, eine Gruppe von Experten für sinnespsychologische Experimente, und zwei Gruppen von Personen ohne Wissen über den Gegenstandsbereich, Novizen und Fortgeschrittene, in diesem Fall Masterstudenten. Alle Gruppen bekamen dieselbe Versuchsplanungsaufgabe und sollten Experimente zu einer vorgegebenen sinnespsychologischen Fragestellung ausarbeiten. Erwartungsgemäß wiesen die Lösungen der Experten für Sinnespsychologie mit Abstand die höchste Qualität auf, gefolgt in dieser Reihenfolge von denjenigen der Designexperten, der Masterstudenten und der Novizen. Schraagen identifizierte eine sogenannte Kontrollstruktur, eine Reihenfolge von Schritten für die Versuchsplanung, wie sie von Experten eingehalten wird: (1) Problem verstehen, (2) Paradigma auswählen, (3) Paradigma verfolgen, (4) Designprinzipien auswählen. Dieses Expertenmodell sieht vor, daß nur zwischen zwei direkt aneinander grenzenden Schritten hin- und hergesprungen werden darf, einzelne Schritte dürfen nicht übersprungen werden. Die Protokolle der Novizen glichen eher einem Modell mit zufälliger Reihenfolge der Schritte, während die Protokolle der beiden Expertengruppen dem Expertenmodell glichen. Die Protokolle der Masterstudenten entsprachen keinem der beiden Modelle.

Es ließen sich zwei Strategien für die Spezifikation eines Experiments herauskristallisieren: zunehmende, schrittweise Vertiefung und mentale Simulation. Schrittweise Vertiefung bedeutet, daß nach Auswahl eines Paradigmas der Versuchsplan nach und nach spezifiziert wird, zunächst etwa die unabhängigen Variablen, dann die Kontrollvariablen, danach die Stichprobengröße. Mit mentaler Simulation ist gemeint, daß die Versuchspersonen sich ein Paradigma vorstellen, dadurch Probleme ihres Versuchsplans aufdecken und so zu weiteren Spezifikationen ihres Versuchsplans gelangen. Alle Fortgeschrittenen und Designexperten

benutzten die Strategie der schrittweisen Vertiefung und der mentalen Simulation. Experten der Sinnespsychologie benutzen die beiden Strategien wesentlich seltener, stattdessen konnten sie die gefragten Versuchspläne aus dem Kopf direkt aufschreiben. Bei den Novizen ließen sich die beiden Strategien kaum nachweisen, was Schraagen (1993) auf eine kognitive Überforderung zurückführt. Insgesamt fanden sich bei den Designexperten Strategien von mittlerer Allgemeinheit, also Strategien, die zwar kein Wissen über den Gegenstandsbereich voraussetzen, jedoch wesentlich spezifischer für das wissenschaftliche Entdecken zugeschnitten sind, als allgemeine Heuristiken wie die Mittel-Ziel-Analyse.

Konnte Schraagen (1993) einige Charakteristika wissenschaftlichen Entdeckens bei Experten identifizieren, so wird aus der vorangegangenen Analyse doch deutlich, daß noch andere Prozesse daran beteiligt sein müssen. Denkbar sind hier die Entscheidungen über Messungen, die Darstellung von experimentellen Ergebnissen, die Interpretation von Ergebnissen oder das Generieren von Hypothesen. Daß dies in Schraagens Untersuchung keine Rolle spielte, lag auch daran, daß die Versuchspersonen Experimente nur planen, aber nicht durchführen konnten und daher kein Feedback über den Erfolg ihrer Ideen bekamen.

Hier setzten Schunn und Anderson (1999) ihr Experiment in einem simulierten wissenschaftlichen Labor an. Um bereichsübergreifende von bereichsspezifischen Fertigkeiten unterscheiden zu können, wurden vier verschiedene Gruppen untersucht: Novizen, anhand eines Vortests unterteilt in eine schwache und eine starke Gruppe, Experten für experimentelles Design im allgemeinen und Experten für experimentelles Design im Gegenstandsbereich, in diesem Fall in der Gedächtnispsychologie. Ihre Aufgabe war es, Experimente durchzuführen, die es ermöglichten, ein gedächtnispsychologisches Phänomen auf eine von zwei möglichen Theorien zurückzuführen, also eine Entscheidung zwischen diesen beiden Theorien zu fällen. Hatten die Versuchspersonen einen Versuchsplan mit bis zu vier variierten und konstantgehaltenen Faktoren definiert, so wurden sie dazu aufgefordert, für jede Zelle ihres Versuchsplans eine quantitative Vorhersage zu treffen. Anschließend generierte das System simulierte Daten und zeigte die Ergebnisse zusammen mit den Vorhersagen in Tabellenform an. Die Versuchspersonen konn-

ten dann neue Experimente planen und durchführen. Während des Experimentierens im simulierten Labor waren die Versuchspersonen zum lauten Denken instruiert.

Zunächst fielen einige Unterschiede bezüglich bereichsunabhängiger Fertigkeiten ins Auge. Während beide Expertengruppen die beiden zu prüfenden Theorien bereits während der Planung des ersten Experiments erwähnten, wurden die beiden Theorien während aller Experimente bei den starken Novizen deutlich seltener und bei den schwachen Novizen fast nie angesprochen. Diejenigen Novizen, die überhaupt die Theorien erwähnten, konzentrierten ihre Experimente dann auch auf die Variablen, deren Zusammenhang mit den Theorien jeweils inhaltlich am plausibelsten erscheinen mußte. Novizen, welche die Theorien nicht erwähnten, orientierten sich bei der Untersuchung der experimentellen Variablen an der Reihenfolge, in der diese auf dem Computerdisplay angezeigt wurden.

Eine bekannte Heuristik beim wissenschaftlichen Experimentieren ist es, die Experimente, soweit dies möglich ist, einfach zu halten. Dies erleichtert die Interpretation der Ergebnisse. Fanden sich bei den Experten für experimentelles Design nur wenige Zellen pro Versuchsplan, so wiesen die Versuchspläne der Novizen eine deutlich höhere Komplexität auf. Die kompliziertesten Versuchspläne entwarfen jedoch die gedächtnispsychologischen Experten. Sie konnten dabei allerdings auf ihre bereichsspezifischen Fertigkeiten vertrauen und hatten daher auch keine Schwierigkeiten bei der Interpretation der Ergebnisse.

Auch hinsichtlich der Auswahl der Faktorstufen bei den experimentell variierten Faktoren unterschieden sich die Gruppen. Bei der Auswahl von Faktorstufen müssen mehrere Aspekte beachtet werden. Die Stufen sollten sich so weit unterscheiden, daß eine Aussage über unterschiedliche Bedingungen möglich ist, und die Abstände zwischen den Stufen sollten jeweils gleich groß sein. Außerdem sollten die Manipulationen einfach gehalten werden, das heißt mindestens denselben Zahlenwert oder dieselbe Maßeinheit berücksichtigen. Die Novizen verletzten alle diese Regeln, wohingegen die Experten sich strikt daran hielten.

Zuletzt wurde als bereichsunabhängige Fertigkeit beobachtet, wie konsequent die Versuchspersonen die experimentellen Variablen über ihre Experimente hinweg konstant hielten. Diese Heuristik wurde in größerem Umfang nur von der Gruppe der schwachen Novizen mißachtet.

Bei der Vorhersage und Interpretation von Ergebnissen zeigten Novizen keine Schwierigkeiten bezüglich der Haupteffekte, um so mehr dafür im Umgang mit Interaktionen. Ihnen fiel es auch schwer, die Ergebnisse auf die beiden Theorien zu beziehen. Meist wurden die Theorien von ihnen, im Gegensatz zu den Experten, bei der Bewertung der Ergebnisse überhaupt nicht erwähnt. Stattdessen zogen die Novizen zur Erklärung ihrer Ergebnisse vermehrt nicht empirische Konzepte heran. Was die bereichsspezifischen Fertigkeiten anbelangt, so zeigten sich die Experten für Gedächtnispsychologie den anderen Gruppen durchweg überlegen.

Liegt der Schwerpunkt der vorgestellten Studien (Schraagen, 1993; Schunn & Anderson, 1999) in der Analyse von Expertenwissen, so interessiert für das Erlernen wissenschaftlichen Problemlösens doch besonders der Vergleich mit den Novizen. Für Novizen gilt die folgende Zusammenfassung: Sie verfügen nicht über die verschiedenen effizienten Heuristiken und Kontrollstrukturen der Experten, und Vorhersage und Interpretation komplexerer Ergebnisse fallen ihnen schwer. In diesen Aussagen zeigt sich eine Sicht auf Novizen, die vor allem deren Defizite im Vergleich mit Experten herausstellt. Um didaktische Maßnahmen zum Erlernen wissenschaftlichen Problemlösens entwickeln zu können, ist es aber notwendig, die Ansatzpunkte dafür möglichst genau zu kennen. Dazu zählen auch die bereits vorhandenen Fertigkeiten der Lernenden. Deshalb erfolgt an dieser Stelle ein Exkurs über die Entwicklung des wissenschaftlichen Entdeckens bei Kindern, bevor verschiedene didaktische Ansätze zum Erlernen des wissenschaftlichen Problemlösens vorgestellt werden.

3.2 Entwicklung wissenschaftlichen Denkens

Die Erforschung wissenschaftlichen Problemlösens bei Kindern kreist von Beginn an um die Frage, ob und, wenn ja, inwiefern sich Kinder in ihrem Denken von Erwachsenen oder gar Wissenschaftlern unterscheiden. Wegweisend für dieses Gebiet waren die Arbeiten Piagets, in denen er qualitativ verschiedene Stadien der

kognitiven Entwicklung von Kindern beschreibt (für einen kritischen Überblick siehe Carey, 1985). Carey leitet aus den Arbeiten Piagets fünf verschiedene Kriterien zur Beurteilung der Unterschiede im kindlichen und erwachsenen Denken ab, die gerade auch für das wissenschaftliche Problemlösen relevant sind: (1) die Datenstrukturen zur Repräsentation von Informationen und die Operationen zur Manipulation dieser Datenstrukturen, (2) metakonzeptuelles Wissen, (3) Grundkonzepte, die in vielen Gegenstandsbereichen gelten, z.B. Kausalität, (4) Werkzeuge mit breiter Anwendung und (5) bereichsspezifisches Wissen. Nach einer detaillierten Analyse mehrerer Experimente kommt Carey allerdings zu dem Schluß, daß der bedeutendste Unterschied zwischen den Denkprozessen von Kindern und Erwachsenen allein darin besteht, daß Kinder über ein sehr viel geringeres bereichsspezifisches Wissen verfügen als Erwachsene. In allen anderen Fällen seien sich Kinder und Erwachsene jedoch sehr ähnlich. Dies gelte auch für das Formulieren und Testen von Hypothesen. Damit stellt sich Carey gegen Piaget, der davon ausging, daß Kinder bis zum Alter von dreizehn Jahren keine Hypothesen aufstellten und auch die Logik nicht beherrschten, Hypothesen zu bestätigen oder zu widerlegen.

In derselben Richtung wie Carey (1985) argumentiert auch Karmiloff-Smith (1988). Sie konnte belegen, daß bereits vierjährige Kinder Theorien über die Welt haben und Experimente daran ausrichten. Sie stellte fest, daß schon Kinder dazu neigen, zunächst ihre Untersuchungsmethode in Zweifel zu ziehen und möglichst lange an ihrer Theorie festzuhalten, wenn die empirischen Ergebnisse ihrer Theorie widersprechen. Sie postuliert, daß sich Kinder bezüglich der Prozesse beim wissenschaftlichen Problemlösen, insbesondere bei der Formulierung von Theorien qualitativ nicht von Erwachsenen unterscheiden.

Die Argumente für qualitative Unterschiede zwischen Kindern und Erwachsenen beim wissenschaftlichen Denken greift Deanna Kuhn in ihren Arbeiten auf (Kuhn, Amsel & O'Loughlin, 1988; Kuhn, 1989). Sie beschreibt zwei Enden des Spektrums dessen, was wissenschaftliches Denken sein kann: Erkundungsverhalten kleiner Kinder versus hochprofessionelle Denkprozesse von Wissenschaftlern. In ihren Untersuchungen fällt es den Kindern schwer, zwischen Theorie und Evidenz zu unterscheiden. Stattdessen nehmen sie die empirischen Daten nur selektiv

wahr. Kuhn erklärt dies mit der Beobachtung, daß im Alltagsdenken Alternativen häufig nicht berücksichtigt werden, weil Theorie und Evidenz nicht unterschieden oder koordiniert werden. Das bedeutet, daß bei der Entwicklung wissenschaftlichen Denkens genau die Konzepte von Theorie und Evidenz stark restrukturiert werden müssen. In den empirischen Untersuchungen fiel es bereits Kindern leichter, Experimente zur Bestätigung eigener Theorien zu entwerfen, als Experimente zur Bestätigung von Theorien zu entwerfen, von denen sie nicht überzeugt waren. Außerdem hatten die Versuchspersonen weniger Schwierigkeiten, Experimente zu entwickeln, um die Korrektheit einer Theorie nachzuweisen, als Experimente zu entwickeln, um eine Theorie zu widerlegen. Nur Schüler mit Collegeausbildung waren zu letzterem durchgängig in der Lage.

Folgende Fertigkeiten sind Kuhn zufolge generell für die Koordination von Theorie und Evidenz von Bedeutung (Kuhn et al., 1988): über eine Theorie nachzudenken und nicht nur mit einer Theorie zu denken; die Evidenz getrennt von der Theorie zu repräsentieren und schließlich von der eigenen Überzeugung hinsichtlich einer Theorie abzusehen, um beurteilen zu können, was die Evidenz für die Theorie bedeutet. Zudem beobachtete Kuhn, daß Kinder nur bis zur 9. Klasse ihr wissenschaftliches Denken verbessern, danach aber kaum noch eine Entwicklung stattfindet, die dann gegebenenfalls stark vom Schultyp abhängig ist. Das bedeutet, daß auch Erwachsene, wenn sie nicht spezifisch ausgebildet wurden, Schwierigkeiten beim wissenschaftlichen Problemlösen bekommen. Die Autorin stellt sich ganz klar gegen die These, daß Kinder Wissenschaftler seien. Kinder verhielten sich in wesentlichen Punkten anders als Wissenschaftler.

Die These von der mangelnden Koordination von Theorie und empirischer Evidenz regte einige Forscherinnen und Forscher zu weiteren Untersuchungen an (Koslowski, 1996; Ruffman, Perner, Olson & Doherty, 1993; Sodian, Zaitchik & Carey, 1991; Schrempp & Sodian, 1999). Sodian et al. konnten nachweisen, daß bereits Zweitklässler Hypothesen richtig prüfen und die Ergebnisse der Prüfung richtig interpretieren können. Diese Fähigkeiten verbessern sich bei Viertklässlern und Erwachsenen nochmals. Die beiden letztgenannten Gruppen können auch spontan die richtigen Tests für Hypothesen vorschlagen, was Zweitklässlern noch schwerfällt. Bereits die Zweitklässler konnten aufschlußreiche und nichtssagende

Experimente voneinander unterscheiden. Diese Befunde zeigten sich sowohl in Experimenten mit vorwissensarmen als auch mit vorwissensreichen Problemen. Sodian folgert aus ihren Ergebnissen daher auch, daß sich metabegriffliche Kompetenzen weitgehend unabhängig vom inhaltlichen Vorwissen entwickeln.

Die Unterschiede zwischen Kindern, erwachsenen Laien und Wissenschaftlern scheinen je nach Perspektive und Definition essentieller Merkmale einmal größer und dann wieder kleiner auszufallen. Klahr, Dunbar und Fay (2000) erklären diese unterschiedlichen Befunde, indem sie diese in ein begriffliches Raster fassen. Ansätze, welche die Ähnlichkeit des wissenschaftlichen Denkens bei Kindern und Erwachsenen betonen, fokussieren auf das zum wissenschaftlichen Problemlösen notwendige bereichsspezifische Wissen, das bei Kindern noch nicht im gleichen Ausmaß vorhanden ist wie bei Erwachsenen. Betrachtet man die bereichsübergreifenden Fertigkeiten für das wissenschaftliche Problemlösen – etwa indem im Experiment Gegenstandsbereiche gewählt werden, die kaum bereichsspezifisches Vorwissen voraussetzen – so wird deutlich, daß sich Kinder und Erwachsene sehr wohl unterscheiden. Eine Entscheidung zwischen diesen beiden Sichtweisen ist nicht möglich, da bereichsspezifisches Wissen und bereichsübergreifende Fertigkeiten beim wissenschaftlichen Denken interagieren.

Um diese Interaktion zwischen bereichsspezifischem und -übergreifendem Wissen zu beleuchten, führten Klahr et al. (2000) weitere Experimente mit der künstlichen Experimentalumgebung Big Trak durch und verglichen dabei Kinder verschiedener Entwicklungsstufen mit Erwachsenen. Im Gegensatz zu Kuhn et al. (1988) konnten die Kinder hier zwischen Theorie und Evidenz unterscheiden. Leistungsunterschiede zu Erwachsenen konnten die Autoren mit Hilfe des SDDS-Modells erklären.

Im Hypothesenraum gehen Erwachsene und Kinder unterschiedlich mit plausiblen und unplausiblen Hypothesen um. Erwachsene überprüfen eine unplausible Hypothese, indem sie ein Experiment definieren, das zwischen der gegebenen unplausiblen Hypothese und einer eigenen plausiblen Alternativhypothese diskriminiert. Auch Kinder erkannten es, wenn eine vorgegebene Hypothese unplausibel war, und stellten selbst eine plausible Alternativhypothese auf. Ihre Experimente legten sie jedoch so an, daß sie eine Bestätigung für ihre plausible Alternativhypo-

these generieren sollten, wodurch die Daten nicht mehr informativ für die eigentliche Fragestellung waren. Im Experimenterraum hatten die Kinder größere Schwierigkeiten als Erwachsene, ihre Suche auf sinnvolle Experimente zu beschränken und bei der Versuchsplanung systematisch vorzugehen.

Zusammenfassend formulieren Klahr et al. (2000) folgende Bedingungen für erfolgreiches wissenschaftliches Problemlösen bei Kindern: Die Hypothesen sollten leicht zugänglich sein oder es sollte nur eine kleine Zahl an Hypothesen vorgegeben werden. Es sollten nur wenige experimentelle Alternativen angeboten werden, und der Gegenstandsbereich sollte aussagekräftige Rückmeldungen über die Unterscheidung zwischen plausiblen Hypothesen geben. Kinder verfügen sehr wohl über Fertigkeiten zur logischen Verifikation und Falsifikation von Hypothesen. Die Überlegenheit der Erwachsenen liegt in Fertigkeiten, die Suche im Hypothesen- und Experimenterraum zu koordinieren.

Die Debatte um die wissenschaftlichen Fertigkeiten von Kindern hat zu einer stetigen Verfeinerung der experimentellen Settings geführt. Dadurch konnten grundlegende Fertigkeiten des wissenschaftlichen Denkens für immer jüngere Kinder nachgewiesen werden. Zusammenfassend zeigt eine zunehmende Zahl an Untersuchungen, daß sich wissenschaftliches Denken eher kontinuierlich als stufenweise entwickelt und daß elementare Verständnisvoraussetzungen für wissenschaftliches Denken bereits früh erworben werden. Andererseits scheint ab dem neunten Schuljahr keine selbständige Weiterentwicklung dieser Fertigkeiten mehr stattzufinden. Nur mit entsprechend anspruchsvollem Unterricht erreichen die Schülerinnen und Schüler die Voraussetzungen für ein Universitätsstudium und selbst dann müssen sie, im Vergleich zu Wissenschaftlern, noch als Novizen bezeichnet werden. Dies zeigt die Notwendigkeit, den Erwerb von Fertigkeiten des wissenschaftlichen Problemlösens didaktisch zu fördern.

Aus der bisher dargestellten Forschung zum wissenschaftlichen Denken bei Kindern, Novizen und Experten lassen sich bereits etliche Empfehlungen für die Vermittlung der zugrundeliegenden Fertigkeiten ableiten. So sollten vor allem bereichsübergreifende Strategien und Kontrollstrukturen erlernt und eingeübt werden. In engem Zusammenhang damit steht eine Förderung metakognitiver Fertigkeiten (Kuhn et al, 1988; Schraagen, 1993). Dies sollte mit einer möglichst frühen

Einübung in die konkrete, wissenschaftliche Praxis und mit einer Verallgemeinerung durch Üben in verschiedenen Gegenstandsbereichen einhergehen. Günstig sind für diesen Zweck insbesondere kooperative beziehungsweise verteilte Settings. Es ist nicht möglich, Fertigkeiten wissenschaftlichen Denkens im abstrakten Unterricht zu erwerben (Dunbar, 2000b; Koslowski, 1996; Kuhn et al., 1988; Neuweg, 1999). Dunbar (2000b) empfiehlt des Weiteren, im Unterricht auf die Bedeutung unerwarteter Ergebnisse (siehe auch Masnick & Klahr, 2003) und die Argumentation mit weit hergeholt Analogien zu fokussieren, da gerade sie den wissenschaftlichen Fortschritt befördern (vergleiche Chinn & Brewer, 1993; Koslowski, 1996). Kuhn et al. (1988) äußern sich auch zur veränderten Rolle der Lehrerinnen und Lehrer. Da die genannten Strategien und metakognitiven Fertigkeiten nicht abstrakt vermittelt werden können und damit Übungen zu deren Vermittlung in den Vordergrund rücken, verschiebt sich die Aufgabe der Lehrenden in Richtung Fördern, Ermöglichen, Unterstützen bei der Lösung der realitätsnahen Übungsaufgaben.

Betrachtet man die Empfehlungen für die Unterstützung des Erlernens von Fertigkeiten für das wissenschaftliche Problemlösen aus dem vorangegangenen Abschnitt, so wird schnell klar, daß sich diese nicht im herkömmlichen Frontalunterricht umsetzen lassen, wie er in Schulen und auch Universitäten noch verbreitet ist. Stattdessen müssen andere didaktische Ansätze genutzt werden, um die nötigen Lernprozesse zu fördern. Hier sind zunächst Konzepte zu nennen, die Lernen als aktiven und konstruktiven Prozeß des Wissens- und Fertigkeitserwerbs betrachten, sogenannte konstruktivistische Theorien zum entdeckenden Lernen und zum situiereten Lernen.

3.3 Entdeckendes Lernen

Die Theorie des entdeckenden Lernens geht davon aus, daß Lernen am besten stattfindet, wenn sich die Lernenden selbst die zum Wissenserwerb nötigen Informationen suchen, sie transformieren und auf diese Weise das Wissen selber aufbauen beziehungsweise konstruieren. „Entdecken“ ist also nicht darauf beschränkt, etwas für die ganze Menschheit Neues herauszufinden, sondern schließt alle Formen des Wissenserwerbs durch Einsatz des eigenen Verstandes ein. Initiiert wurde dieser lerntheoretische Ansatz von Bruner (1961). Die von ihm erforschte Konzeptbil-

dung führte ihn nicht nur, wie oben dargestellt, zur Untersuchung des Hypothesentestens, er betrachtete sie auch als einen Lernprozeß zum Erwerb neuer Konzepte und Regeln. Schon bald wurde die Methode des entdeckenden Lernens auch auf komplexere Probleme angewandt. Aufgaben beim entdeckenden Lernen können sich von selbst ergeben – man denke an ein Kind, das beim Malen mit Wasserfarben etwas über subtraktive Farbmischungen lernt –, oder die Aufgaben werden von Lehrenden gezielt gestellt, so daß die Lernenden beim Entdecken auf das Lernziel hingeführt werden. Lernende sind bei dieser Art des Wissenserwerbs dazu aufgefordert, den Lernprozeß selbst zu organisieren. Selbständiges Problemlösen und Entdecken werden gefördert. Entdeckungen sind dabei immer vom Vorwissen abhängig: je ausgeprägter das Vorwissen, desto reicher die Möglichkeiten für Entdeckungen. Das Vorwissen wird beim Entdecken aktiviert, die neue Information mit dem eigenen Vorwissen in Verbindung gebracht und elaboriert. Dabei werden Verbindungen zum Vorwissen gezogen, und das Vorwissen wird restrukturiert (Pintrich, 2000). Auf diese Weise wird auch der Abruf des Gelernten verbessert. Bruner geht auch davon aus, daß Entdeckungsheuristiken durch eine Einübung in die entsprechenden Methoden erlernt werden und ein formales Verständnis dieser Heuristiken für deren Anwendung nicht ausreicht.

Ausubel, Novak und Hanesian (1981) äußern sich dem entdeckenden Lernen gegenüber ausgesprochen kritisch. Zu leicht verfestigten sich Mißkonzepte der Lernenden und würden nicht korrigiert. Zudem halten sie den Aufwand für ungerechtfertigt, den diese Form des Lernens insbesondere für die Übermittlung großer deklarativer Kenntnissysteme mit sich bringt. Dennoch empfehlen sie gerade entdeckendes Lernen für den Unterricht in wissenschaftlicher Methodik. Es fördere den Erwerb der Fertigkeiten beim Aufstellen und Testen von Hypothesen. Aktiv Lernende seien motivierter, könnten sich die den Lernstoff besser merken und das erworbene Wissen besser auf andere Bereiche übertragen. Auch hier schränken die Autoren jedoch ein, daß beim entdeckenden Lernen nicht so viel verbalisiert werde wie bei anderen Unterrichtsformen, verbalisiertes Wissen aber besser zu transferieren sei.

Reines Lernen durch Entdecken ist auch über die Kritik von Ausubel et al. (1981) hinaus mit Schwierigkeiten verbunden. Gerade bei komplexeren Lernaufgaben, wie sie im Bereich des wissenschaftlichen Denkens gestellt werden, sind Lernende schnell damit überfordert, das selbstgesteuerte Lernen und die Anforderungen der Aufgabe zu koordinieren (Glaser, Duschl, Schulze & John, 1995). Dies führte zur Entwicklung von Konzepten zur Unterstützung des entdeckenden Lernens, dem sogenannten *guided discovery learning*. So stellten Brown und Campione (1994) ein komplettes eigenes Curriculum zusammen. Die Schülerinnen und Schüler arbeiteten selbständig in Gruppen nach einem geregelten Ablauf an spezifischen Aufgaben und bildeten so eine Lerngemeinschaft, in der die Kommunikation zwischen den Lernenden einen hohen Stellenwert einnahm. Besondere Beachtung schenken die Autoren der Lehrerrolle. Durch die dezentrale Unterrichtsform wandelten sich die Aufgaben der Lehrenden. Sie vermittelten nicht mehr hauptsächlich den Stoff, sondern griffen nur ein, wenn dies für den Lernfortschritt notwendig war. Die größte Herausforderung liegt darin, zu erkennen, wann welche Interventionen angebracht sind. Denn gerade diese Interventionen machen neben der Aufgabenstellung und der Aufteilung in Gruppen die zusätzliche Qualität des *guided discovery learning* aus. Der Erfolg dieses Curriculums zeigte sich im Vergleich zu Kontrollgruppen sowohl im Hinblick auf bereichsspezifisches Wissen als auch bei allgemeinen Fertigkeiten wie dem Leseverständnis. Auch Damon und Phelps (1989) stellten fest, daß der Kooperation mit Gleichaltrigen ein stützender Effekt beim entdeckenden Lernen zukommt. Wie das entdeckende Lernen fördert auch das kooperative Lernen tiefe Einsichten und die Entwicklung anderer Perspektiven. Die Kooperation regt das Experimentieren mit neuen Ideen und die kritische Überprüfung eigener Gedanken an. Dem kooperativen Lernen ist daher das nächste Kapitel gewidmet.

Einen Nachweis für die Bedeutung der Aufgabenstellung beim entdeckenden Lernen erbrachte die Studie von Schauble, Klopfer und Raghavan (1991). Sollten Kinder zunächst versuchen, ein praktisches Problem zu lösen, und danach die Zusammenhänge dieses Problems wissenschaftlich experimentell untersuchen, so

explorierten sie einen größeren Teil des Experimenterraums und kamen zu besseren wissenschaftlichen Aussagen als Kinder, die zuerst die wissenschaftliche und dann die praktische Aufgabe lösen sollten.

Auf einer anderen Ebene setzten White und Frederiksen (1998) an. Ihr Unterrichtsziel war der Erwerb eines metakognitiven Modells von Forschung und der dazugehörigen Methode reflektierter Bewertung. Auch sie ließen die Lernenden während ihrer Untersuchung in Gruppen arbeiten. Um das entdeckende Lernen zu unterstützen, bekamen die Lernenden Kriterien an die Hand, mit deren Hilfe sie ihr eigenes Vorgehen beurteilen konnten (z.B. Verbindungen ziehen, systematisch vorgehen, sorgfältig argumentieren). Diese Beurteilungskriterien kamen nach jedem Experiment zyklisch zur Anwendung. Sie lieferten den Lernenden ein präzises Vokabular, um sich angemessen über ihre Aktivitäten auszutauschen. Auch diese Intervention wirkte sich positiv auf den Lernerfolg aus. Eine ähnliche Studie führten Kolodner, Gray und Burks Fasse (2003) durch.

Die Umsetzung entdeckenden Lernens in verschiedensten Gegenstandsbereichen wurde in der Darstellung bis hierher ausgeklammert. Der Aufwand dafür ist teilweise beträchtlich, in manchen Gegenstandsbereichen läßt sich entdeckendes Lernen in der Realität deshalb gar nicht verwirklichen. Einen Ausweg bieten hier computerbasierte Simulationen (etwa Blömeke, 2003; de Jong & van Joolingen, 1998). Wie sie sich für entdeckendes Lernen nutzen lassen und wie spezielle Werkzeuge diese Lernprozesse unterstützen können, ist Gegenstand von Kapitel 5. An dieser Stelle soll zunächst ein weiterer lerntheoretischer Ansatz vorgestellt werden, der des situierten Lernens.

3.4 Situiertes Lernen

Als Kontrast zur kognitionspsychologisch orientierten Lernforschung wurde der Ansatz des situierten Lernens konzipiert. Grundannahmen wurden der Theorie der *situated cognition* (Greeno, 1989) entlehnt: Denken und Lernen finden immer in einem physikalischen und sozialen Zusammenhang statt und können daher nicht nur als die gedankliche Aktivität einer einzelnen Person betrachtet werden. Denken und Lernen sind durch persönliche und soziale Epistemologien, also Vorstellungen über Kognition, bestimmt. Diese Vorstellungen können individuell, zwischen

Situationen und zwischen sozialen Gruppen unterschiedlich sein. Beim Denken und Lernen elaborieren und reorganisieren die Lernenden ihr Wissen und Verständnis aktiv, sie erwerben nicht nur kognitive Strukturen und Prozeduren.

Ein Ausgangspunkt für die Entwicklung des situierten Lernens waren unbefriedigende Ergebnisse beim Transfer von theoretischem Wissen in die Praxis, was als träges Wissen bezeichnet wurde (Gruber, Mandl & Renkl, 2000). Wissen, das in Schule oder Universität erworben wurde, kann oft nicht sinnvoll außerhalb dieser Institutionen angewendet werden. Als Hauptursache dafür machen Gruber et al. ungeeignete Lernsituationen aus: Viele Fertigkeiten können nicht durch theoretischen Unterricht erworben werden, beispielsweise muß das Führen eines Anamnesegesprächs in der sozialen Interaktion eingeübt werden, diagnostisches Wissen um Krankheitsbilder allein reicht nicht aus. Solche Lernsituationen sind geradezu als paradox zu bezeichnen, da die Lernenden die Situation, für die sie sich ihr theoretisches Wissen aneignen, gar nicht kennen und den Lernprozeß nicht darauf abstimmen können.

Eine Lösung für die genannten Schwierigkeiten versprach man sich vom situierten Lernen. Vertreter dieses Ansatzes argumentieren, daß Lernen und Denken immer an das physikalische und soziale Umfeld gebunden seien und jede Bedeutungskonstruktion sozial und kontextabhängig sei. Daher kann Lernen auch nur im entsprechenden praktischen Umfeld stattfinden, dort wo die Fertigkeiten später zur Anwendung kommen sollen. Um diese Ideen umzusetzen, greifen Brown, Collins und Duguid (1989) sowie Lave und Wenger (1991) die Idee der Lehre auf und erweitern sie zur sogenannten *cognitive apprenticeship*, der kognitiven Lehre. Sie sehen Lernen nicht mehr als abgrenzbare Aktivität, sondern als integralen Bestandteil der sozialen Praxis. Lernende nehmen an der sozialen Praxis teil, zunächst peripher, nachahmend und angeleitet, und mit zunehmender Beherrschung der nötigen Fertigkeiten übernehmen sie mehr Aufgaben und nehmen schließlich voll an der Gemeinschaft der Praktiker (*community of practice*) teil. Lernen wird also als Enkulturationsprozeß betrachtet. Durch die Interaktion mit verschiedenen etablierten Mitgliedern der Gemeinschaft und deren Anerkennung wird die Motivation der Lernenden gefördert. Gleichzeitig übernehmen die Lernenden immer wieder neue Rollen und lernen so den Gegenstandsbereich aus

unterschiedlichen Perspektiven kennen, was die Abstraktion des Wissens fördert. Abstraktes Wissen kann dieser Theorie zufolge nur durch Generalisierung aus dem konkreten Tun aufgebaut werden. So erwerben die Lernenden auch ein der Lehr- und Anwendungssituation angemessenes Gesprächsverhalten, anders als während des einseitig lehrerzentrierten Diskurses im Frontalunterricht.

Erscheint der beschriebene Ansatz vor allem als Lernszenario für praktische Tätigkeiten geeignet, so machen Collins, Brown und Newman (1989) Vorschläge, wie das Konzept der Lehre auch für kognitive Prozesse umgesetzt werden kann. Konzeptuelles Wissen wird beim Problemlösen, bei der Ausführung von Aufgaben beispielhaft und in den verschiedensten Zusammenhängen angewendet und dargeboten. Dies fördert ein tieferes Verständnis der Bedeutung dieser Konzepte und schafft gleichzeitig reichhaltige Assoziationen zwischen den Konzepten, was wiederum den Abruf erleichtert.

Im Gegensatz zu traditionellen Gegenstandsbereichen für den Wissenserwerb in der Lehre, ist die Ausführung von Fertigkeiten bei intellektuellen Aufgaben nicht direkt sichtbar, weder für Lehrende noch für Lernende, was die Beobachtung, aber auch die Korrektur und Verbesserung für beide Seiten erschwert. Daher legen Collins et al. (1989) besonderen Wert auf die Externalisierung dieser normalerweise unsichtbaren Prozesse. Als Beispiel führen sie unter anderem *reciprocal teaching* an (Palincsar & Brown, 1984). Hier sollen Kinder ihr Textverstehen trainieren, indem sie abwechselnd Fragen an den Text stellen, Zusammenfassungen formulieren und Begriffe klären. Werden diese Aufgaben zunächst von Lehrerin oder Lehrer modelliert, so übernehmen die Lernenden nach und nach selbst die Rollen. Die Lehrenden blenden also ihre Unterstützung bei komplexen Aufgaben schrittweise aus. Collins et al. identifizierten hier folgende für das situierte Lernen förderliche Prozesse: Modellieren und Artikulieren der angestrebten Fertigkeit auch durch die Lernenden; Trainieren – durch Beobachtung rechtzeitig Unterstützung und Rückmeldung geben, diese langsam ausblenden, angemessene Aufgaben stellen; und Reflexion, um eigene Fortschritte einschätzen und Fehler selbst korrigieren zu können.

Die Kritik der Vertreter des situierten Lernens an den kognitionspsychologischen Lerntheorien blieb nicht unbeantwortet. So wandten sich Anderson, Reder und Simon (1996) anhand empirischer Beispiele gegen die Postulate, daß Lernen immer vom physikalischen und sozialen Kontext abhängig sei, Wissen nicht von einer Situation auf die andere transferierbar und die Lehre abstrakter Fertigkeiten daher nicht sinnvoll sei. Die weitere Diskussion (Anderson, Greeno, Reder & Simon, 2000) führte wieder zu einer Annäherung beider Positionen: Sowohl individuelle als auch soziale Aspekte sind für das Lernen von Bedeutung, auch wenn die beiden Ansätze bei ihren Untersuchungen unterschiedliche Schwerpunkte setzten. Es kann sinnvoll sein, Wissens Elemente auch ohne Bezug zur Anwendung zu vermitteln, weil sie auf andere Anwendungen übertragen werden können. Allgemein werden die praktischen Anwendungen im Unterricht bisher aber zu wenig beachtet. Gruber et al. (2000) stellen fest, daß für den Erwerb inhaltlichen Wissens Formen der direkten Instruktion geeignet sind, während sich Anwendungswissen besser durch situiertes Lernen vermitteln läßt. Die Komplexität des Settings im situierten Lernen birgt jedoch die Gefahr der Überforderung in sich und erfordert eine geeignete didaktische Unterstützung.

Ein expliziter Bezug zum Erwerb von Fertigkeiten für das wissenschaftliche Problemlösen wird in keiner der bisher vorgestellten Arbeiten zum situierten Lernen hergestellt. Dennoch erscheinen dessen Konzepte gerade für die Wissensvermittlung im genannten Gegenstandsbereich sinnvoll, denn auch hier treten Transferprobleme auf, auch hier geht es um den Erwerb von Problemlösefertigkeiten, das Erlernen einer Sprache, die Enkulturation in die Gemeinschaft der Praktiker, in diesem Fall in die Wissenschaftsgemeinde. Stark und Mandl (2000) schildern die Probleme der Methodenlehre im Psychologiestudium und fanden in einer Umfrage auch lernhinderliche Einstellungen der Studierenden zu diesem Fach. Sie stellen ein auf Konzepten des situierten Lernens basierendes Curriculum für dieses Fach vor und betonen vor allem dessen positive motivationale Aspekte. Eine empirische Evaluation des Programms bleiben die Autoren aber schuldig.

Zusammenfassend lassen sich zwischen entdeckendem Lernen und situiertem Lernen einige Ähnlichkeiten feststellen, obwohl die Vertreter der beiden Ansätze aus unterschiedlichen Perspektiven argumentieren. Während das entdek-

kende Lernen vor dem Hintergrund kognitiver Lerntheorien zu sehen ist, entstand das situierte Lernen gerade als Gegenvorschlag dazu. Ähnlich wie beim entdeckenden Lernen wird aber auch beim situierten Lernen die aktive Konstruktion von Wissen in einer realistischen Lernumgebung betont. Auch für das situierte Lernen wurden computerbasierte Lernprogramme mit möglichst realistischen Szenarien entwickelt (Gruber, Law, Mandl & Renkl, 1996). Diese Programme arbeiten mit möglichst vielfältigen Wissensrepräsentationen und beziehen auch soziale Interaktion in das Lernkonzept mit ein. Gruber et al. betonen jedoch, daß diese computerunterstützten Lernprogramme eben gerade kein praktisches Setting böten und daher als Brücke zwischen dem instruktionalen Schulunterricht und der Praxis zu sehen seien.

Beide Theorien sehen eine gewandelte Rolle der Lehrenden, weg von der allwissenden Person, die den Lernenden das Wissen auf eine optimal aufbereitete Weise präsentiert, hin zu einer Moderatorenrolle, einem Vorbild für die soziale Praxis in dem Bereich, der erlernt werden soll. Der Ansatz des situierten Lernens betont darüber hinaus die Bedeutung sozialer Interaktion beim Lernen. Neben den Lehrenden wird daher auch die Rolle der *peers*, der Gleichaltrigen, Mitschülerinnen und Mitschüler in das Lernkonzept einbezogen. Als ein Beispiel dafür wurde bereits das *reciprocal teaching* (Palincsar & Brown, 1984) genannt.

Die Entwicklung des situierten Lernens führte allgemein zu einer größeren Aufmerksamkeit für die Prozesse und Vorteile des kooperativen Lernens. Da kooperative Prozesse auch einen zentralen Bestandteil wissenschaftlichen Arbeitens ausmachen, scheint es sinnvoll, diese schon bei der Vermittlung von Fertigkeiten für das wissenschaftliche Problemlösen mit einzubeziehen. Die Betrachtung kooperativen Lernens ist Gegenstand des nächsten Kapitels.

4

Kooperatives Lernen

Lernen kann in den unterschiedlichsten sozialen Situationen stattfinden, in speziell dafür vorgesehenen Institutionen wie Schule und Universität genauso wie außerhalb; mit Eltern, Lehrern, Tutoren oder Gleichaltrigen (*peers*); in der Klasse oder in kleinen Gruppen. Mit der Betonung der aktiven Konstruktion von Wissen beim Lernen rückte die Kooperation von Gleichaltrigen ins Zentrum des Interesses. Diese ermöglicht es mehreren Personen gleichzeitig, an einem aktiven Lernprozess teilzunehmen. Roschelle und Teasley (1995) definieren Kooperation wie folgt:

*Collaboration is a coordinated, synchronous activity, that is the result of a continued attempt to construct and maintain a shared conception of a problem.*¹

Die Autoren grenzen hierbei „collaboration“ von „cooperation“, also kooperatives von verteiltem Problemlösen ab. Kooperatives Problemlösen besteht darin, eine Aufgabe durch gemeinsame Anstrengungen zu lösen, die in einer übereinstimmenden Auffassung von dem zu lösenden Problem mündet. Beim verteilten Problemlösen übernehmen die Beteiligten Teilaufgaben, lösen diese selbständig und tragen die Teillösungen schließlich zusammen, wobei nicht in jedem Fall ein gemeinsames Problemverständnis aufgebaut wird. Als Grundvoraussetzung für Kooperation zwischen zwei oder mehr Personen nennt Piaget (1976) die Gleichberechtigung zwischen den Teilnehmenden in einer nicht von Wettbewerb geprägten Situation. Baker (1994) betont darüber hinaus die Bedeutung eines gemeinsamen Ziels beim kooperativen Lernen und einer grundsätzlichen Bereitschaft zur Kooperation.

1. „Collaboration“ wird in dieser Arbeit mit „Kooperation“ übersetzt, da der Begriff „Kollaboration“ im Deutschen eine abwertende politische Konnotation hat. „Cooperative“ wird im Unterschied dazu und mit „verteilt“ übersetzt.

Im Folgenden wird es um Lernen beim kooperativen Problemlösen gehen. Die Darstellung soll zeigen, inwieweit sich kooperatives Lernen für den Erwerb von Fertigkeiten für das wissenschaftliche Problemlösen besonders eignet. Dabei werden zunächst theoretische Grundlagen zum kooperativen Lernen dargestellt, soweit sie nicht bereits abgehandelt wurden. Daran schließt sich eine Beschreibung der empirischen Einflußfaktoren auf das kooperative Lernen und der Prozesse während des kooperativen Lernens an. Den Abschluß bilden empirische Arbeiten zum kooperativen Lernen mit besonderem Bezug zum wissenschaftlichen Problemlösen.

4.1 Pädagogisch-psychologische Theorien

Einige grundlegende Überlegungen zum kooperativen Lernen gehen bereits auf den sozio-kulturellen Ansatz von Vygotsky (1962, 1978) zurück. Er postuliert, daß kulturelle Entwicklungsschritte, insbesondere der Gebrauch von Zeichen, immer zunächst in der sozialen Interaktion, also auf der inter-psychischen Ebene, auftreten und dann erst durch Internalisierung auf der intra-psychischen Ebene übernommen werden. So entwickelt sich auch das für das wissenschaftliche Problemlösen zentrale logische Argumentieren zunächst zwischen Kindern oder in der Interaktion mit Erwachsenen. Erst später wird dieses als Reflexion internalisiert (Wertsch, 1985). Der Bereich, innerhalb dessen der Übergang von inter-psychischen zu intra-psychischen Funktionen möglich ist, wird als Zone nächster Entwicklung (*zone of proximal development*) bezeichnet. Es handelt sich um den Bereich zwischen dem, was eine Person allein an Problemlösen leisten kann und dem, was sie mit Hilfe anderer, gegebenenfalls weiter fortgeschrittener Personen meistern kann.

Vygotsky (1962, 1978) betrachtet in seinen Arbeiten vorrangig die Interaktion von Kindern mit ihren Eltern. Dennoch lassen sich seine Konzepte auf das kooperative Lernen zwischen Gleichaltrigen übertragen (Hogan & Tudge, 1999). Sowohl die Beobachtung, daß kulturelle Entwicklung zuerst in der Interaktion stattfindet, als auch das Konzept der Zone nächster Entwicklung können auch zwischen Gleichaltrigen zum Tragen kommen: Sofern ihr Wissensstand und ihre Meinungen zu einem Gegenstandsbereich leicht differieren, was selbst bei Schülerinnen derselben Klasse wahrscheinlich ist, können auch Gleichaltrige im Sinne Vygotskys gemeinsam und voneinander lernen.

Untersuchungen dazu stammen unter anderem von Rogoff (1991). Sie verglich Lernprozesse von Kindern in Interaktion mit Erwachsenen, mit trainierten Gleichaltrigen sowie mit untrainierten Gleichaltrigen und konnte Vorteile für das Lernen mit Erwachsenen und in eingeschränkter Weise auch für das Lernen mit trainierten Gleichaltrigen feststellen. Dabei sind die Lernerfolge sowohl vom Alter der Kinder als auch vom Gegenstandsbereich abhängig. Unabhängig davon, mit wem die Kinder interagierten, zeigte sich, daß sie um so erfolgreicher lernten, wenn die Vorgehensweise beim Problemlösen explizit verbalisiert wurde und gleichzeitig die lernenden Kinder aktiv in den Problemlöseprozeß einbezogen waren (siehe auch Bielaczyc, Pirolli & Brown, 1994). Rogoff betont, daß die Interaktion zwischen Gleichaltrigen in anderen Gegenstandsbereichen wesentlich fruchtbarer sei als bei den von ihr untersuchten räumlichen Planungsaufgaben. Sie nennt diesbezüglich vor allem typische Gegenstandsbereiche für die Interaktion Gleichaltriger wie Spiel und Exploration.

Auch Piaget (1976) sah in der sozialen Interaktion eine notwendige Voraussetzung für die Entwicklung des logischen Denkens. Sie ermöglicht die Entwicklung von Objektivität, Reflexivität und Selbstaufmerksamkeit, die wichtige Voraussetzungen für Lernen und Problemlösen darstellen. Doise und Mugny (1984) vertieften ausgehend von Piagets Ansätzen die Überlegungen zur Bedeutung von Kooperation für kognitive Lernprozesse. Während die sozio-konstruktivistische Perspektive von Vygotsky und seinen Schülern die Internalisierung sozialer Interaktion als integralen Bestandteil von Lernprozessen betrachtet, sahen Doise und Mugny als Vertreter der sozio-kognitiven Schule soziale Interaktion vor allem als Anstoß individueller Lernprozesse. Der Fokus der Betrachtung liegt also bei ersterem Ansatz auf der sozialen Interaktion an sich, wohingegen letzterer sich auf das Individuum konzentriert und soziale Interaktion als einen Impuls zu individuellen Lernprozessen sieht. Kooperation ermöglicht Situationen, in denen neue Kognitionen möglich sind, sie zwingt zur Koordination unterschiedlicher Perspektiven. Als zentraler Mechanismus hierfür wird der kognitive Konflikt angenommen. Er führt auf dreierlei Weise zur Anregung von Lernprozessen: durch explizite Darstellung alternativer Antworten, durch Hinweisreize zur Elaboration neuer Konzepte und durch die kognitive Aktivierung in der Interaktion.

Doise und Mugny (1984) berichten eine Serie von Experimenten, in denen sie untersuchten, auf welcher Entwicklungsstufe welche Form der Interaktion, insbesondere welche Gestaltung kognitiver Konflikte, Lernprozesse besonders fördert. Sie konnten für verschiedene Versuchsbedingungen bestätigen, daß kooperierende Dyaden bessere Lernerfolge erreichen als Einzelpersonen. Die Kooperation ermöglichte es den Versuchspersonen tatsächlich, Aufgaben zu meistern, die sie alleine erst deutlich später hätten erfüllen können. Ohne ausreichendes Vorwissen bei den Teilnehmenden war eine sinnvolle Kooperation jedoch nicht möglich. Auch wenn ein kognitiver Konflikt mit einer hierarchisch höhergestellten Person induziert wurde, führte dies zwar zur Übernahme von deren Überzeugungen, aber nicht zu einsichtigem Lernen und daher auch nicht zu Transferleistungen. Auch in Gruppen, in denen ein kognitiver Konflikt nicht experimentell induziert wurde, konnte er häufig beobachtet werden und führte ebenso wie ein induzierter kognitiver Konflikt zu einem höheren Lernerfolg als in Gruppen, in denen spontan kein kognitiver Konflikt auftrat. Graduell stärkere kognitive Konflikte führten auch zu einem höheren Lernerfolg als schwächere.

Vergleicht man den sozio-kognitiven und den sozio-konstruktivistischen Ansätze, so fällt auf, daß deren Schwerpunkte nicht nur, wie bereits dargestellt, inhaltlich einerseits auf individuelle Lernprozesse und andererseits auf die soziale Interaktion fokussieren, sondern es wird auch deutlich, daß dafür unterschiedliche methodische Herangehensweisen vonnöten sind. Während sich sozio-kognitive Fragestellungen vielfach mit Hilfe experimenteller Vortest-Nachtest-Versuchspläne untersuchen lassen, ist für sozio-konstruktivistische Studien eine eingehendere Analyse der Interaktionsprozesse unerläßlich. Da sich für beide Ansätze Belege finden lassen und es auch nicht nötig ist, sich zwischen beiden zu entscheiden, weil sich diese vielmehr ergänzen, empfiehlt es sich für umfassendere Untersuchungen, auf verschiedene methodische Vorgehensweisen zurückzugreifen (Dillenbourg, Baker, Blaye & O'Malley, 1996; Fischer, 2002).

4.2 Einflußfaktoren auf das kooperative Lernen

Im Folgenden sollen nun empirische Ergebnisse zu den Einflußfaktoren des kooperativen Lernens dargestellt werden. Da es sich beim kooperativen Lernen um die Interaktion einer Gruppe von zwei oder mehr Personen mit einem bestimmten Ziel

handelt, lohnt sich an dieser Stelle auch ein Blick auf die Ergebnisse sozialpsychologischer Forschung. Relevant sind hier insbesondere Untersuchungen zur Zusammensetzung von Gruppen, zur Lösung verschiedener Aufgabenstellungen durch Gruppen und zur Leistungsfähigkeit von Gruppen im Vergleich zu Individuen. Diese werden jeweils Befunden aus der Pädagogischen Psychologie gegenübergestellt.

4.2.1 Zusammensetzung von Gruppen

Bezüglich ihrer Zusammensetzung werden natürliche Gruppen (Familie, Schulklasse) und für Experimente zusammengestellte ad-hoc Gruppen voneinander unterschieden (McGrath, 1984). Als für die Kooperation relevante Merkmale nennt McGrath die Lebensdauer der Gruppe und die Bandbreite der gemeinsam zu bewältigenden Aufgaben. Im Hinblick auf empirische Untersuchungen ist bei natürlichen Gruppen die Kontrolle der Unterschiede zwischen den Gruppen schwierig, weshalb die Großzahl der durchgeführten Studien in diesem Bereich mit ad-hoc Gruppen durchgeführt wurde. Allerdings schränkt gerade dies die Validität solcher Studien deutlich ein (Arrow, McGrath & Berdahl, 2000). Dies gilt insbesondere auch für die Beobachtung von Gruppenprozessen bei komplexen Aufgaben in realistischen Settings (vgl. Boos, Scharpf & Fisch, 1991).

Meist werden sogenannte Kleingruppen mit sechs bis acht Mitgliedern untersucht (McGrath, 1984; Laughlin, Bonner & Altermatt, 1998; Laughlin, Magley & Shupe, 1997). Ihr Verhalten wird in einigen Fällen mit dem von Individuen verglichen (beispielsweise Hill, 1982; McGrath, 1984). Dyaden werden in diesem Zusammenhang selten untersucht. Levine und Moreland (1990) postulieren große qualitative Unterschiede in der Interaktion von Dyaden und größeren Gruppen. Tschan (2002) untersuchte dies empirisch und verglich Dyaden mit Triaden und Einzelpersonen. Sie konnte zeigen, daß Dyaden im Gegensatz zu größeren Gruppen einen geringeren Koordinationsaufwand verlangen, insgesamt aber im Vergleich zu Individuen mehr Ähnlichkeiten als Unterschiede zwischen Dyaden und größeren Gruppen bestehen. Elaboriertere Gruppenstrukturen, wie etwa Hierarchien, können sich bei Dyaden kaum bilden.

Auch in der Pädagogischen Psychologie gibt es viele Untersuchungen zum kooperativen Lernen in Gruppen mit vergleichbarer Größe von 4-8 Schülern. Meist handelt es sich um Untersuchungen im Klassenzimmer (etwa Fleer, 1992; Wegerif, 1996; Wells, 1993; White & Frederiksen, 1998). Aus den Beiträgen der Einzelnen kann (im Gegensatz zum lauten Denken) allerdings nicht auf ihr Wissen geschlossen werden, denn bei einer Gruppengröße von 4-8 Teilnehmenden bleibt den Einzelnen nur wenig Zeit, überhaupt etwas zu sagen und damit Daten für eine Wissensdiagnose zu produzieren. Die Evaluation des Lernerfolgs muß also über Nachtests geschehen.

Einen Kompromiß stellt Hatano und Inagaki (1991) zufolge die Untersuchung von Dyaden dar. Das Setting regt dazu an, die Vorgehensweise beim Problemlösen im natürlichen Dialog abzustimmen, zu begründen und damit zu verbalisieren. Dadurch, daß sich die beiden Beteiligten abwechseln, kann sich keiner zurückziehen und beide müssen eine gewisse Menge an verbalen Daten produzieren, die für eine Diagnose zumindest einige Anhaltspunkte geben können. In einer Dyade kann sich keiner von beiden aus dem Dialog zurückziehen, wie dies in einer größeren Gruppe möglich ist. Dies machen sich Untersuchungen wie beispielsweise Roschelle und Teasley (1995) oder Kneser und Plötzner (2001) zunutze. Gerade bei Aufgaben, die gemeinsam an einem Computer zu lösen sind, scheint es auch realistisch zu sein, maximal Dyaden beim kooperativen Problemlösen zu beobachten.

Betrachtet man die Eigenschaften der Gruppenmitglieder, so beschreiben Moreland, Levine und Wingert (1996) Vor- und Nachteile der Unterschiedlichkeit (*diversity*) der Gruppenmitglieder. Größere Unterschiedlichkeit in Gruppen bringt die positiven Effekte von mehr Flexibilität, Innovation durch neue Mitglieder und damit je nach Aufgabe eine Verbesserung der Gruppenperformanz mit sich. Für den Fall, daß konvergierendes Denken gefragt ist, sollten hingegen homogene Gruppen besser abschneiden als heterogene. Negativ wirkt sich aus, daß mehr Konflikte auftreten, die Energie und Zeit kosten. Auch das Geschlecht muß als Unterschiedsfaktor für Gruppen betrachtet werden, konnten doch mehrere Studien zeigen, daß Männer in gemischtgeschlechtlichen Gruppen eher bei Aufgaben mit

geringer sozialer Komplexität eine Führungsrolle übernehmen. Frauen nehmen dagegen bei sozial komplexeren Aufgaben etwas häufiger eine Führungsrolle ein (für einen Überblick siehe Eagly & Karau, 1991).

Auch in der pädagogischen Psychologie gibt es Befunde darüber, daß sich eine gewisse Heterogenität der Lernenden bezüglich verschiedener Merkmale beim kooperativen Lernen und Problemlösen positiv auswirkt. Dies trifft insbesondere für das Vorwissen und den Leistungsstand der Teilnehmenden zu (Dansereau, 1988; Howe, Tolmie, & Rodgers, 1990, 1992). Bei Vorwissen aus unterschiedlichen Bereichen können die Lernenden ebenso voneinander profitieren, wie wenn starke und schwache Schüler in einer Gruppe gemeinsam eine Aufgabe lösen. Dies bestätigt sich auch für den Fall, daß den Lernenden vor dem kooperativen Problemlösen gezielt unterschiedliches Vorwissen vermittelt wurde (Aronson, Baloney, Stephan, Sikes, & Snapp, 1978; Plötzner, Fehse, Kneser & Spada, 1999). Beim kooperativen Problemlösen zu zweit wirkt sich Heterogenität bezüglich des Geschlechts der Lernenden, sprich ein Junge und ein Mädchen sollen zusammenarbeiten, eher negativ aus (Barbieri & Light, 1992, Underwood, Underwood & Wood, 2000). In gemischtgeschlechtlichen Dyaden dominieren Jungen nämlich meist das gesamte Geschehen.

Über die einzelnen Gruppenmitglieder hinaus kann auch die Gruppe als Ganzes charakteristische Eigenschaften entwickeln. Bei Moreland, Levine und Wingert (1996) wird das als spezielle Chemie bezeichnet. McGrath (1984) widmet dem Thema interpersonelle Anziehung ein eigenes Kapitel. Einflußfaktoren sind demnach physische Anziehung (Schönheit), Ähnlichkeit von sozioökonomischem Status, Alter, Komplementarität der Bedürfnisse, räumliche Nähe und ähnliche inhaltliche Interpretationen.

Azmitia und Montgomery (1993) fanden heraus, daß Paare von Lernenden, die miteinander befreundet sind, beim kooperativen Problemlösen erfolgreicher lernen als Paare, die einander vorher nur entfernt kannten. Freunde kritisieren einander stärker, rechtfertigen ihre Ideen härter und elaborieren die Ideen des Partners mehr. Weitere Gründe für den größeren Lernerfolg könnten Sympathie, höhere

Motivation oder auch eine schon entwickelte gemeinsame Sprache sein. Die zuletzt genannten Gründe sind eher hypothetisch. Die Autoren fordern mehr Forschung in diesem Bereich.

4.2.2 Gruppenaufgaben

Welche Aufgaben für welche Art von Kooperation geeignet sind, stellt McGrath (1984) in einem umfassenden Modell ausführlich dar (Abbildung 1). Die Gruppenaufgaben werden zwei Dimensionen zugeordnet: Kooperation-Konflikt und konzeptuell-behavioral. Zudem klassifiziert er die Aufgaben anhand der Aktivitäten, die zu ihrer Erfüllung ausgeführt werden müssen: Generieren, Auswählen, Verhandeln und Ausführen. Daraus resultieren acht Typen von Aufgaben, die sich in einem Kreismodell anordnen lassen, wobei benachbarte Typen Ähnlichkeiten aufweisen.

Wissenschaftliches Problemlösen umfaßt vielfältige Prozesse, daher können auch mehrere der beschriebenen Gruppenaufgaben dabei eine Rolle spielen. Während Planungs- und Kreativitätsaufgaben (im Modell unter (1) und (2)) wichtige Voraussetzungen für wissenschaftliches Arbeiten bilden, ist das unter (3) aufgeführte Problemlösen sicherlich von zentraler Bedeutung. Dabei faßt McGrath (1984) unter korrekten Antworten sowohl logisch unmittelbar einsichtige und richtige Lösungen als auch Expertenkonsens für den Fall, daß die korrekte Lösung nicht unmittelbar evident ist. Für das wissenschaftliche Problemlösen ist vor allem auch der letztere Fall relevant. Es handelt sich um konzeptuelle und in Kooperation zu lösende Aufgaben. Die Performanz von Gruppen hierfür liegt laut McGrath immer über dem Durchschnitt der individuellen Performanz aller Gruppenmitglieder, aber unter der Leistung, die resultieren würde, wenn immer nur der Beste zum Zug käme.

Allerdings ist im wissenschaftlichen Betrieb der Expertenkonsens nicht weit vom Expertenstreit entfernt oder es ist schwierig, von absoluter Korrektheit zu sprechen. Daher ist Expertenstreit mindestens genauso wichtig für das Fortkommen der Wissenschaft wie Expertenkonsens und läßt sich in den Abschnitten (4) und (5) der Typisierung wiederfinden. McGrath charakterisiert Entscheidungsaufgaben unter (4) als Suche nach der bevorzugten Antwort oder der Mehrheitsant-

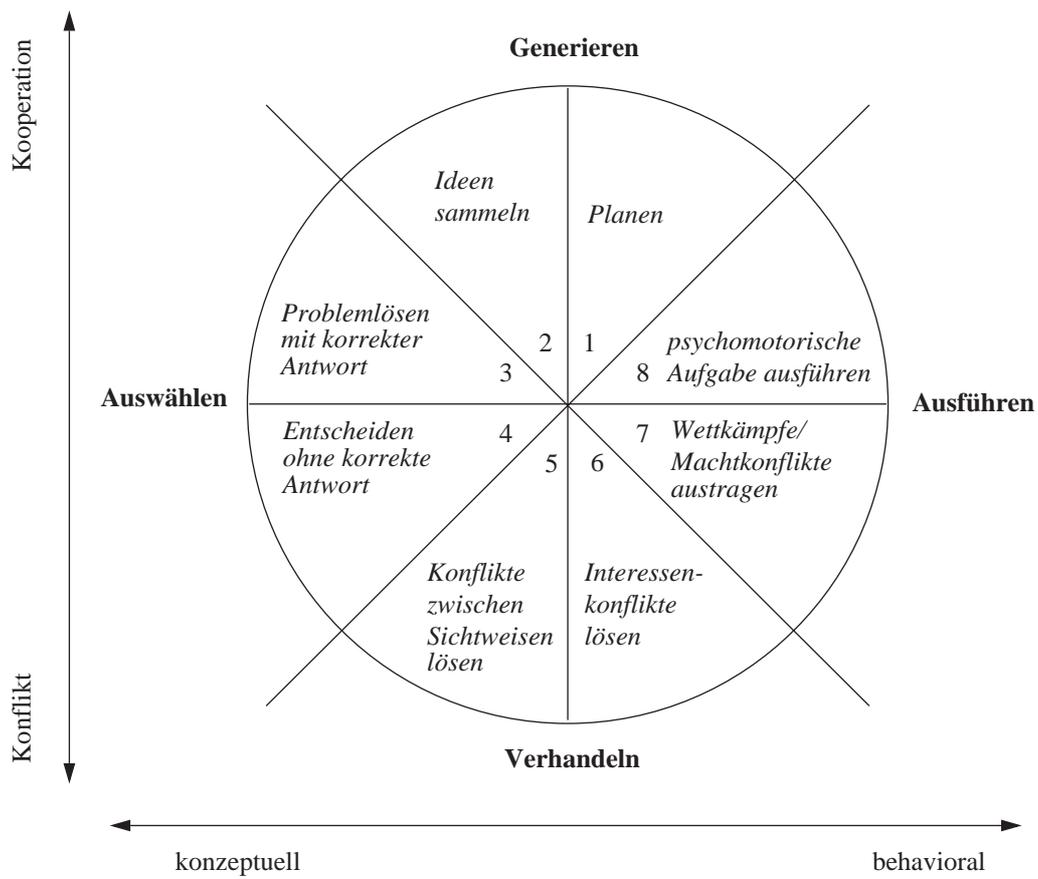


Abbildung 1. Typisierung von Gruppenaufgaben nach McGrath (1984).

wort, etwa in Jurys. Eng damit verwandt sind Konfliktlösungen (5), wenn die Gruppenmitglieder systematisch unterschiedliche Präferenzen haben und daher die zur Verfügung stehende Information unterschiedlich gewichten.

In einer Studie von Stasser (1991) fanden sich unterschiedliche Ergebnisse für das Zusammentragen von Information in einer Gruppe, je nachdem ob die Aufgabe als Problemlösen oder als Entscheidungsaufgabe ausgegeben wurde. Beim Problemlösen wurde mehr Information berücksichtigt als beim Entscheiden. Die Autoren begründen dies damit, daß es beim Problemlösen um das Finden einer richtigen Lösung gehe und dazu soviel Information wie möglich in Erfahrung gebracht werden müsse, wogegen beim Urteilen die Präferenzen nicht bis ins letzte Detail gerechtfertigt werden müßten. Die generelle Bedeutung der Aufgabe für die Qualität von Gruppenergebnissen konnte unter anderen Hackman (1968) nachwei-

sen. In dieser Studie ist der Aufgabentyp, verglichen mit anderen Einflußfaktoren auf den Erfolg von Gruppen, für den bei weitem höchsten Anteil an Varianzaufklärung verantwortlich.

Insgesamt läßt sich anhand der Typisierung aufzeigen, daß sich wissenschaftliches Problemlösen in den Bereichen Generieren und Auswählen abspielt, daß viele der Aufgaben Kooperation erfordern und daß konzeptuelle Aspekte sehr viel stärker einbezogen sind als behaviorale. Planen (1) ist die einzige Aufgabe, in die auch behaviorale Aspekte einfließen. Hier zeigt sich gleichzeitig eine Schwäche der vorgestellten Aufgabenklassifikation und auch anderer Klassifikationen, die hier nicht näher dargestellt werden sollen (für einen Überblick siehe Tschan, 2000): Das Modell sieht gar nicht vor, daß ein Gruppenprojekt mehrere dieser Aufgabentypen umfassen könnte. Somit lassen sich für diesen Fall auch keine Vorhersagen über das Verhalten der Gruppe und deren Leistung ableiten. Das Modell verschließt sich damit einer Übertragung in einen angewandten Kontext. Sobald eine Gruppe außerhalb eines sozialpsychologischen Labors beobachtet werden soll, greift das Modell zu kurz.

Auch für den Lernerfolg beim kooperativen Problemlösen scheinen kooperative Zielstrukturen von großer Bedeutung zu sein (Knight & Bohlmeier, 1990). Mit kooperativen Zielstrukturen ist gemeint, daß die Teilnehmenden eines oder mehrere gemeinsame Ziele verfolgen, die sie alleine oder im Wettbewerb miteinander nicht erreichen können. Johnson und Johnson (1994) ebenso wie Slavin (1995) betonen den förderlichen Effekt einer gegenseitigen positiven Abhängigkeit der Lernenden (siehe auch Johnson, Maruyama, Johnson, Nelson & Skon, 1981), wenn gleichzeitig die Aufgabe auch die Eigenverantwortlichkeit der Lernenden nicht aus den Augen läßt. Individualistische oder kompetitive Zielvorgaben wirken sich im Vergleich dazu nicht so positiv auf den Lernerfolg aus. Wettbewerb kann allenfalls zwischen den Gruppen einen gewissen Ansporn geben, innerhalb einer Lerngruppe ist er keine sinnvolle Motivation für den Wissenserwerb.

4.2.3 Leistungsvergleich von Gruppen und Individuen

Für die Auswahl einer Alternative bei Entscheidungsaufgaben entwickelte Davis (1973) die *Social Decision Schemes* (SDS), die verschiedene Entscheidungsmodelle für konzeptuelle Aufgaben (*intellective tasks*) enthalten: (1) Wahrheit

gewinnt – wenn jemand das Richtige sagt, schließen sich die anderen an, (2) unterstützte Wahrheit gewinnt – wenn mindestens zwei das Richtige sagen, setzt sich das durch, (3) einfache Mehrheit gewinnt und (4) qualifizierte Mehrheit gewinnt. Laughlin und seine Kollegen (Laughlin, 1980; Laughlin & Johnson, 1966; Laughlin, Kerr, Davis, Halff & Marciniak, 1975; Laughlin, Kerr, Munch & Haggerty, 1976) untersuchten die Güte dieser Entscheidungsmodelle für verschiedene konzeptuelle Aufgaben (u.a. Konzepterwerb, Vokabeltests, Analogien). Das Modell „unterstützte Wahrheit gewinnt“ ergibt für fast alle Aufgabenstellungen die besten Vorhersagen und führt zu Ergebnissen, die so gut sind wie das zweitbeste Gruppenmitglied. Einzige Ausnahme bilden Heureka-Aufgaben, bei denen die korrekte Lösung unmittelbar als solche erkannt wird, sobald sie gefunden wurde. Hier macht das Modell „Wahrheit gewinnt“ die besten Vorhersagen.

Weitere Gründe dafür, daß Gruppen meist nicht so gute Leistungen erbringen wie das beste Individuum, liegen in spezifischen Gruppenprozessen: Statusunterschiede entsprechen nicht Kompetenzunterschieden; ein kompetentes Gruppenmitglied besitzt nur ein geringes Durchsetzungsvermögen; eine inkompetente Mehrheit übt sozialen Konformitätsdruck aus; die Gruppe hat Schwierigkeiten bei der Klassifikation von Antworten als richtig oder falsch.

Gruppen können sich allerdings in bestimmten Settings auch positiv auf die Leistung auswirken. Beflügelnd wirken sie unter Umständen als Wettbewerber oder als Publikum (Zajonc, 1965). Wenn Gruppenmitglieder ihren Beitrag zum Gruppenergebnis für einmalig halten, zeigen sie in der Gruppe auch keine verminderte Leistung (Huguet, Charbonnier & Monteil, 1999). Als Vorteile einer Gruppe nennt Hill (1982) zudem gegenseitige Motivation, Entwicklung besserer Strategien und Lernvorteile durch gegenseitige kognitive Stimulation.

Die Darstellung der Entscheidungsmodelle läßt sich nicht ohne Weiteres auf Dyaden übertragen. Für eine Gruppengröße von $N = 2$ würden alle Entscheidungsmodelle außer „Wahrheit gewinnt“ vorhersagen, daß es für die beiden Beteiligten nur Konsensentscheidungen gibt. Die Entscheidungen in einer Dyade sollten also einfacher sein als in größeren Gruppen. Laughlin, VanderStoep und Hollingshead (1991) schlugen weitere Entscheidungsmodelle vor, die sich bei Dyaden anwenden lassen: Bei Uneinigkeit werden die Vorschläge der beiden Beteiligten abwechselnd

berücksichtigt. Hier wird deutlich, daß es sich bei Vorschlägen dieser Art nicht um Pläne größeren Umfangs handeln kann, denn bei größeren Plänen wäre es nicht möglich, sich Schritt für Schritt abzuwechseln. Dieses Entscheidungsmodell bezieht sich vorrangig auf Aufgaben wie Regelinduktion (Laughlin et al., 1998; Laughlin et al., 1997). Zwei weitere Entscheidungsmodelle nach Laughlin et al. (1991) sind bei Uneinigkeit das Aufstellen einer gemeinsamen neuen Hypothese oder die argumentative Demonstration, daß eine Hypothese der anderen überlegen ist. Kerr, MacCoun und Kramer (1996) fanden heraus, daß Gruppen, verglichen mit Individuen, dazu tendieren, ungenauer zu arbeiten und Informationen zu übergehen. In dieser Untersuchung zeigte sich auch, daß Gruppen jede Art von Verzerrungen noch verstärken, etwa die einseitige Wahrnehmung einer Situation oder das Ignorieren bestimmter Informationen. Ein ausführlicher Überblick zur Qualität von Gruppenleistungen findet sich auch bei Tschan (2000).

In der pädagogischen Psychologie finden Vergleiche zwischen Gruppen und Individuen meist unter der Hypothese statt, daß kooperatives Lernen effektiver sei als individuelles Lernen (Barron, 2000; Blaye, Light, Joiner & Sheldon, 1991; Cohen, 1994; Dansereau, 1988; Fantuzzo, Riggio, Conelly & Dimeff, 1989; Teasley, 1995). Die Überlegenheit kooperativen Lernens zeigt sich allerdings nicht in allen Studien (Guntermann & Tovar, 1987; für einen Überblick siehe Slavin, 1995). Diese Überlegenheit wird einerseits mit einer erhöhten Motivation begründet (Fantuzzo et al., 1989), andererseits mit kognitiven Aspekten, wie der Konfrontation mit anderen Ansichten und der Situation, die eigenen Überzeugungen erklären, begründen und verteidigen zu müssen, sowie der Tatsache, daß sich alle Beteiligten auf einem ähnlichen Niveau befinden und daher Erklärungen meist eine für die Partner angemessene Komplexität besitzen (Okada & Simon, 1997; Webb, 1989). Die Vorteile kooperativen Lernens und Problemlösens treten allerdings nur bei geeigneten Aufgaben zutage (De Lisi & Golbeck, 1999), das heißt die Aufgaben müssen hinreichend komplex sein, etwa Planungsaufgaben, umfangreichere Textaufgaben, Diskussion eines konträren Themas oder auch Hypothesentesten (Moshmann & Geil, 1998). Kooperatives Lernen ist vor allem dann effektiv, wenn bei einer geeigneten Aufgabe den Lernenden gewisse Strukturen und Strategien für die Kooperation nahegelegt werden (Fantuzzo et al., 1989; Gillies, 2004; Hron,

Hesse, Reinhard & Picard, 1997; King, 1997). Dieser Befund deckt sich mit einer Untersuchung von Stasser, Taylor und Hanna (1989), bei der sich ergab, daß in strukturierte Diskussionen mehr Informationen einfließen als in unstrukturierte.

Dieser Überblick zeigt die komplexen Einflußvariablen, die in Kleingruppen und beim kooperativen Lernen wirken: die Zusammensetzung der Gruppe und die Eigenschaften ihrer einzelnen Mitglieder, die Charakteristika der Gruppenaufgaben und gegebenenfalls die Strukturen und Strategien zur Bewältigung einer Gruppenaufgabe. Dennoch bleibt die Betrachtung dieser Variablen für die pädagogische Forschung und Praxis unbefriedigend. Häufig ist es in Studien nicht möglich, das komplexe Zusammenspiel der genannten Faktoren ausreichend zu kontrollieren, um effektives kooperatives Lernen zu gewährleisten. Andererseits weisen selbst bei Kontrolle dieser Faktoren die Lernerfolge eine beträchtliche Heterogenität auf. Beides spricht dafür, die Prozesse während des kooperativen Lernens einer genaueren Betrachtung zu unterziehen (Dillenbourg et al., 1996), um herauszufinden, inwiefern bestimmte Interaktionsmuster mit dem Lernerfolg zusammenhängen und ob diese Interaktionsmuster gar induziert werden können.

4.3 Prozesse kooperativen Lernens und ihre didaktische Unterstützung

Zur Analyse der Prozesse beim kooperativen Lernen schlagen Kumpulainen und Mutanen (1999) drei Ebenen vor: die kognitive, die soziale und die verbale. Während die verbale Ebene durch die sprachlichen Äußerungen der Teilnehmenden unmittelbar zugänglich ist, müssen die sozialen und vor allem die kognitiven Funktionen anhand von Äußerungen erschlossen werden. Weil die drei genannten Ebenen so ineinander verwoben sind und eine Trennung künstlich erscheint, werden sie in der Forschungspraxis häufig nicht getrennt analysiert.

Einen zentralen Teil des Lernens, ob in einem Lehrer-Schüler-Verhältnis oder mit Gleichaltrigen, machen Frage-Antwort-Dialoge aus. In einer Metaanalyse stellt Webb (1989) fest, daß Erklärungen meist nicht ohne Aufforderung gegeben werden, sondern erfragt werden müssen. Je präziser und direkter diese Fragen gestellt werden, desto größer ist die Chance auf eine ausführliche Erklärung. Graesser und Person (1994) nehmen an, daß Fragen ein Mittel der Selbstregulation des Lernens darstellen, indem sie den Lernenden erlauben, sich aktiv um die Beseitigung von Wissensdefiziten zu bemühen. Hierfür ist beim kooperativen Lernen

viel mehr Raum als im herkömmlichen Unterricht. In der Untersuchung von Graesser und Person zeigte sich, daß erfolgreiche Lernende seltener Fragen stellen, dafür aber in ihren Fragen komplexe und ausführliche Erklärungen einforderten. Selbst wenn Fragen nicht beantwortet werden, können sie noch zum Lernerfolg beitragen, indem sie die fragende Person dazu anregen, selbst nach einer Erklärung zu suchen (Kneser & Plötzner, 2001).

Webb (1989) faßt in ihrer Arbeit auch zusammen, wie Erklärungen während des kooperativen Lernens beschaffen sein sollten, um den Lernerfolg zu gewährleisten: Eine Erklärung soll relevant sein, auf dem Niveau des Lernpartners und für diesen verständlich, in zeitlicher Nähe zur Frage abgegeben werden, für den Lernenden mit der Möglichkeit verbunden sein, das neu gewonnene Wissen gleich anzuwenden (siehe auch Bielaczyc et al., 1994; Salomon & Perkins, 1998). Kurze oder nicht gegebene Erklärungen wirken sich immer negativ auf den Lernerfolg aus. Komplexe Erklärungen wirken nicht nur auf den Lernerfolg der fragenden Person positiv aus, sondern auch auf denjenigen der erklärenden Person. Rückfragen verstärken diesen Effekt. Howe, Tolmie, Duchak-Tanner und Rattray (2000) fanden, daß auch Konsens zwischen den Lernenden zum Lernerfolg beiträgt.

Von großer Bedeutung für erfolgreiches Lernen sind auch metakognitive Prozesse (Berardi-Coletta, Buyer, Dominowski & Rellinger, 1995). Als solche sind die Reflexion des eigenen Verständnisses und sogenanntes *Monitoring* (Überwachen) des Problemlösens und des Lernfortschritts zu betrachten. Metakognition wurde von Chi, Bassok, Lewis, Reimann und Glaser (1989) als wichtiger Faktor für den Lernerfolg beim individuellen Lernen durch Selbsterklärungen identifiziert. Einen Zusammenhang mit dem Lernerfolg beim kooperativen Problemlösen fanden Dansereau (1988) und Kneser und Plötzner (2001).

Eindeutig auf die verbale und soziale Ebene bezogen sind zwei weitere Aspekte der sozialen Interaktion beim kooperativen Lernen, die Kohärenz der dabei stattfindenden Dialoge und die wechselseitig eingenommenen Rollen. Die Kohärenz eines Dialogs ist um so größer, je stärker sich die Beiträge der Teilnehmenden aufeinander beziehen. Daß sich eine gute Kohärenz positiv auf den Wissenserwerb auswirkt, bestätigten die Arbeiten von Kneser und Plötzner (2001) sowie Roschelle und Teasley (1995).

Rollen können durch externe Gegebenheiten zugeteilt werden, etwa wenn aufgrund von räumlicher Enge beim kooperativen Problemlösen nur jeweils eine Teilnehmerin den Problemraum manipulieren kann und die andere Teilnehmerin dadurch zur Beobachterin wird (Miyake, 1986) oder wenn vor dem Computer nur einer der Kooperierenden die Maus navigieren kann (Blaye et al. 1991). Diese Rollenverteilung hat auch Auswirkungen auf die Verteilung der kognitiven Funktionen der Lernenden, so findet man bei Personen in der beobachtenden Rolle verstärkt reflektierende Äußerungen. Wechselnde Rollen mit ähnlichen Auswirkungen auf den Lernerfolg lassen sich auch im freien Dialog beim kooperativen Problemlösen beobachten (Kneser & Plötzner, 2001). Die förderliche Wirkung von Rollen, aber auch von einzelnen Dialogstrategien, konnte auch durch verschiedene Trainingsprogramme, für Schülerinnen und Schüler der unteren Klassen, nachgewiesen werden (Herrenkohl & Guerra, 1998; King, 1997; Meloth & Deering, 1994; Mercer, Wegerif & Dawes, 1999; Palincsar & Brown, 1984).

Es wurde bereits erwähnt, daß dem Vergleich von Dyaden mit Individuen für die Erforschung des kooperativen Lernens ein großer Stellenwert zukommt. Auch für diesen Vergleich sind die Prozesse während des Wissenserwerbs von besonderem Interesse. Dennoch finden sich bisher wenige Studien, die sich dieser Aufgabe stellen. Neben dem Aufwand einer solchen Untersuchung mag einer der Gründe in den methodischen Problemen eines solchen Vergleichs liegen, denn auch er stützt sich auf verbale Daten, die aber von unterschiedlicher Herkunft und Qualität sind. Beim kooperativen Problemlösen findet ein natürlicher Dialog statt. Die Teilnehmenden äußern aber nicht unbedingt alle ihre Überlegungen, vielleicht weil sie im Eifer des Gefechts untergehen oder weil der Partner sie nicht zu Wort kommen läßt. Wenn eine Teilnehmerin ihrer Lernpartnerin zustimmt oder Aspekte aus den Beiträgen ihrer Partnerin wiederholt, so ist nicht sichergestellt, daß sie diese auch verstanden hat. Der Dialog stellt auch zusätzliche kognitive Anforderungen, weil er koordiniert werden muß. Zur Untersuchung von Dialogen kommen diskursanalytische Verfahren (Pilkington & Parker-Jones, 1996; van Boxtel, van der Linden & Kanselaar, 2000) zur Anwendung, mit denen auch typische Dialogmerkmale wie Kohärenz und Initiative im Dialog untersucht werden können.

Verglichen werden Dialoge dann mit Protokollen lauten Denkens von Einzelpersonen (Ericsson & Simon, 1993). Dabei verbalisiert die Versuchsperson ihre Denkschritte, ohne sie selbst dabei zu analysieren. Ericsson und Simon gehen davon aus, daß lautes Denken bei geeigneter Instruktion die kognitiven Prozesse einer Person qualitativ nicht verändert, lediglich verlangsamt. Obwohl die Autoren die Ähnlichkeit von Protokollen lauten Denkens mit Dialogen betonen, sind einige bedeutende Unterschiede zwischen beiden offensichtlich. Die Koordination mit einer anderen Person entfällt und damit auch wichtige Indikatoren wie Kohärenz und Initiative im Dialog. Protokolle lauten Denkens werden als Ausdruck des verbalisierbaren Inhalts des Arbeitsgedächtnisses betrachtet. Sie müssen nicht von einem Dialogpartner verstanden werden und nicht im Zusammenhang mit vorherigen Äußerungen stehen. Sie entsprechen daher noch häufiger als Dialoge nicht den linguistischen Normen für vollständige Sätze. Ziel einer Protokollanalyse ist es, aus den verbalen Daten Rückschlüsse darüber zu erhalten, wie sich die Versuchsperson durch den Problemraum arbeitet. Im besten Fall können ihre Äußerungen direkt mit Operatoren und spezifischen Zuständen im Problemraum in Verbindung gebracht werden.

Trotz der genannten Unterschiede wurden bei der Analyse von Protokollen lauten Denkens ähnliche Strukturen für erfolgreiche Lernende gefunden wie in Dialogen, allerdings meist nicht mit Hilfe protokollanalytischer Methoden, sondern durch an die Diskursanalyse angelehnte Kategoriensysteme (*verbal analysis* nach Chi, 1997). So berichten Chi et al. (1989) in ihrer Untersuchung zu Selbsterklärungen mit Einzelpersonen bei erfolgreichen Lernenden mehr Erklärungen und metakognitive Äußerungen zu spezifischen Verständnisproblemen als bei weniger erfolgreichen Lernenden. Deren metakognitive Äußerungen waren allgemeiner und zogen seltener Erklärungsversuche nach sich.

Aufgrund der unterschiedlichen Datenqualität wählten Vye, Goldman, Voss, Hmelo, Williams und die Cognitive and Technology Group at Vanderbilt (1997) jeweils ein eigenes Analyseschema für die verbalen Daten von Einzelpersonen und Dyaden und beschränkten ihren Vergleich auf qualitative Beobachtungen von Operationen im Problemraum.

Teasley (1995) verglich erstmals direkt zwischen den verbalen Daten von Einzelpersonen und Dyaden. Dafür verwendete sie die computerbasierte Experimentalumgebung Big Trak (Klahr & Dunbar, 1988) mit der Aufgabe, die Funktion einer unbekanntesten Taste des Roboters herauszufinden, und untersuchte vier Gruppen von Viertklässlern ($N = 70$): Dyaden mit Erlaubnis zu sprechen, Dyaden ohne Erlaubnis zu sprechen, Einzelpersonen mit Instruktion zum lauten Denken und Einzelpersonen ohne lautes Denken. Für die Beurteilung der Leistung der Versuchspersonen wurden Prozeßdaten herangezogen, vor allem die letzte Hypothese zur Funktion der unbekanntesten Taste. Ein Nachttest fand nicht statt. Dyaden mit Sprecherlaubnis schnitten am besten ab, gefolgt von Einzelpersonen mit der Instruktion zum lauten Denken und Einzelpersonen ohne lautes Denken. Die schwächste Leistung zeigten Dyaden, die nicht miteinander sprechen durften.

Bis auf die beiden spezifisch für die Dialoge vorgesehenen Kategorien „mit dem Partner abklären“ und „Aufgabenverteilung“ bezogen sich alle für die Analyse der verbalen Daten vorgesehenen Kategorien auf Äußerungen zum Problem und zum Gegenstandsbereich. Sie lassen sich daher eher der Protokollanalyse als einer verbalen Analyse zuordnen. Dyaden redeten signifikant mehr als Einzelpersonen beim lauten Denken, weshalb für die weitere Analyse Prozentwerte herangezogen wurden. Als sinnvolle Auswertungseinheiten stellten sich die Anzahl der durchgeführten Experimente und die pro Experiment gemachten Äußerungen je Kategorie heraus. Für beide Gruppen galt, daß die erfolgreichereren unter ihnen mehr Äußerungen machten als die weniger erfolgreichen. Rein beschreibende Äußerungen korrelierten negativ mit den Leistungswerten, interpretierende und elaborierende Äußerungen dagegen positiv. Dyaden machten außerdem mehr Vorhersagen und mehr Aussagen über ihr Programm.

Eine weitere Untersuchung mit einem Vergleich der Verbaldaten von Dyaden und Einzelpersonen führten Okada und Simon (1997) durch. Sie arbeiteten mit 27 männlichen Studierenden im Grundstudium eines naturwissenschaftlichen Faches und griffen auf dasselbe Problem aus der Genetik zurück, das bereits in der oben beschriebenen Studie von Dunbar (1993) benutzt wurde, einen für das wissenschaftliche Problemlösen validen Gegenstandsbereich. Auch hier wurden nur Prozeßdaten und verbale Daten erhoben und keine individuellen Leistungstests

nach dem Experiment durchgeführt. Als Kriterium für die Leistung wurde wieder die letzte Hypothese über die Funktion der in Frage stehenden Gene herangezogen und von zwei voneinander unabhängigen Beurteilern auf einer fünfstufigen Rating-skala eingeschätzt. Die Interrater-Reliabilität dieser Einstufung war hinreichend, aber nicht sehr gut (75% Übereinstimmung). Dyaden zeigten sich auch in dieser Untersuchung gegenüber Einzelpersonen überlegen. Sie kamen zu den besseren abschließenden Hypothesen.

Auf der Suche nach den Gründen für die unterschiedlichen Ergebnisse der beiden Experimentalgruppen verglichen die Autoren zunächst die Operationen im Experimenterraum, die Breite der Suche, den Informationsgehalt der Experimente und die Systematik der Suche. Hier fanden sich keinerlei bedeutsame Unterschiede zwischen den beiden Gruppen. Deutliche Unterschiede fanden sich im Hinblick auf den Hypothesenraum: Dyaden stellten ungefähr dreimal mehr Hypothesen auf als Einzelpersonen und unterhielten häufiger mehrere Hypothesen gleichzeitig. Besonders effektiv waren Dyaden, wenn sie sowohl die kritischen Experimente im Gegenstandsbereich durchführten als auch häufig über die Begründung ihrer Ergebnisse aus dem simulierten genetischen Labor diskutierten. Erklärungen dazu wurden um so öfter gegeben, je mehr danach gefragt wurde. Als Auslöser für solche Fragen identifizierten Okada und Simon (1997) fünf Mechanismen: ein überraschendes empirisches Ergebnis, die Zusammenfassung von Ergebnissen, die Vervollständigung einer genetischen Teilkomponente, den Widerspruch zu einer Erklärung des Partners oder Nicht-Verstehen einer solchen und schließlich metakognitive Einsichten über das Mißtrauen gegenüber oder Nicht-Verstehen von eigenen Erklärungen. Damit ist mindestens ein Mechanismus eindeutig von der Anwesenheit eines Lernpartners abhängig, bei den anderen erhöht die Anwesenheit des Lernpartners mit Sicherheit deren Auftretenswahrscheinlichkeit. Diese Form der Analyse bezieht Elemente der verbalen Analyse mit ein und geht damit über die Arbeit von Teasley (1995) hinaus. Einige Ergebnisse deuten in dieselbe Richtung, indem nochmals die Bedeutung aufgestellter Hypothesen und tiefgründiger Erklärungen bestätigt wurde.

An dieser Stelle seien noch zwei weitere Untersuchungen erwähnt, in denen kooperatives Lernen erfolgreich für das Erlernen von Fertigkeiten des wissenschaftlichen Problemlösens bei Studierenden eingesetzt wurde. Kaartinen und Kumpulainen (2002) gaben eine Aufgabe im sozialwissenschaftlichen Kontext vor, Derry, Levin, Osana, Jones und Peterson (2000) entwarfen ein Curriculum für die Statistik. In beiden Fällen werden Lernerfolge der Teilnehmenden berichtet, allerdings sind die Analysen nicht so detailliert wie in den oben geschilderten Arbeiten, weshalb sich eine ausführlichere Darstellung an dieser Stelle erübrigt.

Zusammenfassend läßt sich über die Prozesse beim kooperativen Lernen und Problemlösen sagen, daß sich über mehrere Studien hinweg immer wieder dieselben Merkmale als besonders förderlich für den Wissenserwerb herausgestellt haben. Hierbei handelt es sich insbesondere um Fragen nach tiefgehenden Erklärungen sowie um diese tiefgehenden Erklärungen selbst. Einen positiven Einfluß haben auch metakognitive Prozesse. Methodisch sind für den Vergleich von Dyaden und Einzelpersonen zwei Ebenen von Bedeutung, einerseits die verbalanalytische und andererseits die kognitiv-protokollanalytische. Die beiden Studien, die einen direkten Vergleich von Dyaden und Einzelpersonen auf der Ebene der Kooperationsprozesse anstrengen, umgehen die Frage nach dem individuellen Lernerfolg ihrer Teilnehmenden und stellen daher auch keinen Bezug zwischen Prozeßdaten und extern erhobenen Leistungsdaten her.

Aufgrund dieser Befunde geht die Pädagogische Psychologie davon aus, daß kooperatives Lernen dem individuellen Lernen überlegen ist und die aufgeführten positiven Effekte der Kooperation die aus der Sozialpsychologie bekannten negativen Effekte überwiegen. Dies trifft insbesondere für komplexe Aufgaben zu. Desweiteren sollte die Gruppenzusammensetzung beachtet werden. Hier wirkt sich heterogenes Vorwissen positiv aus. Gruppen mit Lernenden desselben Geschlechts sollte der Vorzug gegeben werden.

Nachdem bisher verschiedene didaktische Ansätze dargestellt wurden, die die Lernprozesse und das soziale Setting des Lernens in den Vordergrund stellten, soll im nächsten Kapitel das Medium Computer im didaktischen Einsatz zur Vermittlung von Fertigkeiten für das wissenschaftliche Problemlösen dargestellt wer-

den. Wie bereits bei der Beschreibung des situierten und entdeckenden Lernens anklang, geben Simulationen offenbar geeignete Lernumgebungen für diesen Zweck abzugeben.

5

Lernen am Computer

Die Idee, Computer für das Lernen einzusetzen, entstand, sobald sich technische Möglichkeiten dafür abzeichneten. Während man sich zunächst von einer möglichst engmaschigen Führung der Lernenden mit immer intelligenteren Diagnosemöglichkeiten die Entwicklung sinnvoller Lernmedien versprach (Kerres, 2001), begann man parallel zur Entwicklung der didaktischen Ansätze des entdeckenden und des situierten Lernens, offenere und flexiblere Computerlernumgebungen zu schaffen, in denen aktives und selbstgesteuertes Lernen unterstützt wurde (Andriessen & Sandberg, 1999; Derry & Lajoie, 1993; Reusser, 1993). Die Erkenntnis setzte sich durch, daß die Lernenden selbst in der Lage sein sollten, ihren Wissensstand zu diagnostizieren, sich Ziele zu setzen und zu planen, wie sie beim Lernen vorgehen (Scardamalia, Bereiter, McLean, Swallow & Woodruff, 1989). Nach Land und Hannafin (2000) sollten solche lernerzentrierten Lernumgebungen fünf Bedingungen erfüllen: (1) Der Fokus liegt auf den Lernenden und ihren Anstrengungen, selbst Wissen zu konstruieren; (2) Lernkontexte sind situiert und authentisch Lernkontexte; (3) Diskussion persönlicher Überzeugungen und multiple Perspektiven werden gefördert; (4) frühere Lernerfahrungen werden in die Konstruktion von Wissen einbezogen; (5) Technologie wird eingesetzt, um höhere mentale Prozesse zu unterstützen. Mikrowelten und insbesondere Simulationen wurden als geeignete Lernmedien eingesetzt, um wissenschaftliche Fertigkeiten zu erlernen und einzuüben (Alessi, 2000; Shute & Glaser, 1990; Windschitl & Andre, 1998).

Immer wieder stellt sich die Frage nach einem Vergleich von verschiedenen Unterrichtsmedien und unterschiedlichen Szenarien für deren Einsatz. Schulmeister (1997) faßt zusammen, daß zahlreiche ältere und neuere Studien die einerseits die Schwierigkeiten bei der Untersuchung solcher Fragestellungen belegen, etwa überhaupt vergleichbare Unterrichtsmedien und Einsatzszenarien zu entwickeln

und diese so zu evaluieren, daß die Ergebnisse nicht vom Hawthorne-Effekt geprägt sind. Andererseits können nach Schulmeister vergleichende Studien in den meisten Fällen, wie durch Meta-Analysen belegt, keine Leistungsunterschiede für Teilnehmerinnen und Teilnehmer an computervermittelten und herkömmlichen Kursen feststellen. Aus mehreren Gründen ist es dennoch sinnvoll, sich weiterhin um die Entwicklung computerunterstützter Lehr-Lernsysteme zu bemühen. Sie ermöglichen Lernen in anderen Szenarien als herkömmlicher Unterricht und erreichen damit einen anderen Personenkreis als dieser. Sie können bei sehr guter Planung und Nutzung unter Umständen kostensparend wirken. Vor allem sprechen auch qualitativ-inhaltliche Aspekte für das Lernen mit neuen Medien. Computerunterstützte Lehr-Lernsysteme haben eigene Qualitäten, die im herkömmlichen Unterricht nicht angeboten werden können. So können durch den Computer eine Vielzahl von Materialien zur Verfügung gestellt werden, die den Lernenden sonst kaum zugänglich wären. Dies fördert die Reflexion und pluralistische Sichtweisen. Darüberhinaus ermöglichen Simulationen den Lernenden die direkte Manipulation von Experimenten, die sie sonst nur von Lehrerin oder Lehrer vorgeführt bekommen hätten. Dadurch kann ein aktiver Wissenserwerb stattfinden. Simulationen lassen sich beliebig oft abspielen und können alle gewünschten Daten präzise protokollieren.

Der Einsatz von Simulationen als Lernumgebungen wird im Folgenden als erstes beschrieben. Da Simulationen hohe Ansprüche an die Lernenden stellen, kamen früh Überlegungen auf, wie das Lernen mit Simulationen zu unterstützen wäre. Eine interessante Möglichkeit dafür sind kognitive Werkzeuge, die im Anschluß an die Simulationen besprochen werden.

5.1 Simulationen als offene Lernumgebungen

Die Entwicklung von Mikrowelten und Simulationen für die Vermittlung wissenschaftlicher Fertigkeiten wurde stark von Papert (1980) geprägt. Mit der Entwicklung der Programmiersprache „Logo“ sprach er sich vor allem für den Einsatz von Computern als spielerische Lernumgebung für Kinder aus. Er betonte, daß Simulationen die Motivation für schwierige Fächer wie Mathematik und Physik fördern können, indem die Lernenden selbst aktiv mit der Materie umgehen. Dabei ist unter einer Simulation ein Computermodell einer realen Situation oder eines Aus-

schnittes davon zu verstehen, das von den Lernenden manipuliert werden kann. Laurillard (1987) ergänzte als weitere Vorteile von Simulationen den direkten Zugang zum Gegenstandsbereich und das intrinsische Feedback, welches Simulationen durch die Ergebnisse der Lernermanipulationen zurückmelden. Gleichzeitig gibt sie zu bedenken, daß Simulationsprogramme die Ziele der Lernenden nicht direkt unterstützen, keine Ratschläge geben, und Lernende daher oft nicht die konzeptuell zentralen Experimente durchführen. Hierfür forderte sie didaktische Unterstützung ein.

Rivers und Vockell (1987) sehen diese Unterstützung im Ansatz des *guided discovery learning*. Sie beschreiben den experimentellen Zyklus in einer Simulation in den drei Schritten Planen des simulierten Experiments, Durchführen des Experiments und Evaluieren der simulierten Experimentaldaten und wählten zur Umsetzung des *guided discovery learning* zunächst verbale Instruktionen, die sich dann auch empirisch als wirkungsvoll erwiesen. Als weitere Vorteile des Lernens mit Simulationen gegenüber echter Laborforschung stellen sie heraus, daß die Lernenden mehr Probleme lösen können als im realen Labor und den experimentellen Zyklus mehrmals durchlaufen können. Dadurch haben die Lernenden bei geringem Ressourceneinsatz mehr Möglichkeiten zu üben. Zudem sind die Datensätze in Simulationen kontrolliert erhoben worden, womit grobe Fehler im Datensatz vermieden werden.

Simulationen können recht unterschiedlich gestaltet werden. So unterscheidet Gray (2002) vier Arten von Simulationen: Als erstes nennt er hochrealistische Simulationen (1) für komplexe Aufgaben, beispielsweise Flugsimulatoren. Gray betont, daß die Realitätsnähe einer Simulation immer im Verhältnis zur Aufgabe zu sehen ist. So kann beispielsweise der Gruppenzusammenhalt in einer militärischen Simulationsaufgabe ein anderer sein als im Feld. Hochrealistische Simulationen mit einfachen Aufgaben können zum Beispiel Teile eines komplexen Systems sein, etwa ein GPS-Navigator aus einem Cockpit. Skalierte Welten (2) konzentrieren sich auf ein Teilsystem funktioneller Beziehungen aus einem komplexen System und versuchen, diese funktionellen Beziehungen realistisch abzubilden, während andere ausgeblendet werden. Auf diese Weise sind für ein und denselben Gegenstandsbereich verschiedenste skalierte Welten vorstellbar. Die

Gestaltung einer skalierten Welt setzt daher eine genaue Definition der Lernziele und eingehende Kenntnis des Gegenstandsbereiches voraus. Ein Beispiel findet sich bei Opwis (1992). Es handelt sich um die Lernumgebung DiBi, ein virtuelles Physiklabor zur Simulation elastischer Stoßvorgänge. In DiBi liegt der Schwerpunkt dezidiert auf den qualitativen Aspekten des Gegenstandsbereiches, etwa das Nachvollziehen von Beziehungen zwischen Größen, die in den dazugehörigen Gleichungen vorkommen. Quantitative Berechnungen übernimmt die Lernumgebung, damit sich die Lernenden auf die genannten qualitativen Aspekte konzentrieren können. Während das Lernziel bei den skalierten Welten eine spezifische Anwendung ist, sollen synthetische Lernumgebungen und Mikrowelten (3) abstraktere Fertigkeiten vermitteln, beispielsweise Planungsaufgaben. Synthetische Lernumgebungen und Mikrowelten sind vor allem bei Forschenden beliebt, da dieser Typ der Simulation große Gestaltungsfreiräume gestattet und es daher erlaubt, flexible Experimentalumgebungen zu schaffen. Laboraufgaben und simulierte Aufgabenumgebungen (4) werden herangezogen, um genau einen spezifischen Aspekt einer komplexen Aufgabe zu untersuchen. Hier wären beispielsweise Aufgaben ähnlich denjenigen von Wason (1966) zu nennen, die dann als Beispiele für übergeordnete Konzepte herangezogen werden.

Um seine Einteilung noch plastischer zu machen, führt Gray (2002) drei Dimensionen ein, auf denen sich die vier Simulationstypen darstellen lassen: Handhabbarkeit mit den Polen schwierig und leicht, Korrespondenz zur Realität mit den Ausprägungen „viele Aspekte eines Systems“ gegenüber „ein Aspekt vieler Systeme“ und Engagement, was von langweilig bis anregend reichen kann. Es zeigt sich, daß vor allem für skalierte Welten und Mikrowelten große Spielräume auf den Dimensionen Handhabbarkeit und Engagement bestehen und für deren Entwicklung genutzt werden sollten. Hier wird deutlich, daß die Gestaltung von Simulationen große Freiräume läßt und daher vor der Entwicklung die Ziele für den Einsatz einer Simulation genau definiert werden müssen (Reigeluth & Schwartz, 1989).

Für die Vermittlung von Fertigkeiten für das wissenschaftliche Problemlösen an Schüler oder Studierende im Grundstudium erscheinen die skalierten Welten als der angebrachte Simulationstyp. Sie erlauben es, die bekannte Komplexität der

Aufgabe auf angemessene Weise zu regulieren und dennoch die notwendige Realitätsnähe zu wahren, um eine Vorbereitung für die Forschungspraxis zu gewährleisten. Aber auch absolut realistische Simulationen können für die Vermittlung von Fertigkeiten für das wissenschaftliche Problemlösen geeignet sein. So fanden Triona und Klahr (2003), daß Schülerinnen und Schüler mit einer realistischen physikalischen Simulation denselben Lernerfolg erzielten wie bei realen Experimenten. Es gilt also, für den Zweck der Lernumgebung bei der Entwicklung das passende Abstraktionsniveau zu wählen.

Eine Klassifikation der Ziele beim Lernen mit Simulationen stellten van Berkum, Hijne, de Jong, van Joolingen und Njoo (1991) auf und fokussierten dabei auf die Art des zu erwerbenden Wissens. Für diese Klassifikation schlagen sie drei Dimensionen vor (Abbildung 2): (1) Konzeptuelles Wissen wird operationalem Wissen gegenübergestellt. Konzeptuelles Wissen wird als Wissen über die Definition und Klassifikation von Begriffen definiert. Operationales Wissen beinhaltet Abläufe von Operationen, hier Sequenzen in einer Simulation. (2) Deklaratives und kompiliertes Wissen sind nicht zu verwechseln mit den beiden Polen der ersten Dimension. Deklaratives Wissen ist leicht zu verbalisieren und auf verschiedene Probleme anwendbar, dafür muß es für die Anwendung kompiliert werden. Kompiliertes Wissen wird durch Benutzung beim Problemlösen gebildet, ist schwerer zu verbalisieren und daher nur für einen begrenzten Kontext anwendbar. Dafür laufen Problemlöseprozesse mit diesem Wissen automatisiert ab.

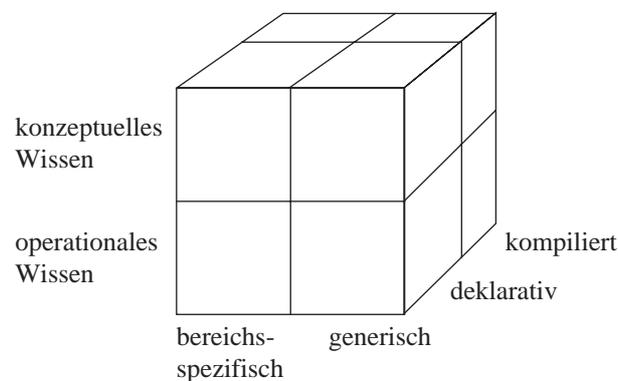


Abbildung 2. Klassifikation von Lernzielen beim Lernen mit Simulationen nach van Berkum, Hijne, de Jong, van Joolingen und Njoo (1991).

(3) Bereichsspezifisches und generisches bzw. bereichsübergreifendes Wissen bilden die Pole der dritten Dimension. Während die Handhabung von Simulationen einiges an bereichsspezifischem Wissen voraussetzt, können mit ihrer Hilfe auch bereichsübergreifende Fertigkeiten erworben werden. So ist es etwa in einem simulierten Labor möglich, allgemeine Heuristiken wie den sinnvollen Ablauf eines experimentellen Zyklus' zu erlernen. Als weiteres übergeordnetes Lernziel nennen van Berkum et al. (1991) den Erwerb von Metawissen über den Lernprozeß als solchen, Wissen über kognitive Prozesse und deren Produkte.

Goodyear, Njoo, Hijne und van Berkum (1991) entwickelten ein Beschreibungsschema für die Lernprozesse beim Problemlösen mit Computersimulationen. Sie beziehen sich auf die Klassifikation von van Berkum et al. (1991), legen ihren Schwerpunkt aber auf die Lernprozesse und unterscheiden diese explizit von Lernzielen, Instruktion und Lerneraktivitäten. Im Vordergrund steht für sie entdeckendes Lernen. Aber auch situiertes Lernen läßt sich mit Simulationen teilweise realisieren, zumindest versprechen Simulationen eine sehr realitätsnahe Lernumgebung und schaffen damit eine wichtige Voraussetzung für situiertes Lernen, auch wenn die sozialen Umstände andere sein können.

Goodyear et al. (1991) beschreiben auf der ersten Ebene ihres Schemas (Tabelle 1) im Grunde genommen die Phasen des experimentellen Zyklus' für wissenschaftliche Forschung: Orientierung, Formulierung von Hypothesen, Überprüfung und Evaluation. Für die zweite Ebene beschreiben die Autoren eine mögliche Aufteilung in sogenannte Transaktionen, die auf der dritten Ebene in basale Gedächtnisoperationen aufgegliedert werden. Diese Aufteilung läßt sich auf weiteren Ebenen noch verfeinern, wird aber von den Autoren nicht mehr detailliert ausgeführt.

Die erste Ebene ist so allgemein gehalten, daß sie sich auf unterschiedlichste Lernvorgänge anwenden läßt. Betrachtet man zum Beispiel das Lernen mit Texten, so wäre die Orientierungsphase diejenige, in der Vorwissen aktiviert wird und gegebenenfalls der Lerntext ausgesucht wird. Die Hypothesengenerierung bestünde darin zu antizipieren, was von dem Text zu erwarten ist, etwa indem man sich anhand von Zwischenüberschriften über den Inhalt orientiert. Der Test

Tabelle 1

Beschreibungsschema für Lernprozesse beim explorativen Lernen mit Simulationen nach Goodyear, Njoo, Hijne und van Berkum (1991).

Ebene 1	Ebene 2	Ebene 3
Orientierung	Beobachtung	Enkodierung Abruf
	Analyse	Enkodierung Abruf Abgleich
	Modellexploration	...
Aufstellen von Hypothesen	–	
Überprüfung	Versuchsplanung	
	Vorhersage Manipulation Interpretation	
Evaluation	Beobachtung	
	Generalisierung	

bestünde dann darin, den Text zu lesen, die neuen Informationen aufzunehmen und in der Evaluation mit den Hypothesen abzugleichen, um dann diese zu verändern, erneut über den Text zu gehen etc.

So breit sich das Schema auf unterschiedliche Lernprozesse anwenden läßt, so genau paßt es doch auf wissenschaftliches Problemlösen. Hier spiegelt sich die Feststellung Bechtels (1988) wider, daß Wissenschaft der Wissenserwerb auf sehr hohem Niveau ist. Lernprozesse und Lernziele stimmen hier in weiten Teilen überein. Der Klassifikation von van Berkum et al. (1991) zugeordnet fallen diese Lernziele vorrangig in die Kategorien bereichsübergreifend-generisch und operational. Das Schema zeigt auf, daß es möglich ist, die hier beschriebenen Schritte des wissenschaftlichen Problemlösens bei der Entwicklung von simulierten Lernumgebungen angemessen zu berücksichtigen.

Den Schwierigkeiten bei der Messung des Lernerfolges durch das Lernen mit Simulationen widmen sich Swaak und de Jong (2001). Vor allem das dabei erworbene sogenannte intuitive, nicht deklarative Wissen, wie es von Neuweg (1999) beschrieben wurde, ist mit herkömmlichen Methoden kaum faßbar. Es wurde in einer perzeptuell reichen, dynamischen Situation erworben, ist schwer verbalisierbar und mit der Wahrnehmung und Repräsentation der Simulation ver-

knüpft (Baumgartner, 1999). Der Abruf erfolgt auf andere Weise als bei explizitem, deklarativem Wissen. Betrachtet man die genannten Eigenschaften intuitiven Wissens, so bietet sich vor allem eine Messung in einem ähnlichen Kontext an, etwa anderen Aufgaben oder einem anderen Gegenstandsbereich in derselben Simulation, Neuweg spricht hier von einer direkten Leistungsbeurteilung. Auf diese Weise lassen sich jedoch keine sauberen Vortest-Nachtest-Designs erheben. Aus diesem Grund schlagen Winne und Perry (2000) eine Mehrfachmessung vor, bei der Fragebögen, Protokolle lauten Denkens und Leistungstests zur Beurteilung herangezogen werden. Derry et al. (2000) erhoben die genannten Fertigkeiten in Interviews, in denen sie anwendungsbezogene Fragen stellten. Diese Interviews wurden später von Ratern bewertet. Swaak und de Jong entwickelten Richtlinien für Tests, die unabhängig von der Simulationsumgebung auch mit Papier und Bleistift durchgeführt werden können. Sie schlagen vor, die Test-Items mit visuellen Repräsentationen, hauptsächlich in Form von Graphiken, zu gestalten. Sie empfehlen situative Fragen und solche nach Vorhersagen. Zudem sollten diese Tests unter Geschwindigkeitsbedingungen durchgeführt werden. Leider wurden zwar mit Hilfe dieser Tests gemessene Lernzuwächse von durchschnittlich 10% berichtet, Gütekriterien für die Tests aber nicht berichtet.

In der Arbeit von Goodyear et al. (1991) werden auch die Eigenschaften der Lernenden berücksichtigt. Dabei geht es den Autoren nicht darum, einzelne Eigenschaften als entscheidend für den Lernerfolg beim Lernen mit Simulationen herauszustellen. Stattdessen sehen sie diese eher als ganze Cluster und warnen auch davor, diese als statisch zu betrachten. Dennoch stellen sie einige Eigenschaften als einflußreich heraus. Es handelt sich um Intelligenz (siehe auch Renkl & Schweizer, 2000; Süß, 1999), Vorwissen (siehe auch Schauble, Glaser, Raghavan & Reiner, 1991; Süß, 1999), Leistungsmotivation, Ängstlichkeit, Selbstkonzept, externale oder internale Kontrollüberzeugung und kognitiven Stil (Feld(un)abhängigkeit). Die Autoren versuchen, die genannten Eigenschaften in ein Lernermodell zu integrieren und somit Aussagen über das Verhalten der Lernenden zu gewinnen. Das Modell ist allerdings so komplex, daß es kaum zu handhaben ist. Dennoch scheint der Hinweis auf die Eigenschaften der Lernenden einerseits im Hinblick auf eine

geeignete Diagnostik im Rahmen von empirischen Untersuchungen des Lernens mit Simulationen und andererseits im Rahmen der Entwicklung adaptiver Lernumgebungen von Bedeutung (vergleiche Shute, 1993; Shute & Towle, 2003),

Eine große Studie zum Lernen mit Simulationen führten Shute und Glaser (1990) durch. Sie untersuchten wissenschaftliches Problemlösen und Lernen mit Studierenden in der Simulationsumgebung *Smithtown*. Als Gegenstandsbereich wurde die Mikroökonomie aus den Wirtschaftswissenschaften herangezogen. In *Smithtown* wird ein Markt simuliert, Studierende können Variablen von Angebot und Nachfrage verändern und die wirtschaftlichen Auswirkungen auf die Jahreszyklen beobachten. Dabei werden die Studierenden durch zwei Maßnahmen unterstützt. Zum einen ist die Reihenfolge eines experimentellen Zyklus festgelegt, und die Studierenden werden auf diese Weise geführt, können aber die Variablen für ihre Experimente frei wählen. Zum anderen enthält das System eine Diagnosekomponente, basierend auf einer Fehlerbibliothek und dem Vergleich mit einem Expertenmodell (Polson & Richardson, 1988). Dadurch ist das System in der Lage, das Vorgehen beim wissenschaftlichen Problemlösen zu überwachen und den Lernenden sowohl positive Rückmeldungen zu geben, als auch gegebenenfalls Korrekturen anzubringen. Zusätzlich erlauben mehrere Archive den Zugriff auf alle bereits von den Lernenden in der Sitzung mit der Simulation erhobenen Daten. Wird die Diagnosekomponente ausgeschaltet, so können die Lernenden auch im sogenannten Entdeckungsmodus den Gegenstandsbereich frei explorieren, ohne Hypothesen aufzustellen.

In einem ersten Experiment mit *Smithtown* wurden drei Gruppen untersucht. Die erste arbeitete mit *Smithtown*, die zweite besuchte die herkömmlichen Vorlesungen zur Mikroökonomie und bei der dritten Gruppe handelte es sich um eine Kontrollgruppe ohne jegliche Instruktion. Jede Gruppe bestand aus $N = 10$ Studierenden im Grundstudium ohne Vorwissen in Wirtschaftswissenschaften. Die parallelen Vor- und Nachtests bestanden aus je zwei Teilen, einem Untertest mit Multiple-Choice-Items und einem Untertest mit Fragen, zu denen freie Antworten gegeben werden mußten. Obwohl das erklärte Ziel des Lernens

mit Smithtown der Erwerb methodischen Wissens und allgemeiner Problemlösefertigkeiten war, prüften die Parallelltests nur bereichsspezifisches Wissen ab, womit sich die Autoren den obengenannten Meßschwierigkeiten entzogen.

Insgesamt ergab sich kein Unterschied zwischen Gruppe 1 und 2, aber diese beiden Gruppen schnitten besser ab als die Kontrollgruppe. Bei ausschließlicher Betrachtung der freien Antworten zeigte sich eine Überlegenheit der Smithtown-Gruppe. In einer Prozeßanalyse innerhalb der Smithtown-Gruppe ergab sich, daß erfolgreich Lernende kohärenter und zielgerichteter vorgehen als weniger erfolgreich Lernende, daß erstere Variablen systematischer veränderten, sich also eher daran hielten, nur eine Variable gleichzeitig zu verändern, und insgesamt mehr Vorhersagen machten. Shute, Glaser und Raghavan (1989) berichten weitere Zusammenhänge zwischen dem Umgang mit bereichsspezifischen Variablen und dem Lernerfolg. Je systematischer und tiefer die Studierenden einen Markt explorierten, desto besser wurde ihr Gesamtergebnis. Die oberflächliche Exploration mehrerer Märkte brachte hingegen nichts. Lernende, die theorie- und hypothesengeleitet vorgehen, waren erfolgreicher als solche, die eine datengeleitete Herangehensweise bevorzugten.

In einer groß angelegten Evaluation mit 530 Rekruten wurde auch der Einfluß allgemeiner Intelligenz auf den Lernerfolg mit Smithtown untersucht. Der Einfluß des Umgangs mit Hypothesen auf den Lernerfolg stellte sich jedoch als deutlich stärker heraus als derjenige der Intelligenz (siehe auch Vollmeyer & Burns, 1996). Hieraus schlossen Shute und Glaser (1990), daß das Lernen mit Simulationen allgemein gefördert werden könne, indem man gezielt den sinnvollen Umgang mit Hypothesen trainiere.

Shute und Glaser (1990) stoßen in ihrer Untersuchung auf dieselben Mechanismen erfolgreichen wissenschaftlichen Problemlösens, wie sie von Klahr und Dunbar (1988) oder Schunn und Anderson (1999) berichtet werden. Während die letzteren Arbeiten aber explizit an der Untersuchung des wissenschaftlichen Problemlösens an sich interessiert waren, geht es Shute und Glaser um die Vermittlung von Fertigkeiten für das wissenschaftliche Problemlösen. Allerdings überprüfen ihre Tests ausschließlich bereichsspezifisches Wissen, so daß bestenfalls ein indirekter Nachweis des Wissenserwerbs im Bereich der Methodologie zuerkannt wer-

den kann. Da die Abfolge der Schritte innerhalb des experimentellen Zyklus vom System vorgegeben waren, ist nicht klar, ob die Lernenden auch selbständig in der Lage gewesen wären, mit sinnvollen experimentellen Zyklen zu arbeiten. Für den Erwerb bereichsspezifischen Wissens ist die Simulation *Smithtown* durchaus geeignet und erscheint sogar effizienter als herkömmliche Vorlesungen, da die Studierenden im Umgang mit der Simulation ihr Wissen in der Hälfte der Zeit erworben wie die Studierenden, die an den Vorlesungen teilnahmen.

Eine weitere Studie zum Lernen mit Simulationen führten Njoo und de Jong (1993) durch. Als realistisches Setting für das Lernen an der Universität wurden Dyaden von Studierenden im Grundstudium herangezogen. Die Lernumgebung simulierte die Steuerung eines Schiffes. Durch verschiedene Aufgabenstellungen sollte die Wirkung angeleiteten und nicht angeleiteten entdeckenden Lernens (*guided* und *unguided discovery learning*) beleuchtet werden. In einem ersten explorativen Experiment mit acht Dyaden konnte kein Unterschied zwischen angeleitetem und nicht angeleitetem Lernen aufgezeigt werden, wobei die Anleitung lediglich in einer anfänglichen expliziten Instruktion zur Formulierung von Vorhersagen, deren Überprüfung und zur Begründung von Schlußfolgerungen bestand.

Im zweiten Experiment sollten die Versuchspersonen sich beim explorativen Lernen von einem Schema leiten lassen, das auf Papier vorgegeben wurde. In das Schema sollten unter anderem Hypothesen, Variablen, Daten und Schlußfolgerungen eingetragen werden. Zwei unabhängige Variablen wurden manipuliert: Die Zellen der Protokollbögen wurden entweder frei ausgefüllt oder die Hypothesen waren vorgegeben. Bei der Instruktion wurde entweder nur global oder detailliert in den Gegenstandsbereich eingeführt. Untersucht wurden 5 Gruppen von Dyaden ($N = 44 \times 2$), vier Gruppen mit den oben angegebenen kombinierten unabhängigen Variablen und eine Kontrollgruppe, die ohne weitere Anleitung mit der Simulation experimentierte. Die erwähnten Protokollblätter wurden auf fünf Ebenen analysiert: (1) Globale Aktivität – Anzahl der ausgefüllten Formulare und Zellen; (2) Lernprozeßvalidität – Abgleichung der Inhalte der ausgefüllten Zellen mit formalen Kategorien, ist in der Zelle Hypothese auch eine Hypothese enthalten? (3) Kor-

rektheit im Gegenstandsbereich; (4) Konsistenz zwischen den Zellen innerhalb eines Formulars – passen Hypothese und Experiment zusammen? (5) Gesamtstrategie – Entwicklung des Inhaltes derselben Zellen über mehrere Formulare hinweg.

Die experimentellen Ergebnisse fielen eher unbefriedigend aus, da sich für die verschiedenen Bedingungskombinationen kaum signifikante Unterschiede ergaben. Auch konnte kein Vorteil der Experimentalgruppen gegenüber der Kontrollgruppe festgestellt werden. Insgesamt schnitten die Versuchspersonen dann gut ab, wenn ihre Lernprozeßvalidität hoch bewertet worden war, wenn die Kategorien auf dem Protokollblatt also mit korrespondierendem Inhalt ausgefüllt worden waren. Die Studierenden zeigten mehr Aktivität als im ersten Experiment, was von den Autoren auf die bessere Anleitung zurückgeführt wird. Verglichen mit einer traditionellen Unterrichtsmethode schnitten die Studierenden nicht unbedingt besser ab. Allerdings hatten sie nur sehr begrenzte Zeit, um sich an die neue Lernmethode zu gewöhnen. Hier zeigt sich, daß auch Untersuchungen zum Lernen mit Simulationen zu heterogenen Ergebnissen kommen können, wie sie sich etwa auch bei Carlsen und Andre (1992) finden.

Auch bei dieser Studie zum kooperativen Lernen, im Falle von Njoo und de Jong (1993) mit Simulationen, fällt auf, daß kein externes Maß für den Lernerfolg erhoben wurde. Die Autoren beschränken sich auf die Analyse von Prozeßdaten, wobei sie die Dyaden als Einheiten betrachten und damit differentielle Effekte innerhalb der Dyaden vernachlässigen. Allein der Bezug auf Quellen, die für Einzelpersonen den Nutzen der als positiv herausgestellten Prozesse belegen, läßt in diesem Fall Rückschlüsse über Lernprozesse zu. Auch die eingesetzten Protokollblätter scheinen keine sinnvolle Umsetzung des angeleiteten entdeckenden Lernens zu sein. Dennoch hat diese Studie Beachtung verdient, da sie erstmals kooperatives wissenschaftliches Problemlösen mittels einer Simulation untersuchte.

Nachdem die unbestritten positiven Eigenschaften von Simulationen als Lernumgebungen dargestellt wurden, sei an dieser Stelle auf einige Probleme dieser Systeme im pädagogischen Kontext hingewiesen. Sie gleichen teilweise den Schwierigkeiten beim entdeckenden Lernen. So nennen de Jong und Njoo (1991), Goodyear et al. (1991) und Laurillard (1987) jeweils ähnliche Kritikpunkte: Das Lernziel ist beim Lernen mit Simulationen häufig nicht bekannt. Die oft beträchtli-

che Komplexität der Systeme gepaart mit einer mangelnden didaktischen Einbettung (etwa durch Übungen und Lehrtexte) führt leicht zur Überforderung der Lernenden, insbesondere der Anfänger. So erklärt Schulmeister (1997) auch die heterogenen empirischen Ergebnisse, welche nicht in allen Fällen Vorteile für das Lernen mit Simulationen belegen können. Zudem bleiben Zusammenhänge häufig unklar, weshalb nur implizites Wissen erworben wird. Ein Mangel an Problemlösefertigkeiten kann zu unsystematischem Vorgehen, mangelhaften Vorhersagen, unvollständiger Exploration und letzten Endes zur Beibehaltung von Mißkonzepten führen. Hier wirkt sich negativ aus, daß Simulationen die Lernenden nicht führen und ihre Ziele nicht adaptiv unterstützen können.

Zum Abschluß dieses Abschnittes sei noch auf andere Herangehensweisen zur computerunterstützten Vermittlung methodischen Wissens hingewiesen. Diese greifen jeweils einen gesonderten Aspekt der Methodologie heraus und behandeln diesen dann ausführlicher, können aber nicht den gesamten experimentellen Zyklus vermitteln. So gibt es mittlerweile zahlreiche interaktive Statistiklehrmedien, beispielsweise VISUALSTAT (Bodemer, Plötzner & Feuerlein, 2001) oder StatSoft (StatSoft Inc., 2003). Einen anderen Ansatz verfolgen MacWhinney, James, Shunn, Li und Schneider (2001) sowie Reips und Neuhaus (2002). Sie gehen jeweils von einer Software aus, die entwickelt wurde, um Experimente am Computer zu definieren und über diese Software die Datenerhebung durchzuführen. Die Autoren schlagen vor, diese Software als Lehrmedium für die Versuchsplanung zu nutzen. Die langwierige Datenerhebung und Auswertung müssen die Lernenden dann allerdings selbst durchführen, ohne didaktische Unterstützung der Software.

Einige Autoren empfehlen didaktische Unterstützung für das Lernen mit Simulationen. Während viele Ansätze über das eigentliche Medium der Computersimulation hinausgehen, etwa die Anleitung durch Tutorinnen und Tutoren, erscheint für eine Verbesserung von Computersimulationen als solche eine Erweiterung durch sogenannte kognitive Werkzeuge (*cognitive tools*, siehe Jonassen, 1991, Thurman, 1993) besonders geeignet. Diese sind Gegenstand des folgenden Unterkapitels. Dort werden zunächst einige theoretische Grundlagen beschrieben.

Es folgen praktische Erwägungen zur Umsetzung des Ansatzes sowie eine Darstellung erster empirischer Untersuchungen zur Wirkungsweise kognitiver Werkzeuge.

5.2 Kognitive Werkzeuge für das Lernen am Computer

Kognitive Werkzeuge haben ganz allgemein den Zweck, kognitive Prozesse zu unterstützen. Während ein handwerkliches Werkzeug materiell greifbar ist, muß es sich bei einem kognitiven Werkzeug nicht zwangsläufig um ein reales Objekt handeln, es könnte stattdessen auch beispielsweise ein Symbolsystem sein (Salomon, 1988). Ein Werkzeug enthält in sich bereits den Zweck, dem es dient, und es ist nicht autonom, wie eine Maschine, sondern muß gekonnt gehandhabt werden, um seinen Zweck zu erfüllen. Ein kognitives Werkzeug dient einerseits als Mittel, um in der Welt zu handeln, und andererseits als kognitive Stütze, die eben diese Handlung erleichtert (Salomon & Perkins, 1998). Während diese allgemeine Definition auf unzählige Artefakte zutrifft, wurde der Begriff des kognitiven Werkzeugs vorrangig im Zusammenhang mit dem Lernen am Computer geprägt.

Die Idee, Computer in diesem Sinne als kognitive Werkzeuge zu nutzen oder sogar kognitive Werkzeuge dafür zu schaffen, äußerte zunächst Papert (1980). Über die Beschaffenheit der Computer als Kulturwerkzeuge schrieb Pea (1985), daß Computer das Denken nicht nur erweiterten, sondern dessen Reorganisation ermöglichen. Andere mentale Prozesse rückten in den Vordergrund. So sei beispielsweise Kopfrechnen oder von Hand ausrechnen heute nicht mehr so wichtig, weil diese Aufgabe von Taschenrechnern und Computern erledigt werde. Dafür gelte die Bedienung eines Textverarbeitungsprogramms heute als Kulturtechnik. Pea nimmt die Perspektive des situierten Lernens ein, daß Intelligenz nicht eine Leistung des Geistes allein ist, sondern ein Produkt der Beziehung zwischen mentalen Werkzeugen und intellektuellen Werkzeugen, die durch die Kultur geschaffen wurden.

Als konkrete Funktionen kognitiver Werkzeuge nannte Brown (1985), die Aufmerksamkeit zu fokussieren und das Lernen zu strukturieren, so daß wichtige Denkprozesse an die Oberfläche kommen und relevante Inhalte herausgefiltert, betont und annotiert werden können. Van Joolingen (1999) stellte fest, daß es kognitive Werkzeuge den Lernenden erlauben, kognitive Prozesse und deren

Ergebnisse explizit zu machen. Pea (1985) und Collins und Brown (1988) hoben die Rolle des Computers als Speichermedium hervor. Archive ermöglichen das Lernen aus Vergangenen. Gleichzeitig können Hypothesen für die Zukunft festgehalten werden.

Collins und Brown (1988) betonen auch die Rolle des Computers für die Bereitstellung multipler Repräsentationen für das Lernen. Am Computer ist es technisch einfach, verschiedene externe Repräsentationen des Lernstoffes darzubieten und es den Lernenden damit zu ermöglichen, den Stoff elaboriert zu verarbeiten oder auch diejenige externe Repräsentation auszuwählen, die ihnen am eingängigsten ist. Nach Reusser (1993) erlauben es externe Repräsentationen, zwischen ähnlichen Problemen zu differenzieren, operative, ikonische und symbolische Lösungswege zu finden sowie relevante Informationen durch perzeptuelle Vorgänge schnell wiederzuerkennen und abzurufen. Zudem strukturieren externe Repräsentationen das Problemlösen und erleichtern den Übergang von idiosynkratischen zu kulturell verbreiteten Repräsentationen. Eine Aufgabe mit Beteiligung externer Repräsentationen kann jedoch nur gelöst werden, wenn die Information von externen und internen Repräsentationen auf vernetzte, integrierte und dynamische Art verarbeitet wird (Zhang, 1997). Externe Repräsentationen werden dabei in interne transformiert und umgekehrt. Je mehr relevante Merkmale der abstrakten Struktur eines Problems extern repräsentiert werden, desto leichter fällt dessen Lösung. Zhang weist darauf hin, daß externe Repräsentationen es ermöglichen, Information zu teilen, und daher auch für das kooperative Lernen von Bedeutung sind. Er spricht damit als weitere theoretische Grundlage den Ansatz der verteilten Kognition (*distributed cognition*) an.

Der Ansatz der verteilten Kognition macht Aussagen darüber, wie kognitive Aktivität über interne menschliche geistige Prozesse, externe kognitive Artefakte und Gruppen von Menschen sowie über Raum und Zeit verteilt ist. Es wird nicht mehr eine einzelne Person und ihre individuelle kognitive Leistung betrachtet, sondern eine Person in Interaktion mit den Artefakten und gegebenenfalls mit anderen Personen, die an der Lösung eines Problems beteiligt sind, werden als ein kognitives System betrachtet (Rogers & Ellis, 1994). Hierdurch wird deutlich, welcher integraler Bestandteil des Lernens kognitive Werkzeuge sein können. Sie übernehmen

für die Lernenden gewisse kognitive Aufgaben, beispielsweise als externer Speicher, und ermöglichen dadurch neue kognitive Operationen, etwa eine Verarbeitung auf einem höheren kognitiven Niveau (Mayes, 1992).

Kognitive Werkzeuge können die kognitive Belastung (*cognitive load*) reduzieren (Sweller, 1988) und damit wesentlich zum Lernerfolg beitragen. Eine Entlastung wird erreicht, indem beispielsweise einfachere kognitive Prozesse vom System übernommen werden (Lajoie, 1993). Die Autorin berichtet aus einer Untersuchung, daß die Lernenden bei einer Aufgabe jeden einzelnen Teilprozeß sehr wohl selbst ausführen konnten, daß aber bei der Integration in einer komplexeren Aufgabe Fehler gemacht wurden, weshalb komplexere Aufgaben auf diese Weise kaum geübt werden konnten. Das computerbasierte System übernahm es in diesem Fall, einfache Tests durchzuführen und deren Ergebnisse zu berechnen. Dadurch wurde gegenüber der realen Situation auch Zeit für die Erhebung der Tests gespart. Die Lernenden erhielten die Testergebnisse und mußten diese selbst interpretieren. Auf diese Weise waren sie dazu in der Lage, auch die komplexe Interpretation der Testergebnisse zu meistern, was vorher nicht der Fall war. Salomon (1988) geht sogar soweit, auf der Basis von Vygotsky zu postulieren, daß gute kognitive Werkzeuge internalisiert werden, wenn sie in der Zone nächster Entwicklung liegen. Angesichts der Bedeutung kognitiver Werkzeuge in der Gesellschaft fordert Pea (1993) konsequenterweise, daß schon in der Schule gelernt werden sollte, wie man kognitive Hilfsobjekte entwirft.

Zusammenfassend beschreibt Jonassen (1991) kognitive Werkzeuge folgendermaßen: Sie erleichtern die kognitive Verarbeitung und können sowohl mentale, als auch computergestützte Hilfsmittel sein. Diese Werkzeuge werden genutzt, um die Lernenden zu einer sinnvollen kognitiven Informationsverarbeitung zu veranlassen. Sie sind Werkzeuge des Wissensaufbaus, sollen kognitive und metakognitive Lernstrategien aktivieren und können in verschiedenen Gegenstandsbereichen eingesetzt werden. Als computerbasierte Werkzeuge sollten sie Denkprozesse ergänzen und erweitern sowie dazu anregen, neue Information mit bereits bekannter Information in Bezug zu setzen. Kognitive Werkzeuge repräsentieren Lernen *mit* Informationsverarbeitungstechnologien im Gegensatz zum Lernen *von* Informationsverarbeitungstechnologien und sind konstruktivistisch, weil sie die Lernen-

den zum aktiven Aufbau von Wissen anregen, indem sie die Reflexion der eigenen Konzepte und des eigenen Verständnisses anregen und nicht die Präsentation objektiver Information in den Vordergrund stellen.

Aus übergeordneter Perspektive bemerkte Schulmeister (1997), daß kognitive Werkzeuge heuristische Komponenten in den Lernprozeß integrierten und einen ganzheitlichen Zugang zum Lernen erlaubten. Die Lernenden könnten sich größeren kognitiven Gesetzen allmählich durch entdeckendes Verhalten nähern, Konzepte generieren und erst im Prozeß einzelne Subkonzepte erschließen, während Instruktionssysteme einen analytischen Zugang favorisierten.

Van Joolingen (1999) beschäftigte sich mit der konkreten Gestaltung kognitiver Werkzeuge und griff als ein Beispiel ein Hypothesenscratchpad auf, eine Art Notizblock, mit dessen Hilfe Hypothesen formuliert werden können. Das Scratchpad enthielt eine Liste der verfügbaren Variablen und eine Auswahl verfügbarer Funktionen und Relationen, die durch Anklicken mit der Maus zu einer Hypothese zusammengestellt werden konnten. In einem zusätzlichen Archiv-Feld wurden die bereits generierten Hypothesen und ihr Status (getestet ja/nein, belegt/widerlegt) angezeigt (vergleiche Meyer, Miller, Steuck und Kretschmer, 1999). Van Joolingen schlägt vor, die Lernenden über ansteigende Schwierigkeitsgrade an die Formulierung heranzuführen, indem nach einer Demonstration des Hypothesenformulierens zunächst bereits fertige Hypothesen präsentiert werden, dann das beschriebene Scratchpad eingesetzt wird und schließlich nur noch Felder zur freien Formulierung vorgegeben werden.

Eine systematische Untersuchung verschiedener Arten von Hypothesenscratchpads legten van Joolingen und de Jong (1991) vor. Sie bezogen drei verschiedene Scratchpads in ihre Studie ein, erstens ein strukturiertes, bei dem Variablen und Funktionen angeklickt und zu einer Hypothese kombiniert werden konnten (van Joolingen, 1999). Das zweite Scratchpad war semi-strukturiert und bot den Lernenden auch eine Liste mit Variablen zur Auswahl an. Ihre Hypothesen mußten sie allerdings frei formulieren. Das dritte Scratchpad war unstrukturiert und bestand lediglich aus einem Eingabefeld für die freie Formulierung der Hypothesen. Die Scratchpads wurden in der Untersuchung zusammen mit einer Simula-

tion zur Fehleranalyse in der Chemie eingesetzt, waren allerdings nicht in die Simulation integriert, sondern existierten lediglich in Gestalt einer Papier-und-Bleistift-Form.

Die Untersuchung mit $N = 31$ Studierenden, auf drei Gruppen mit jeweils anderem Scratchpad verteilt, zeigte, daß die mit dem strukturierten Scratchpad formulierten Hypothesen eine bessere Syntax aufwiesen als die Hypothesen der anderen beiden Gruppen. Hierbei könnte es sich jedoch um ein Artefakt handeln, da die Formulierungen im strukturierten Scratchpad bis auf die einzusetzenden Variablen vorgegeben waren. Auch Studierende in der Gruppe mit dem semi-strukturiertem Scratchpad formulierten bessere Hypothesen als die Gruppe mit dem unstrukturierten Scratchpad. Zusätzlich waren deren Hypothesen detaillierter als diejenigen aus der Gruppe mit strukturiertem Scratchpad. Beide strukturierten Scratchpads führten zu einer weiträumigeren Exploration des Hypothesenraums. Auch in dieser Studie wurden nur direkte Maße für die Hypothesen erhoben. Der Lernerfolg wurde nicht gemessen. Eine Schlußfolgerung, daß die Scratchpads das Lernen unterstützen, ist also nicht angemessen. Ähnliche Ergebnisse auch mit einem Scratchpad in Papier-und-Bleistift-Form berichten Dantas, Kemm und Weaver (2003). Auch hier wirkte sich die intensive Nutzung des Scratchpads positiv auf die endgültige Interpretation der Daten aus, unabhängige Tests für den Lernerfolg wurden jedoch nicht durchgeführt.

Eine Simulation mit kognitiven Werkzeugen für einen Gegenstandsbereich aus der Physik setzte Reimann (1991) ein. Das System ermöglichte es, Hypothesen mit Hilfe simulierter Experimente anhand vereinfachter Daten zu testen. Der experimentelle Zyklus wird in mehrere Schritte zerlegt, Vorhersagen mit unterschiedlicher Präzision sind möglich, und über ein Archiv können Informationen aus früheren Experimenten eingesehen und organisiert werden. Das System kam in einer Beobachtungsstudie zum Einsatz, deren Ergebnisse sich allerdings nur auf interindividuelle Unterschiede in der Formulierung von Hypothesen und in der Exploration des Experimenterraums beziehen. Eine Evaluation der Lernumgebung wurde nicht durchgeführt.

Während noch keine kognitiven Werkzeuge dafür entwickelt wurden, die Versuchsplanung in experimentellen Simulationen zu unterstützen, stellen Lewis, Stern und Linn (1993) eine Möglichkeit vor, die Interpretation von empirischen Ergebnissen im Hinblick auf aufgestellte Hypothesen zu fördern. Dazu mußten die Lernenden ihre Hypothesen neben der schriftlichen Formulierung bereits in ein Schaubild eintragen. Diese Graphen wurden dann zusammen mit den Resultaten im Ergebnisschaubild angezeigt, wodurch ein direkter Vergleich möglich wurde. Leider bietet die Untersuchung von Lewis et al. keine systematische Evaluation dieses kognitiven Werkzeugs.

Eine bedeutende Rolle für das Lernen in Computerlernumgebungen spielt auch die unmittelbare Verfügbarkeit von Informationen (Berry & Broadbent, 1987; Lajoie, 1993; Leutner, 1993; Shute, 1993). Dies kann einerseits durch die bereits erwähnten Archive geschehen, in denen Informationen über den bisherigen Verlauf des Lernprozesses in einer Lernumgebung eingesehen werden können. Andererseits können deklarative Informationen zum Gegenstandsbereich und zur Methodologie auch über Hyperlinks oder gar in einem kompletten Hypertext-Wörterbuch zur Verfügung gestellt werden. Sie sind damit genau in dem Augenblick verfügbar, wenn sie gebraucht werden, und müssen nicht aus der Instruktionsphase erinnert werden.

Zusammenfassend stellen sich kognitive Werkzeuge als theoretisch fundierte Maßnahme zur Unterstützung computerbasierter Lernumgebungen dar. Sie lassen sich einsetzen, um gezielt spezifische Lernprozesse zu fördern und die Lernenden dennoch zum selbstgesteuerten Lernen anzuregen. Dies wurde hier vor allem im Zusammenhang mit Simulationen betrachtet, gilt aber auch für andere Lernumgebungen, beispielsweise Hypermedia. Empirisch validiert wurde die Wirkung kognitiver Werkzeuge bisher nicht systematisch. Für einige weiter verbreitete kognitive Werkzeuge konnten positive Auswirkungen auf die Arbeit in der Lernumgebung nachgewiesen werden, der Nachweis verbesserter Lernerfolge steht in einigen Fällen jedoch noch aus. Negative Erfahrungen mit kognitiven Werkzeugen werden in der Literatur nicht berichtet. Es ist daher sinnvoll, kognitive Werkzeuge

für Simulationen zur Vermittlung von Fertigkeiten für das wissenschaftliche Problemlösen einzusetzen, auch wenn ihre Wirkungsweise letztlich noch nicht vollständig untersucht ist.

6

Hypothesen

Die vorliegende Arbeit verfolgt zweierlei Ziele. Einerseits wurde mit *virtue* ein simuliertes psychologisches Labor entwickelt, das die Beobachtung wissenschaftlichen Problemlösens in einem definierten Setting erlaubt. In der vorliegenden Studie sollen damit Novizen untersucht werden. Andererseits wurde *virtue* als eine computerbasierte Lernumgebung entwickelt, und es sollen Lernprozesse beobachtet werden. Diese beiden Ziele spiegeln sich auch in den Hypothesen wider, die sich mit der Beobachtung wissenschaftlichen Problemlösens und mit der Vermittlung von Fertigkeiten dafür befassen.

Das simulierte Labor *virtue* wurde mit kognitiven Werkzeugen ausgestattet, um das Erlernen von Fertigkeiten für das wissenschaftliche Problemlösen zu unterstützen. Sowohl das Lernen mit Simulationen als auch die Unterstützung mit kognitiven Werkzeugen sollte den Lernerfolg fördern (Lajoie, 1993; Shute & Glaser, 1990), auch wenn Simulationen anderen Lernformen nicht in jedem Fall überlegen sind (Carlsen & Andre, 1992). Daher sollte in einer Untersuchung der Nachweis erbracht werden, daß der Erwerb der genannten Fertigkeiten mit *virtue* möglich ist.

In der Untersuchung von Schunn und Anderson (1999) wird beim wissenschaftlichen Problemlösen bereichsspezifisches und bereichsübergreifendes Wissen unterschieden, Experten mit und ohne bereichsspezifisches Wissen werden getrennt untersucht. Für Novizen ist davon auszugehen, daß weder bereichsspezifisches noch bereichsübergreifendes Wissen auf hohem Niveau vorhanden ist. Die methodischen Fertigkeiten bei der Planung, Durchführung und Auswertung psychologischer Experimente gehören größtenteils zum bereichsübergreifenden Wissen. Hier ist von Fertigkeiten die Rede, da es sich nicht um rein deklaratives Wissen handelt, sondern um Handlungswissen, das im simulierten Labor auch durch Wiederholung des experimentellen Zyklus eingeübt werden kann.

Das bereichsspezifische Wissen bezieht sich vor allem auf den Gegenstandsbereich der Experimente, auf den theoretischen Hintergrund, die Operationalisierung und den Einfluß der verschiedenen Faktoren und deren Interaktionen. Hier gibt es Schunn und Anderson (1999) zufolge spezifische Kenntnisse, die sich positiv auf die Versuchsplanung und Auswertung auswirken können, sowohl auf die Operationalisierung als auch auf die Interpretation empirischer Daten. Diesen Kenntnissen liegen umfangreiches Wissen und Erfahrung zugrunde liegen, welche im Rahmen einer als Lernumgebung entwickelten Simulation kaum vermittelt werden können. Da das didaktische Ziel bei der Entwicklung von *virtue* vor allem die Ausbildung experimenteller Methodologie war, wird bereichsübergreifendes Wissen eigens abgefragt.

Hypothese 1: Die Experimente in *virtue* führen zum Erwerb von bereichsübergreifenden Fertigkeiten. Dies äußert sich in einer Verbesserung bei den bereichsübergreifenden Items zwischen Vortest und Nachtest und gilt für beide Experimentalgruppen.

Methodologisches Wissen ist nicht leicht zu erfassen (Swaak & de Jong, 2001). Daher wird als weiterer Indikator für den Lernerfolg auch das bereichsspezifische Wissen überprüft. Auch dieses kann in Simulationen durch entdeckendes Lernen erworben werden, einige Simulationen wurden sogar hauptsächlich für diesen Zweck entwickelt (Reimann, 1991; Shute & Glaser, 1990). Die zweite Hypothese bezieht sich daher auf das bereichsspezifische Wissen.

Hypothese 2: Durch das Experimentieren im simulierten Labor *virtue* wird auch bereichsspezifisches Wissen erworben. Es wird eine Verbesserung in den bereichsspezifischen Items vom Vortest zum Nachtest erwartet, und zwar für beide Experimentalgruppen, also für Einzelpersonen wie für Dyaden.

Der Leistungsvergleich von Dyaden und Einzelpersonen bringt heterogene Ergebnisse. Vor allem aus der Sozialpsychologie werden auch Prozeßverluste in Gruppen berichtet (Kerr, MacCoun & Kramer, 1996; Laughlin, 1980; Laughlin, Kerr, Munch & Haggerty, 1976). Untersuchungen zum kooperativen Lernen haben im Gegensatz dazu häufig Befunde erbracht, die zu der Annahme berechtigen, daß

kooperatives Lernen dem individuellen Lernen überlegen ist (Barron, 2000; Blaye, Light, Joiner & Sheldon, 1991; Fantuzzo, Riggio, Conelly & Dimeff, 1989). Dies gilt insbesondere auch für den Erwerb von komplexem Wissen, wie es das Einüben in Methoden wissenschaftlichen Experimentierens darstellt. Die Prozesse kooperativen Lernens lassen sich in einem Papier-und-Bleistift-Test kaum messen. Sie sollten sich aber trotz allem auch in der vorliegenden Untersuchung auf den Lernfortschritt auswirken und in den Wissenstests bemerkbar machen.

Hypothese 3: Dyaden profitieren vom wissenschaftlichen Problemlösen mit *virtue* mehr als Einzelpersonen. Sie verbessern sich daher vom Vor- zum Nachtest stärker als Einzelpersonen.

Die bisherigen Hypothesen bezogen sich auf den als Differenz zwischen Nachtest und Vortest gemessenen Lernerfolg. In diesem Zusammenhang stellt sich die Frage, welche Prozesse genau diesem Lernerfolg bei Dyaden und Einzelpersonen zugrunde liegen. Hier lassen sich zwei Datenebenen unterscheiden, die verbalen Daten und die Verhaltensdaten, wie sie in den Logfiles des simulierten Labors registriert werden.

Die Prozeßdaten aus den Logfiles von *virtue* enthalten die aufgestellten Hypothesen, die durchgeführten Experimente und die Evaluation der Ergebnisse und Hypothesen. Einerseits spiegeln sie die Vorgehensweise der Versuchspersonen wider und sind ein Indikator für die Überlegungen, die zu diesen Verhaltensdaten geführt haben. Sie zeigen somit die methodische Kompetenz der Versuchspersonen. Andererseits bilden die durchgeführten Experimente, wie sie in den Prozeßdaten dokumentiert werden, die Basis für den Aufbau des bereichsspezifischen Wissens. Denn nur über diejenigen Faktoren, für die aussagekräftige Experimente durchgeführt wurden, kann sinnvolles Wissen aufgebaut werden. Qualitative Unterschiede in diesen Daten können die Abdeckung von Hypothesen- und Experimenterraum, die Durchführung kritischer Experimente zur Überprüfung einer Theorie oder auch die Korrektheit bei der Bewertung von Hypothesen angesichts empirischer Ergebnisse betreffen. Wenn sich Dyaden durch den Austausch beim kooperativen wissenschaftlichen Problemlösen von Einzelpersonen unterscheiden, dann sollte sich dies auch auf der Ebene der Prozeßdaten niederschlagen.

Hypothese 4: Dyaden zeigen sich in ihrem Vorgehen beim wissenschaftlichen Problemlösen Einzelpersonen überlegen. Dies schlägt sich in einer höheren Qualität der Prozeßdaten bei Dyaden verglichen mit Einzelpersonen nieder.

Die genannten Prozeßdaten sind ein Indiz für kooperative und kognitive Prozesse, die sich am offensichtlichsten in den verbalen Daten manifestieren. Hier zeigen sich auch die Merkmale, auf welche der Erfolg beim kooperativen Lernen zurückgeführt wird. Er wird unter anderem damit erklärt, daß die Lernenden durch die Kooperation zu Verbalisierungen und zur Elaboration des Stoffes angeregt werden, wie dies für Einzelpersonen nicht möglich ist (Bielaczyc, Pirolli & Brown, 1994; Webb, 1989). Dies schlägt sich in den Aussagen zu Hypothesen und Ergebnissen ebenso nieder wie in Fragen und Erklärungen. Dyaden sollten gemeinsam auch zu Reflexion und Kritik auf einem höheren Niveau fähig sein als Einzelpersonen.

Hypothese 5: Dyaden und Einzelpersonen unterscheiden sich in ihren verbalen Äußerungen beim wissenschaftlichen Problemlösen. Bei Dyaden findet sich eine intensivere verbale Auseinandersetzung mit dem wissenschaftlichen Problemlösen sowohl auf der inhaltlichen Ebene als auch auf der Metaebene.

Aus den bisherigen Ausführungen wird deutlich, daß die verschiedenen angesprochenen Datenebenen untereinander zusammenhängen. Der Lernerfolg hängt davon ab, was während des wissenschaftlichen Problemlösens im simulierten Labor geschah, und das wird wiederum davon beeinflusst, ob Dyaden oder Einzelpersonen dies tun. Solche Zusammenhänge werden in der Literatur zwar postuliert, wurden empirisch jedoch selten nachgewiesen. Dies hängt unter anderem mit methodischen Problemen zusammen. Lassen sich die verschiedenen Datenebenen bei Einzelpersonen einander direkt zuordnen, so ist dies bei den Daten der Dyaden nicht möglich. Jede der beiden Versuchspersonen innerhalb einer Dyade absolviert individuell Vor- und Nachtests sowie die Fragebögen und Tests der begleitenden Diagnostik. Die Prozeßdaten generieren jedoch beide Personen innerhalb einer Dyade gemeinsam. Die Logfiledaten sind als gemeinsames Produkt der beiden Versuchspersonen zu betrachten, und die verbalen Daten lassen sich zwar den

jeweiligen Sprechern zuordnen, ihre Entstehung ist jedoch interdependent. Hier stellt sich die Frage, inwiefern eine Zuordnung von Testdaten und Prozeßdaten sinnvoll ist und wie sie gegebenenfalls aussehen könnte. Auf jeden Fall sind bei einer solchen Zuordnung schwächere Zusammenhänge zu erwarten als bei Einzelpersonen, da die Prozeßdaten einer Dyade immer den Einfluß beider Versuchspersonen abbilden. Werden gar die Durchschnittswerte der beiden an einer Dyade beteiligten Versuchspersonen für weitere Vergleiche herangezogen, so ist in den Daten der Effekt einer Tendenz zur Mitte unvermeidlich.

Weitere Fragen ergeben sich im Zusammenhang mit den Daten aus der begleitenden Diagnostik. Sie werden zur Kontrolle der experimentellen Daten erhoben, können jedoch auch wertvolle inhaltliche Hinweise enthalten: Gibt es einen Zusammenhang zwischen der Einstellung zum kooperativen Lernen und dem Verhalten während des kooperativen wissenschaftlichen Problemlösens? Zeigt sich ein Zusammenhang von Vorwissen und Intelligenz mit dem Lernerfolg? Stehen Vorkenntnisse im Umgang mit Computern im Zusammenhang mit der Beurteilung der Lernumgebung oder gar mit dem Lernerfolg? Bei der Fülle der erhobenen Daten lassen sich viele offene Fragen formulieren, die neben dem experimentellen Ansatz der vorliegenden Untersuchung weitere Ergebnisse von explorativem Charakter versprechen.

7

virtue: das simulierte psychologische Labor

Mit *virtue* (der Name steht für „virtuelle Experimente“) wurde eine offene, internetbasierte Lernumgebung geschaffen, um entdeckendes Lernen für den Erwerb wissenschaftlicher Fertigkeiten zu fördern. Zielgruppe für *virtue* sind Psychologiestudierende im zweiten Studienjahr, die am Projektseminar beziehungsweise am Experimentalpraktikum teilnehmen. In *virtue* können Studierende selbstständig Experimente zum wahrnehmungspsychologischen Phänomen der visuellen Suche planen und durchführen. Ihre Aufgabe ist es, die vorgegebene Theorie der visuellen Suche zu überprüfen und die den empirischen Ergebnissen zugrunde liegenden theoretischen Gesetzmäßigkeiten zu entdecken.

In diesem Kapitel wird zunächst der Gegenstandsbereich des virtuellen Labors dargestellt. Es folgt ein Überblick über dessen Gestaltung und technische Implementierung. Den Abschluß bildet der Bericht über eine experimentelle Evaluation des Hypothesenscratchpads in *virtue*.

7.1 Gegenstandsbereich visuelle Suche

Wie findet eine Person ihren Freund bei einem Fußballspiel in einer riesigen Menge von Zuschauern wieder? Wie kann man ein getarntes Tier entdecken? In beiden Fällen soll ein bestimmtes Zielobjekt in einer visuellen Anordnung gefunden werden, deren Elemente Ähnlichkeiten mit dem Zielobjekt aufweisen. Dieser Prozeß wird als visuelle Suche bezeichnet. Sie stellt einen wichtigen Aspekt der alltäglichen Wahrnehmung dar.

Der Gegenstandsbereich für *virtue* mußte einigen Kriterien entsprechen. Er durfte nicht im Curriculum des Grundstudiums enthalten sein, da die Studierenden kein bereichsspezifisches Vorwissen in die Experimente einbringen sollten. Für den Gegenstandsbereich sollte mindestens eine empirisch belegte Theorie existie-

ren. Die empirische Untersuchung des Gegenstandsbereiches sollte streng experimentell möglich sein und zwar mit mehreren frei kombinierbaren Faktoren, von denen einige für die Theorie relevant sind, andere jedoch nicht. Die statistischen Effekte müssen klar zu erkennen sein. Die Einführung in den Gegenstandsbereich und dessen Theorie darf nicht zu kompliziert sein, da sie für alle Studierenden gut und möglichst ohne Rückfragen zu verstehen sein sollte und nicht zu viel Zeit in Anspruch nehmen darf.

Zur visuellen Suche gibt es zahlreiche Modelle (Doshier, 1998, Wolfe, Cave & Franzel, 1989). Als theoretische Grundlage für *virtue* wurde die Merkmals-Integrations-Theorie von Treisman und Gelade (1980) herangezogen (Abbildung 3). Nach dieser Theorie läuft die visuelle Wahrnehmung in zwei Stufen ab. Die präattentive, parallele Stufe operiert sehr schnell und über das gesamte visuelle Feld.

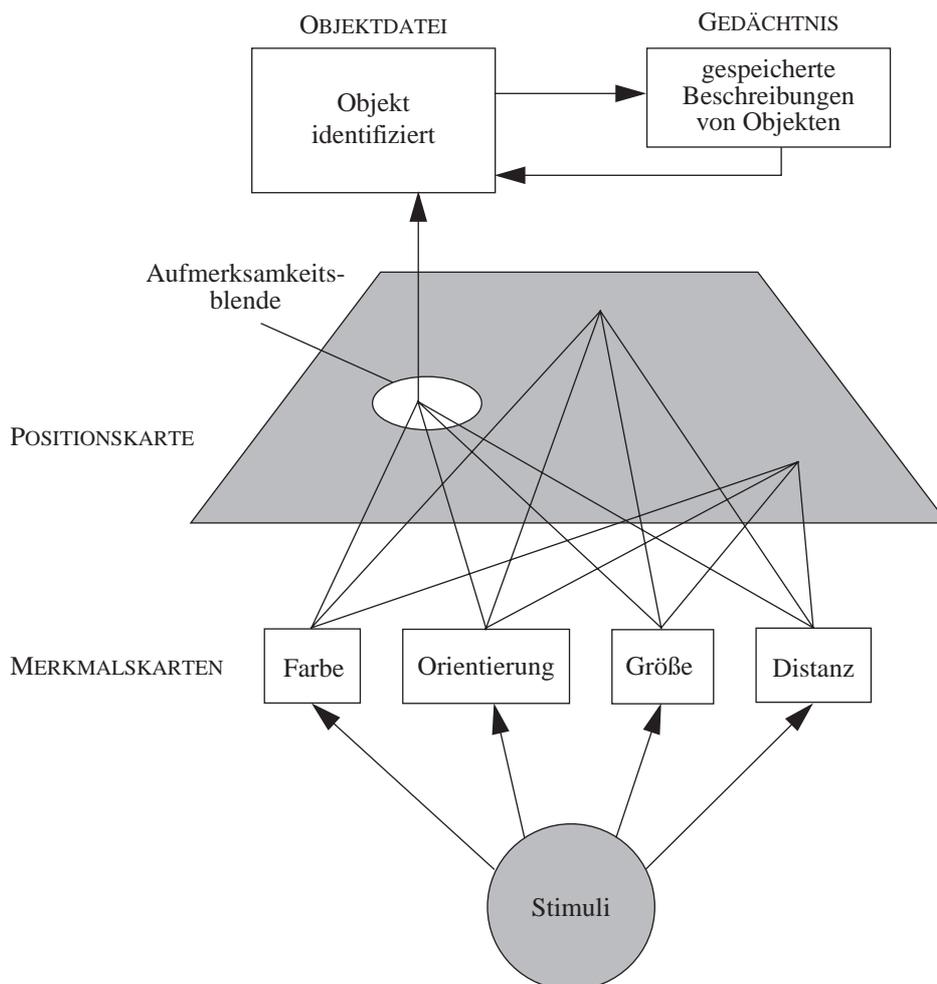


Abbildung 3. Schematische Darstellung der Merkmals-Integrations-Theorie von Treisman (nach Posner & Raichle, 1996).

Hier werden Elementarmerkmale eines visuellen Objekts wie Farbe, Orientierung, Größe oder räumliche Distanz auf einzelnen Merkmalskarten gleichzeitig kodiert. Diese erste Stufe verläuft unbewußt und entzieht sich der willentlichen Kontrolle. Die zweite Stufe verläuft seriell und ist aufmerksamkeitsgesteuert. Die Elementarmerkmale werden nun zu einem visuellen Objekt zusammengesetzt. Dabei richtet sich die Aufmerksamkeit innerhalb der Positionskarte jeweils auf die Position des aktuellen Objektes. Da die Ressourcen für die Aufmerksamkeit begrenzt sind, wird jedes Objekt einer komplexen visuellen Anordnung nacheinander einzeln verarbeitet. Wurde das Objekt korrekt registriert, so kann das Objekt wahrgenommen, als Einheit im Gedächtnis gespeichert und gegebenenfalls mit dem Zielobjekt verglichen werden.

Untersucht wird die visuelle Suche mit folgendem Paradigma: Die Versuchspersonen bekommen zunächst ein Zielobjekt (*target*) präsentiert. Anschließend sehen sie eine visuelle Anordnung (*display*) mit mehreren Objekten. Nur ein Teil der präsentierten Anordnungen enthält einen Zielreiz. Aufgabe der Versuchspersonen ist es, über zwei Tasten so schnell wie möglich anzugeben, ob das Zielobjekt auf der visuellen Anordnung vorhanden ist oder nicht. Experimentell variiert werden die oben genannten Elementarmerkmale der Objekte auf der visuellen Anordnung sowie die Anzahl der Objekte (*set size*). Als abhängige Variablen werden vor allem die Reaktionszeit, aber auch die Fehlerrate gemessen.

Unterscheidet sich ein Zielobjekt nur durch ein einziges Merkmal von den restlichen Objekten (Distraktoren) auf der Anordnung, so sticht dieses sofort ins Auge. Das heißt, es kann ohne aufmerksamkeitsgesteuerte, serielle Suche identifiziert werden. Ein Beispiel wäre die Suche nach einem magentafarbenen senkrechten Rechteck als Zielobjekt zwischen lauter grünen senkrechten Rechtecken als Distraktoren (Abbildung 4). Dieser Fall wird als Merkmalsuche bezeichnet. Mit steigender Anzahl an Distraktoren auf der Anordnung ergibt sich höchstens eine minimale Erhöhung der Reaktionszeit.

Unterscheidet sich das Zielobjekt andererseits durch eine spezifische Kombination von zwei oder mehreren Elementarmerkmalen von den Distraktoren, so ist ein unmittelbares Urteil darüber, ob es sich bei einem Objekt um das Zielobjekt handelt, nicht möglich. Ein Beispiel dafür wäre die Suche nach einem magentafar-

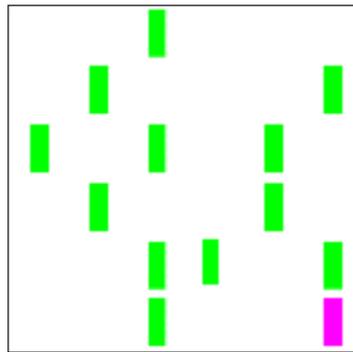


Abbildung 4. Visuelle Anordnung zur Merkmalsuche.

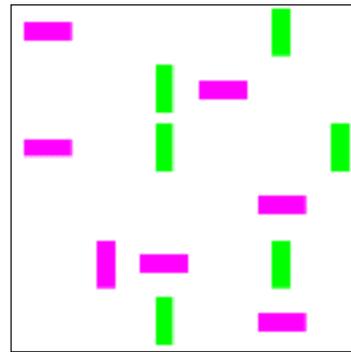


Abbildung 5. Visuelle Anordnung zur Verbindungssuche.

benen senkrechten Rechteck als Zielobjekt zwischen magentafarbenen waagrechten und grünen senkrechten Rechtecken (Abbildung 5). Jeweils ein Elementarmerkmal haben Zielobjekt und Distraktor gemeinsam, die Farbe oder die Position. Die Informationen über die Merkmale des Objekts sind auf verschiedenen Merkmalskarten gespeichert und müssen erst in einem seriellen Prozeß miteinander verknüpft werden, bevor eine Identifikation als Zielobjekt oder Distraktor möglich ist. Dieser Fall heißt Verbindungssuche. Dadurch, daß jedes Objekt der Anordnung einzeln kodiert werden muß, nimmt die Reaktionszeit mit steigender Anzahl der Objekte auf der Anordnung zu. Die Vorhersagen der Merkmals-Integrations-Theorie sind in Abbildung 6 zusammenfassend dargestellt.

Datengrundlage für das simulierte Labor *virtue* bildet ein Experiment von Reijnen (2001). Es flossen die Daten von $N = 47$ Personen im Alter von 18 bis 28 Jahren ein. Für das Experiment wurden folgende Faktoren und Faktorstufen vari-

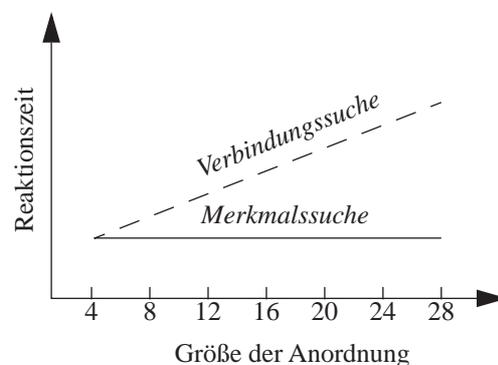


Abbildung 6. Vorhersagen der Merkmals-Integrations-Theorie.

iert: (1) *Suchtyp* mit den Stufen Merkmalssuche und Verbindungssuche – während Anordnungen zur Merkmalssuche nur aus senkrechten Rechtecken von zweierlei Farbe bestanden, enthielten Anordnungen zur Verbindungssuche senkrechte und waagrechte Rechtecke mit zwei verschiedenen Farben; (2) *Größe der Anordnung* mit 7, 13, 19, 25 und 31 Objekten. (3) und (4): Die Farbe der Objekte wurde in zweierlei Hinsicht variiert, die *Farbkombination* und die *Farbähnlichkeit*. Die beiden Farbkombinationen waren Karmin-Magenta-Grün und Eisblau-Blau-Gelb. Als farbähnliche Kombinationen kamen Anordnungen mit Objekten in den Farben Magenta und Karmin oder Blau und Eisblau zum Einsatz. Anordnungen mit farbunähnlichen Kombinationen enthielten Objekte in den Farben Magenta und Grün oder Blau und Gelb. Die Zielobjekte hatten entweder die Farbe Magenta oder die Farbe Blau. (5) *Exzentrizität*: Die Entfernung des Zielreizes von der Mitte der Anordnung wurde in drei Stufen variiert: klein, mittel, groß. Die visuelle Anordnung wurde in drei konzentrische Zonen aufgeteilt, innerhalb derer das Zielobjekt positioniert wurde. Um für diesen Faktor eine zuverlässige Messung zu gewährleisten, wurde das Zielobjekt als Fixationspunkt dort angezeigt, wo die Mitte der nächsten visuellen Anordnung präsentiert wurde. Von den visuellen Anordnungen enthielten 25% kein Zielobjekt. Insgesamt wurden 320 Anordnungen in allen Faktorkombinationen in einer zufälligen Reihenfolge präsentiert. Alle Daten wurden also in Meßwiederholung erhoben.

Die Effekte für die Faktoren Suchtyp und Größe der Anordnung bestätigten die Vorhersage der Merkmals-Integrations-Theorie. Darüber hinaus fand Reijnen (2001) bedeutsame Effekte für die Farbähnlichkeit und die Exzentrizität. Sowohl für die Merkmalssuche als auch für die Verbindungssuche wurden bei ähnlichen Farben längere Reaktionszeiten gemessen als bei unähnlichen Farben, wobei der Effekt bei der Verbindungssuche deutlicher zutage trat. Auch bei der Exzentrizität zeigten sich stärkere Effekte für die Verbindungssuche als für die Merkmalssuche. Eine Interaktion ergab sich mit dem Faktor Farbähnlichkeit: Bei ähnlichen Farbkombinationen wirkte sich die Exzentrizität stärker aus als bei unähnlichen Farbkombinationen. Für den Faktor Farbkombination wurden keine systematischen Effekte gefunden.

7.2 Gestaltung der Lernumgebung *virtue*

Bei der Entwicklung von *virtue* wurde versucht, eine klare, sparsame und damit lernförderliche und ansprechende Gestaltung zu verwirklichen (Blömeke, 2003; Mayer, 2001). Das virtuelle Labor *virtue* besteht aus einem Hauptbildschirm, über den verschiedene Funktionen in jeweils eigenen Fenstern aufgerufen werden können. Bei einer Anmeldeprozedur wird eine Benutzeridentität vergeben und zwischen der englischen und der deutschen Version von *virtue* gewählt. Mit der Benutzeridentität ist eine wiederholte Anmeldung möglich. Die Daten bisheriger Hypothesen und Experimente werden gespeichert und sind für den Benutzer wieder verfügbar.

Nach der Anmeldung präsentiert das System zunächst automatisch die Instruktion. Von dort wird man auf den Hauptbildschirm weitergeleitet (Abbildung 7). Über diesen sind fünf wesentliche Funktionen mit ihren kognitiven Werkzeugen (Lajoie, 1993) verfügbar: (1) Theorie nachschauen, (2) Hypothesen formulieren und verwalten, (3) Versuchsplan aufstellen, (4) Experiment durchführen, (5) Ergebnisse anzeigen und interpretieren.

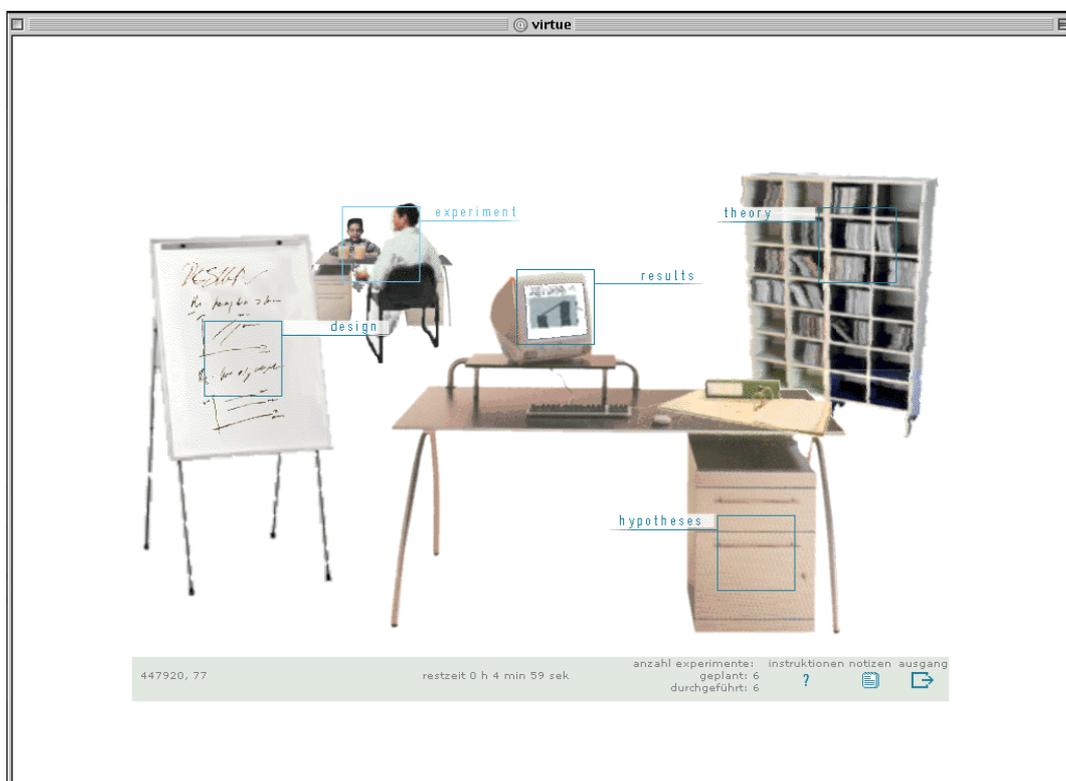


Abbildung 7. Hauptbildschirm von *virtue*.

Die Funktion zum Nachschauen der Theorie enthält denselben Text, wie er den Versuchspersonen im Vorfeld der experimentellen Untersuchung als Instruktion auf Papier präsentiert wurde. So wurde vermieden, daß lange Texte am Bildschirm gelesen werden müssen (Niegemann, 2001), zum Nachschlagen war die Theorie dennoch online verfügbar. Einige Begriffe (beispielsweise Zielreiz oder Distraktor) sind sowohl in der Theorie als auch in anderen Fenstern als Hypertextlinks erkennbar. Werden diese angeklickt, so erscheint eine virtuelle Karteikarte mit einer Definition dieser Begriffe. Auf diese Weise wurde ein spezifisches Online-Lexikon zur Verfügung gestellt.

Eine Leiste am unteren Rand des Bildschirms zeigt als weitere Informationen an, wieviele Versuchspersonenstunden aus dem großzügig bemessenen, aber dennoch begrenzten Pool noch verfügbar sind und wieviele Experimente bereits geplant und durchgeführt wurden. Außerdem kann man über diese Leiste einen virtuellen Notizblock und eine Hilfefunktion aufrufen und sich aus dem System wieder ausloggen. In den Funktionsfenstern sind als Navigationsfelder Pfeile für Vor- und Zurückgehen sowie Häkchen und Kreuze für definitive Eingabe beziehungsweise Schließen eines Fensters vorgesehen.

7.2.1 Hypothesenscratchpad

Als Hypothesenscratchpad wurde ein kognitives Werkzeug entwickelt, das die Lernenden beim Aufstellen von Hypothesen unterstützen soll. Wird das Hypothesenscratchpad aufgerufen, so erscheint zunächst das Hypothesenarchiv (Abbildung 8). In ihm sind die bisher aufgestellten Hypothesen aufgelistet mit den Faktoren, auf

Nr.	Faktoren	Beschreibung	Status
1	similarity_of_colours, set_size	set-size - similarity	beibehalten
2	similarity_of_colours, set_size	set-size, similarity - combination	verworfen
3	color_combination	iiii	getestet
4	similarity_of_colours, color_combination	combination - similarity	getestet
5	set_size, eccentricity	set size	beibehalten
6	set_size, eccentricity	eccentricity	beibehalten

[neue Hypothese](#)

close window X

Abbildung 8. Hypothesenarchiv von *virtue*.

die sie sich beziehen, ihrem Kurztitel und dem Status (ungetestet, getestet, beibehalten oder verworfen). Über einen Link läßt sich der Volltext der Hypothesen aufrufen. Der Text einer einmal eingegebenen Hypothese kann nicht mehr editiert und verändert werden.

Vom Hypothesenarchiv aus wird das Fenster für die Formulierung einer neuen Hypothese aufgerufen (Abbildung 9). Es handelt sich um ein semi-strukturiertes Scratchpad (van Joolingen & de Jong, 1991). Hier wird das Formulieren einer Hypothese in drei Schritte gegliedert. Zunächst wählt eine Lernende aus einem Menü die Faktoren aus, über die sie in ihrer Hypothese eine Aussage machen möchte. Im zweiten Schritt formuliert sie die Hypothese frei als Fließtext. Dieser Schritt ist in Abbildung 9 festgehalten, im oberen Teil des Fensters sieht man die im ersten Schritt ausgewählten Faktoren. Schließlich muß noch ein Titel für die Hypothese formuliert werden, der im Hypothesenarchiv erscheint. Solange eine Hypothese noch nicht endgültig verabschiedet ist, kann immer zu den vorhergehenden Schritten zurückgesprungen werden, und die dortigen Eintragungen können revidiert werden. Eine einmal verabschiedete Hypothese kann nicht mehr verändert werden. Im Hypothesenarchiv wird sie zunächst mit dem Status „ungete-

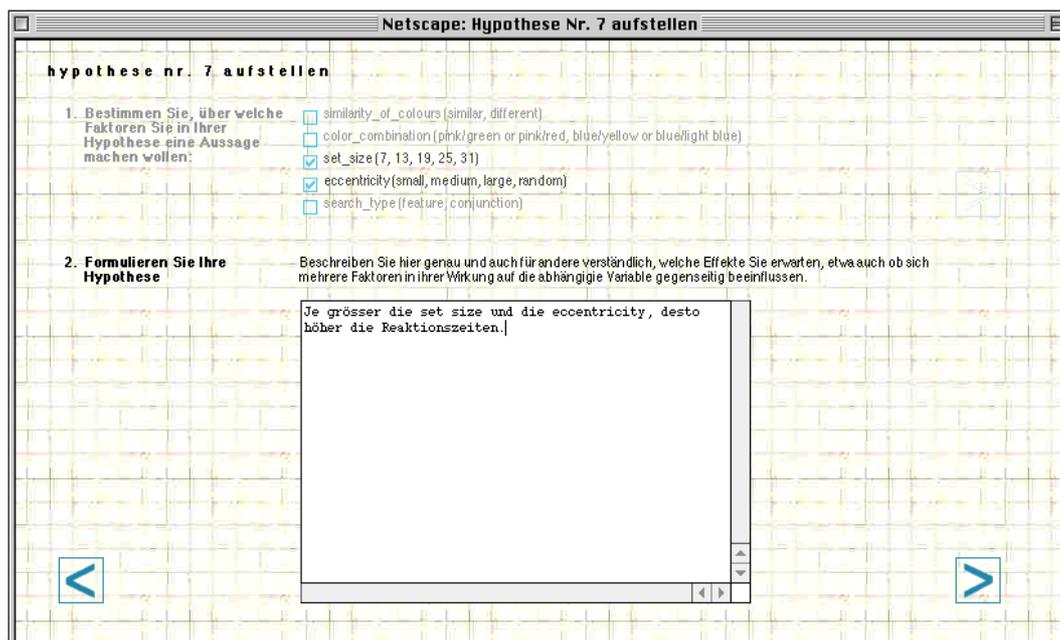


Abbildung 9. Hypothesenscratchpad von *virtue*.

stet“ abgelegt. Wurde sie dann in einem Experiment überprüft, so wechselt der Status zu „getestet“, oder die vom Benutzer vorgenommene Bewertung wird an dieser Stelle eingetragen.

7.2.2 Versuchsplanung

Ein Versuchsplan kann in *virtue* jederzeit aufgestellt werden unabhängig davon, ob eine Hypothese formuliert wurde oder nicht. Wurden bereits Hypothesen aufgestellt, so können diese im ersten Fenster zur Überprüfung ausgewählt werden. Die Versuchsplanung läuft dann in vier Schritten ab. Die ersten drei Schritte sind in Abbildung 10 durch die drei Spalten des Fensters repräsentiert. Ist ein Schritt vollzogen, so wird diese Spalte schattiert dargestellt. Zunächst werden die Faktoren festgelegt, die experimentell variiert werden sollen (linke, bereits schattierte Spalte). Dies können ein oder zwei Faktoren sein. Versuchen die Lernenden mehr als zwei Faktoren einzugeben, so erscheint ein Pop-Up-Fenster mit der Mitteilung, daß nur zwei Faktoren experimentell variiert werden können. Diese Beschränkung wurde eingeführt, da die Zielgruppe, Psychologiestudierende im zweiten Studienjahr, mit der Interpretation von Experimenten mit mehr als zwei Faktoren noch überfordert sind.

Im zweiten Schritt werden die jeweils mindestens zwei Ausprägungen des variierten Faktors beziehungsweise der variierten Faktoren bestimmt. In Schritt drei wird je eine Stufe festgelegt, auf welcher die jeweiligen nicht variierten Faktoren konstant gehalten werden.

1 Zu variiende Faktoren auswählen	2 Ausprägungen der variierten Faktoren auswählen	3 Ausprägungen der konstant gehaltenen Faktoren auswählen
<input type="checkbox"/> similarity_of_colours		<input type="radio"/> similar <input checked="" type="radio"/> different
<input type="checkbox"/> color_combination		<input checked="" type="radio"/> pink/green or pink/red <input type="radio"/> blue/yellow or blue/light blue
<input checked="" type="checkbox"/> set_size	7 13 19 25 31	
<input type="checkbox"/> search_type		<input type="radio"/> feature <input checked="" type="radio"/> conjunction
<input checked="" type="checkbox"/> eccentricity	small medium large	

Abbildung 10. Definition des Versuchsplans in *virtue*.

Der auf diese Weise definierte Versuchsplan wird auf einer neuen Seite nochmals übersichtlich tabellarisch dargestellt (Abbildung 11). Hier müssen die Studierenden auch die Anzahl der Versuchspersonen für das geplante Experiment angeben. Gleichzeitig mit der Eingabe wird im Feld daneben die verbliebene Versuchspersonenzeit berechnet. Anfangs standen den Lernenden 50 Versuchspersonenstunden zur Verfügung. Pro Versuchsperson und Zelle wurden 30 Sekunden von der Versuchspersonenzeit abgezogen. Das Zeitbudget war also großzügig bemessen und sollte die Lernenden dennoch zu einer effizienten Versuchsplanung anregen.

In *virtue* ist es auch möglich, mehrere Experimente am Stück zu planen, ohne diese zwischendurch jeweils auszuführen. Sie können dann anschließend ebenso am Stück durchgeführt werden.

The screenshot shows a Netscape browser window with the following content:

Versuchsplan

	7	19	19	25	31
small	n = 30				
eccentricity medium	n = 30				
large	n = 30				

constant factors

similarity_of_colours: different
 color_combination: pink/green or pink/red
 search_type: conjunction

Sie haben einen zweifaktoriellen Versuchsplan aufgestellt. Ihre abhängige Variable ist, wie in allen Experimenten zur visuellen Suche, die Reaktionszeit der Versuchspersonen für die Entscheidung, ob auf dem Display ein Zielreiz vorhanden ist oder nicht.

Mit diesem Versuchsplan können Sie untersuchen, ob die Faktoren *set_size* und *eccentricity* oder deren Kombination einen Einfluss auf die Reaktionszeit der Versuchspersonen hat.

Bestimmen Sie nun die Größe der Stichprobe für Ihr Experiment. Bitte beachten Sie dabei: Alle Experimentalbedingungen werden in Messwiederholung erhoben. Das heisst, jede Versuchsperson produziert Daten für jede Zelle des Versuchsplans. Daher verändert sich die Stichprobengröße *n* für alle Zellen parallel. Pro Experimentalbedingung verbraucht eine Versuchsperson 30 Sekunden Ihres Zeitbudgets.

Wählen Sie hier Ihre Stichprobengröße:

RESTZEIT: 34 : 00 : 00

Abbildung 11. Festlegung des Stichprobenumfangs in *virtue*.

7.2.3 Ergebnisse

Damit die Ergebnisse eines geplanten Experiments ausgewertet werden können, müssen die Lernenden es zuerst durchführen. Dazu rufen sie ein eigenes Fenster auf, in dem alle geplanten und noch nicht durchgeführten Experimente aufgelistet sind. Das durchzuführende Experiment muß angeklickt werden. Nach Bestätigung der Auswahl zeigt ein Fortschrittsbalken die Datenerhebung an. Je komplexer das experimentelle Design ist und je mehr Versuchspersonen für die visuelle Suche vorgesehen sind, desto länger dauert die Datenerhebung. Pro Versuchsperson für die visuelle Suche und Zelle werden 100 Millisekunden berechnet. Dies entspricht

nicht der tatsächlichen Berechnungszeit der Ergebnisse, diese ist wesentlich kürzer. Stattdessen soll hiermit der Aufwand symbolisiert werden, den eine tatsächliche Datenerhebung mit sich brächte.

Die Ergebnisse eines Experiments können in einem eigenen Fenster angeschaut werden (Abbildung 12). In der linken Spalte des Ergebnisfensters ist ein weiteres Archiv für die durchgeführten Experimente positioniert. Hier kann einerseits nochmals der Versuchsplan für ein Experiment eingesehen werden, andererseits werden über dieses Archiv die Experimente ausgewählt, deren Ergebnisse angezeigt werden sollen. In einem zweiten Auswahlschritt wird festgelegt, in welchem Format die Ergebnisse angezeigt werden sollen. Als Formate stehen eine Tabelle mit den deskriptiven Werten, eine Tabelle mit den Ergebnissen einer Varianzanalyse mit Meßwiederholung und ein Interaktionsplot zur Auswahl. Es können auch alle drei Formate gleichzeitig für die Darstellung gewählt werden.

Wurde bei der Versuchsplanung eine Hypothese zur Überprüfung angegeben, so erscheint unterhalb der Resultate ein weiteres Feld, in dem die Hypothese noch einmal aufgeführt wird. Darunter sind drei Möglichkeiten zur Interpretation der Ergebnisse und zur Evaluation der Hypothese vorgesehen. Die Hypothese kann beibehalten oder verworfen werden. Als dritte Möglichkeit ist die Antwort „weiß nicht“ vorgesehen. In einem Eingabefeld wird noch um die Begründung dieser Entscheidung gebeten. Das Schließen des Ergebnisfensters ist nicht möglich, ohne die vorhandenen Hypothesen bewertet zu haben.

Obwohl die einzelnen Funktionen nun in der Reihenfolge dargestellt wurden, wie sie für einen experimentellen Zyklus sinnvoll sein können, handelt es sich um eine offene Lernumgebung, in der die Reihenfolge nirgends vorgeschrieben ist. Es können mehrere Hypothesen nacheinander aufgestellt werden, ohne daß Experimente dazu geplant würden, und es können Experimente ohne Hypothesen durchgeführt werden. Die kognitiven Werkzeuge strukturieren allerdings die Reihenfolge innerhalb einer Funktion, was eine kognitive Entlastung zum Ziel hat. Ohne Experimente können freilich keine Ergebnisse abgerufen werden. Theorie, Glossar, Hilfs- und Notizfunktionen sind jederzeit verfügbar. Auf diese Weise sind die Versuchspersonen in ihrer Exploration frei und nur durch die Instruktion zur Untersuchung der Merkmals-Integrations-Theorie angeleitet.

Netscape: results

EXPERIMENT 7
Deskriptive Statistik

eccentricity	small			medium			large		
	M	SD	n	M	SD	n	M	SD	n
7	838.33	219.80	30	811.33	188.43	30	834.00	156.56	30
13	796.67	208.92	30	970.33	227.24	30	972.67	211.99	30
19	826.00	243.78	30	1024.33	274.78	30	1036.67	306.11	30
25	948.67	291.67	30	1007.00	197.47	30	1072.33	245.90	30
31	1010.33	341.16	30	1178.00	301.28	30	1247.33	349.52	30

Varianzanalyse (Messwiederholung)

source	df	F
between subjects (S)	29	(380528.09)
within subjects:		
set_size (A)	4	27.83***
eccentricity (B)	2	18.17***
A x B	8	2.11*
A x S	116	(45049.70)
B x S	58	(49961.82)
A x B x S	232	(45571.41)

Anmerkungen: Werte innerhalb der Klammern repräsentieren die mittleren Quadratwerte.
*p<.05, **p<.01, ***p<.001

INTERAKTIONSPLOT

zu bewertende hypothesen

Hypothese 7 **set size - eccentricity**
Je grösser die set size und die eccentricity, desto höher die Reaktionszeiten.

Wollen Sie aufgrund der vorliegenden Resultate die Hypothese 7...

...beibehalten ...verwerfen weiss nicht

Vor allem ab 25 Objekte auf dem Display wirkt sich die eccentricity stark auf die Reaktionszeit aus. Auch ansonsten entsprechen die Ergebnisse unseren Erwartungen.

Bewerten Sie vor dem Schliessen des Fensters all Ihre mit den Experimenten verbundenen obigen Hypothesen!

submit

Abbildung 12. Ergebnisfenster mit Hypothesenbewertung in *virtue*.

7.3 Technische Implementierung

Die Lernumgebung *virtue* lässt sich über einen Internet-Browser (etwa Netscape ab Version 4.0) über den URL <http://virtue.psychu.unibas.ch/> aufrufen und ist als Client/Server-Applikation implementiert. Für die Benutzer sind daher weder Hard-

ware- und Softwareanschaffungen noch gesonderte Installationen (Thin-Client) nötig. Zudem müssen keine programmspezifischen Navigationsfertigkeiten erworben werden, da Studierenden die Bedienung eines Browsers vertraut ist.

Betrieben wird *virtue* auf einem einfachen Web-Server (Apache 1.3.9). Es wurde in Form verschiedener Perl-Skripten (Perl 5.0) implementiert. Die Kommunikation mit dem Client erfolgt über das Internet. Eine genaue Beschreibung der Kommunikation findet sich bei Rügsegger (2002). Jede Client-Anfrage wird auf dem Server mit Zeitstempel in einem Logfile dokumentiert. Dabei kann es zu leichten Verzögerungen kommen. Da für die Art der untersuchten Lernprozesse jedoch keine Zeitmessungen im Millisekundenbereich relevant sind, ist die Genauigkeit der Messung für diesen Zweck ausreichend.

Die Daten für die simulierten Experimente zur visuellen Suche sind in einem Textfile abgelegt, das nach Bedarf ausgelesen wird. Für jede der 160 möglichen Stufenkombinationen existieren für $N = 47$ Versuchspersonen Werte, die in der Untersuchung von Reijnen (2001) ermittelt wurden. Wird in *virtue* ein Experiment durchgeführt, so werden entsprechend der angegebenen Versuchspersonenzahl aus diesen 47 Werten Reaktionszeiten ausgelesen.

Da die Daten für den Gegenstandsbereich in einem Textfile abgelegt sind, ist dessen Austausch ohne größeren technischen Aufwand möglich. Auf diese Weise können auch neue Gegenstandsbereiche in *virtue* integriert werden, wodurch sich der Einsatzbereich deutlich erweitern läßt.

Die Protokolle über die Benutzer werden in zwei Dateien abgelegt, in einem sogenannten Datenfile und in einem Logfile. Das Datenfile enthält Informationen über den aktuellen Zustand der Hypothesen und Experimente inklusive deren Beschreibungen, Ergebnisse, Bewertungen und Relationen zueinander. Im Logfile wird jeder Serverkontakt eines Benutzers aufgezeichnet. Auf diese Weise wird dokumentiert, wie lange ein einzelner Schritt im System dauert.

Auf die Verwendung einer komplexen Datenbankstruktur wurde bewußt verzichtet, da Perl-Skripten für die Arbeit mit Textfiles optimiert sind. Gleichzeitig ermöglicht es diese Art der Implementation, daß Versuchsleitung und Lehrpersonen die virtuellen Experimente in Echtzeit verfolgen können.

Das System verhält sich server-seitig ausgesprochen stabil. Während der Untersuchungen ereigneten sich beim Server keinerlei Zwischenfälle. Auch client-seitig kann man von einer akzeptablen Stabilität sprechen. In über vierzig Stunden kam es zu 13 Zwischenfällen. Diese konnten jedoch jeweils durch kurze Interventionen der Versuchsleitung behoben werden, und die Untersuchungen konnten in jedem dieser Fälle zu Ende geführt werden.

7.4 Wirksamkeit des Hypothesenscratchpads

In einer eigenen Studie untersuchte Rüeegsegger (2002) die Wirksamkeit des Hypothesenscratchpads in *virtue*. Dabei kam neben der oben beschriebenen Variante auch ein voll strukturiertes Scratchpad zum Einsatz. Eine Hypothese wird hier in mehreren Schritten komplett über die Auswahl vorformulierter Elemente definiert. Einzig Vorhersagen über Interaktionen müssen die Lernenden frei formulieren, falls sie Aussagen darüber machen wollen.

Bei dem Experiment wurden drei Gruppen ($N = 35$) von Psychologiestudierenden im zweiten Studienjahr an der Universität Basel untersucht. Sie nahmen freiwillig an dem Experiment teil und bekamen eine Bestätigung für die Versuchsteilnahme, die sie sich zusammen mit anderen solchen Bestätigungen als Studienleistung anrechnen lassen konnten. Die Versuchspersonen in jeder Gruppe arbeiteten jeweils mit einer anderen Version von *virtue*: mit strukturiertem Scratchpad, mit semi-strukturiertem Scratchpad (demselben wie in der vorliegenden Arbeit) oder ohne Scratchpad. Ihre Aufgabe bestand darin, die in der Instruktion präsentierte Merkmals-Integrations-Theorie (Treisman, 1980) zu überprüfen. Alle Versuchspersonen arbeiteten individuell mit *virtue* und wurden zum lauten Denken instruiert, so daß die Hypothesen aus der Gruppe ohne Scratchpad den auf Video aufgezeichneten Protokollen entnommen werden konnten. Der Lernerfolg wurde durch parallele Vor- und Nachtests erhoben.

Während Rüeegsegger (2002) für die drei Experimentalgruppen keine verbesserte Lernleistung nachweisen konnte, zeigten sich deutliche Unterschiede bezüglich der Prozeßdaten: Die beiden Gruppen mit den Scratchpads beschäftigten sich deutlich länger ($M = 60$ min.) mit *virtue* als die Gruppe ohne Scratchpad ($M = 45$ min., nach Kruskal und Wallis: $H(2) = 7.681$, $p < .05$). Zeigten sich hinsichtlich der Anzahl durchgeführter Experimente keine Unterschiede zwischen den

Gruppen, so stellte die Gruppe mit strukturiertem Scratchpad die meisten Hypothesen auf ($M = 5.3$), gefolgt von der Gruppe mit semi-strukturiertem Scratchpad ($M = 3.8$). Die wenigsten Hypothesen fanden sich bei der Gruppe ohne Scratchpad ($M = 1.3$, $F(2,32) = 24.34$, $p < .01$). Dies hat Auswirkungen auf die Exploration des Hypothesenraums. Er wird in den beiden Gruppen mit Scratchpad vor allem in Bezug auf die Haupteffekte weiträumiger abgedeckt als in der Gruppe ohne Scratchpad ($F(2,32) = 19.6$, $p < .01$). Alle Gruppen deckten die Haupteffekte im Experimenterraum gleichmäßig und fast vollständig ab. Die Gruppe ohne Scratchpad untersuchte jedoch signifikant mehr Interaktionen als die beiden anderen Gruppen ($F(2,32) = 3.6$, $p < .05$). Hier zeigten sich Effekte analog zu Klahr und Dunbar (1988): Hypothesengeleitete Versuchspersonen führten weniger Experimente durch als explorativ vorgehende Versuchspersonen.

Das Experiment von Rüegegger (2002) konnte also eine positive Wirkung der Scratchpads auf die Formulierung von Hypothesen aufzeigen. Ein Lerneffekt konnte nicht nachgewiesen werden. Dies wird von Rüegegger hauptsächlich auf die mangelnde Qualität des Meßinstruments für die parallelen Vor- und Nachtests zurückgeführt. Ein weiterer möglicher Grund für diese Ergebnisse könnte die mangelnde Motivation der Teilnehmenden gewesen sein. Diese wurde im Experiment leider nicht gemessen, aber vom Versuchsleiter berichtet. Ebenfalls für diese Erklärung spricht die Tatsache, daß das Experiment vielen Teilnehmenden als ausgesprochen anstrengend erschien. Dies wurde im abschließenden Fragebogen nämlich erhoben ($M = 4.1$ auf einer fünfstufigen Likertskala). Das Experiment dauerte mit allen Tests auch zweieinhalb Stunden, wofür die Studierenden sich mit den Bestätigungen für die Teilnahme wahrscheinlich nicht entsprechend entschädigt fühlten. Dies könnte sich negativ auf die Resultate des Nachtests ausgewirkt haben.

8

Methode

Die vorliegende Untersuchung zum wissenschaftlichen Problemlösen und Lernen in einem simulierten psychologischen Labor wurde mit der in Kapitel 7 beschriebenen Lernumgebung *virtue* durchgeführt. Hier wird der methodische Aufbau der Untersuchung beschrieben, zunächst der Versuchsplan und das über die Lernumgebung hinaus zur Untersuchung verwendete Material. Anschließend wird die Stichprobe charakterisiert. Den Abschluß bildet eine Darstellung des Versuchsablaufes.

8.1 Versuchsplan

In dieser Studie wurden zwei Gruppen untersucht (Tabelle 2). Die eine Gruppe arbeitete beim wissenschaftlichen Problemlösen mit *virtue* individuell, die andere setzte sich aus Dyaden zusammen, deren beide Teilnehmende beim wissenschaftlichen Problemlösen jeweils kooperierten. Vor dem wissenschaftlichen Problemlösen durchliefen die Versuchspersonen jeweils individuell verschiedene Tests und Fragebögen zur begleitenden Diagnostik (für den Fragebogen siehe Anhang A). Die Theorie zur visuellen Suche bekamen sie in Papierform vorgelegt (Anhang B), ebenso die parallelen Vor- und Nachtests (Anhänge C und D). Die abschließende Bewertung der Versuchspersonen für das Problemlösen mit *virtue* wurde in einem

Tabelle 2
Versuchsplan

Bedingung	Begleit- diagnostik	Theorie visuelle Suche	Parallel- test 1	wissenschaftliches Problemlösen mit <i>virtue</i>	Parallel- test 2
Einzelpersonen $n = 12$				individuell	
Dyaden $n = 12 \times 2$		individuell		kooperativ	individuell

weiteren Fragebogen erhoben (Anhang E). Das wissenschaftliche Problemlösen mit *virtue* fand an einem Computer statt, an dem die Versuchspersonen je nach Experimentalgruppe alleine oder zu zweit saßen.

8.2 Material

In diesem Abschnitt werden die Materialien beschrieben, die für die vorliegende Untersuchung zusätzlich zum simulierten psychologischen Labor *virtue* (siehe Kapitel 7) verwendet wurden. In den parallelen Vor- und Nachtests wurde der Lernfortschritt durch das wissenschaftliche Problemlösen mit *virtue* erhoben. Zur Kontrolle verschiedener Einflußvariablen wie Intelligenz und Einstellungen zu bestimmten Lernformen wurden verschiedene Tests beziehungsweise Fragebögen durchgeführt.

8.2.1 Paralleltests zum bereichsspezifischen und bereichsübergreifenden Lernen beim wissenschaftlichen Problemlösen

Der Wissenszuwachs beim Problemlösen mit *virtue* wurde durch parallele Vor- und Nachtests erhoben (Anhänge C und D). Beim Lernen mit Simulationen können zwei Arten von Wissen beziehungsweise Fertigkeiten erworben werden: bereichsspezifische und bereichsübergreifende. Beim wissenschaftlichen Problemlösen mit *virtue* entspricht bereichsspezifisches Wissen den Inhalten über die visuelle Suche, der Merkmals-Integrations-Theorie (Treisman & Gelade, 1980) und der Wirkung der in Kapitel 7 beschriebenen Variablen aus dem Experiment von Reijnen (2001). Dieses Wissen läßt sich gut beschreiben und abgrenzen, es ist daher leicht verbalisierbar und kann mit einem Papier-und-Bleistift-Test abgefragt werden.

Das bereichsspezifische Wissen wurde ausschließlich in Multiple-Choice-Items abgefragt. Mit den Items 1-6 wurde das Verständnis für die in der Instruktion erklärten theoretischen Begriffe aus dem Gegenstandsbereich erfaßt (Untertest Theorie). Die Items 7-10 erheben das Wissen über die beiden zentralen Faktoren der Merkmals-Integrations-Theorie Suchtyp und Größe der Anordnung und deren empirische Wirkungen (Untertest Empirie). Dabei handelt es sich bei Item 10 um eine Auswahl zwischen drei Schaubildern, während die Alternativen in allen anderen Aufgaben schriftlich formuliert waren. Im Nachtest fragten fünf Items zusätzlich das Wissen über die in der Instruktion nicht näher beschriebenen Faktoren Exzentrizität, Farbähnlichkeit und -kombination ab. Diese Items wurden im Vortest

nicht vorgegeben, da die Testprozedur nicht unnötig verlängert werden sollte und davon ausgegangen werden konnte, daß sie zu spezifisch sind, als daß die Versuchspersonen hierzu über Vorwissen verfügen könnten.

Als bereichsübergreifendes Wissen werden in dieser Untersuchung die Fertigkeiten zur Planung, Durchführung, Auswertung und Interpretation von psychologischen Experimenten betrachtet (Untertest Methodologie). Die Schwierigkeiten beim Erfassen von Lernprozessen für diesen Bereich wurden bereits in Kapitel 5 dargestellt. Das hier Erlernte ist großenteils nicht deklarativ, sondern eher intuitiv, schwer verbalisierbar und mit der Wahrnehmung und Repräsentation der Simulation verknüpft. Für die Messung solcher Fertigkeiten schlagen Swaak und de Jong (2001) vor, bei den Items auch mit visuellen Repräsentationen zu arbeiten, Vorhersagen formulieren zu lassen und situative Fragen zu stellen.

Für die Beschreibung der Items wird im Folgenden auf die Nummern im Vortest Bezug genommen. Parallele Items, teilweise in anderer Reihenfolge, finden sich im Nachtest. In den Items 11-13 werden allgemeine methodische Fertigkeiten überprüft. Die Versuchspersonen müssen ein Beispiel für einen Versuchsplan geben sowie Haupteffekte und Interaktionen in Interaktionsplots erkennen. Die Items 14 und 15 fragen methodische Begriffe und Zusammenhänge über Multiple-Choice-Aufgaben ab. Den Abschluß des Tests bilden zwei komplexe Anwendungsaufgaben. In der ersten werden zwei Versuchspläne als Operationalisierung derselben Fragestellung einander gegenübergestellt. Bei der einen Alternative handelt es sich um ein ausgesprochen komplexes Design, dessen Ergebnisse schwer interpretierbar wären. Die andere Alternative sieht zunächst einen zweifaktoriellen Versuchsplan vor mit der Möglichkeit für Folgeexperimente. Die Versuchspersonen sollten sich für eine der Alternativen entscheiden und ihre Entscheidung begründen. In der letzten Aufgabe werden die Daten eines Experiments und die dazu aufgestellten Hypothesen präsentiert. Die Daten stammen aus der Simulation *virtue*, werden jedoch in einem anderen inhaltlichen Kontext dargestellt. Aufgabe der Versuchspersonen war es, die Gültigkeit der Hypothesen im Lichte der empirischen Ergebnisse zu beurteilen und ihr Urteil zu begründen.

Für die Parallelisierung der Vor- und Nachtests wurde die Reihenfolge der Items teilweise geändert. Bei den Multiple-Choice-Items wurden die Antwortalternativen größtenteils ersetzt. Die letzten beiden komplexen Aufgaben wurden jeweils inhaltlich anders eingekleidet. Die Zahlen der letzten Aufgabe stammten aus einer Replikation desselben Experiments in *virtue* mit denselben statistischen Effekten und wurden um eine Dezimalstelle verschoben.

8.2.2 Leistungsdiagnostik und Fragebögen

Den Einfluß der Intelligenz, des Vorwissens und verschiedener Einstellungen auf komplexe Lernprozesse beschreiben unter anderen Goodyear et al. (1991), Schauble, Glaser et al. (1991) und Süß (1999). Um die Intelligenz und das Vorwissen der Versuchspersonen sowie einige ihrer Einstellungen zu kontrollieren, wurden mit verschiedenen Verfahren Daten dazu erhoben. In der vorliegenden Untersuchung wurden Items aus verschiedenen Skalen inklusive Fragen zu soziodemographischen Angaben zu einem Fragebogen zusammengestellt, der den Versuchspersonen gleich nach der Begrüßung vorgelegt wurde (Anhang A). Anschließend wurden Aufgaben aus dem Test für medizinische Studiengänge (Institut für Test- und Begabungsforschung, 1995) und der Raventest durchgeführt. Nach dem wissenschaftlichen Problemlösen mit *virtue* und dem Nachtest beantworteten die Versuchspersonen noch einen kurzen Fragebogen zur Evaluation des Experiments (Anhang E).

Die Skala Need for Cognition

Mit dem Fragebogen „Need for Cognition“ (Cacioppo & Petty, 1982) werden individuelle Unterschiede im Engagement und der Freude bei Denkaufgaben erfaßt. Er wurde entwickelt, um den Einfluß der Bereitschaft zur Informationsverarbeitung auf soziales Urteilen und Verhalten zu bestimmen. Für die vorliegende Untersuchung wurden 8 Items aus der deutschsprachigen Kurzfassung dieser Skala nach Bless, Wänke, Bohner, Fellhauer und Schwarz (1994) verwendet. Zur Validierung wurden von Bless et al. auch die Korrelationen zwischen Need-for-Cognition-Werten und wissenschaftlichem Interesse ($r = .55$), Leistungsmotivation ($r = .57$), Mißerfolgsmotivation ($r = -.20$) und Stimmungslage ($r = .21$) erhoben, die sich allesamt als signifikant erwiesen.

Skala zur Standfestigkeit gegenüber Gruppen und bedeutsamen anderen aus den Frankfurter Selbstkonzeptskalen

Die *Standfestigkeit gegenüber Gruppen und bedeutsamen anderen* wurde erhoben, um deren möglichen Einfluß auf die Kooperation der Dyaden zu kontrollieren. Dafür wurden drei Items aus den Frankfurter Selbstkonzeptskalen (Deusinger, 1986) entnommen. Die Versuchspersonen werden hier nach ihren Einstellungen zu sozialen Situationen befragt, insbesondere dazu, wie sie ihre eigenen Fähigkeiten einschätzen, eigene Vorstellungen und Meinungen zu äußern und durchzusetzen. Niedrige Werte auf der Skala zeigen geringe Sicherheit und hohe Ängstlichkeit und Irritierbarkeit im sozialen Kontakt mit anderen an und dürften beim kooperativen Problemlösen die Person zu zurückhaltendem Verhalten veranlassen. Ein hoher Wert auf dieser Skala signalisiert eine gewisse Standfestigkeit der betreffenden Person. Beim kooperativen Problemlösen sollte eine Person mit hohen Werten auf dieser Skala ihre Positionen in der Diskussion eindeutig vertreten und die Ansichten des Lernpartners gegebenenfalls auch in Frage stellen.

Plan- und Zielorientierung

Wissenschaftliches Problemlösen sollte planvoll und zielorientiert betrieben werden, zumindest solange dabei deduktiv vorgegangen wird. Da sich Personen in ihren Handlungen durch Pläne und Ziele interindividuell unterschiedlich stark bestimmen lassen, schlagen Frese, Stewart und Hannover (1987) das Konzept der Handlungsstile vor, konkret der Planorientierung und der Zielorientierung. Handlungsstile werden definiert als (1) Neigung, in einer bestimmten Weise zu handeln, sie seien (2) zu einem gewissen Grad erlernbar und (3) bidirektional. Ein Handlungsstil sollte an eine Situation angepaßt werden können. Frese, Albrecht, Kreuzer, von Papstein, Prümper und Schulte-Göcking (1995) führten mehrere Experimente zur Rolle von Plan- und Zielorientierung in Problemlöse- und Lernsituationen durch. Es zeigte sich, daß sich Intelligenz und Zielorientierung bei der Anwendung einer neu erlernten Software mit unklar definierten Zielen gegenseitig kompensieren. Chaotische Informationen können bei zielorientierten Personen zu einer geringeren Performanz führen. Allerdings können zielorientierte Personen mehr Nutzen aus graphischen Informationen ziehen. Planorientierung wirkt sich für einen Lernstil, der als „learning by studying“ bezeichnet wird, positiv aus, zeigt

aber keine Korrelationen bei einem „learning by doing“ genannten Lernstil. Für die vorliegende Untersuchung wurden die Items von Frese et al. (1995) direkt übernommen.

Einstellung zum kooperativen Lernen

Die Einstellung der Versuchspersonen zum kooperativen Lernen und ihre Vorerfahrung damit wurde in sechs selbst formulierten Items erhoben. Dabei sollten die Versuchspersonen unter anderem angeben, ob sie das Lernen mit anderen anregend oder eher ablenkend empfinden und wie sie den Nutzen kooperativen Lernens einschätzen.

Nachbefragung

Nach dem Nachtest wurden den Versuchspersonen abhängig von ihrer jeweiligen Zugehörigkeit zu einer der beiden Experimentalgruppen unterschiedliche Versionen eines Fragebogens vorgelegt (im Anhang E findet sich die Version für Dyaden). Beide Versionen enthielten Items zur Vorerfahrung mit Computern und zur Evaluation der Lernumgebung *virtue*. Nur Dyaden wurden danach gefragt, wie sie die Kooperation beurteilten. Diese Item dieses Fragebogens wurden in Anlehnung an Cornelius (2001) sowie Nägele, Gurtner, Tschan und Semmer (2001) formuliert.

Naturwissenschaftliches Grundverständnis und Textverständnis aus dem Test für Medizinische Studiengänge

Die spezifischen Grundfertigkeiten wissenschaftlichen Arbeitens wurden mit Aufgaben aus zwei Untertests des *Tests für Medizinische Studiengänge* (TMS, Institut für Test- und Begabungsforschung, 1995) erhoben. Diese Aufgaben erfordern kein bereichsspezifisches Vorwissen und gelten daher als kaum trainierbar. Für die Zwecke der vorliegenden Untersuchung wurden aus dem Untertest zum medizinisch-naturwissenschaftlichen Grundverständnis und aus dem Untertest zu Diagrammen und Tabellen je 5 Aufgaben mit inhaltlicher Nähe zur Psychologie gewählt.

Im Untertest zum medizinisch-naturwissenschaftlichen Grundverständnis wird überprüft, inwiefern eine Person in der Lage ist, aus einem kurzen Lehrbuchtext logische Schlüsse zu ziehen. In einer schweizerischen Stichprobe fand sich eine eindeutige Überlegenheit von Männern gegenüber Frauen bei diesem Aufga-

bentyp (Hänsgen & Spicher, 1999). Im Untertest zu Diagrammen und Tabellen wird die Fähigkeit geprüft, diese richtig zu analysieren und zu interpretieren. Auch in diesem Untertest zeigten sich in der Untersuchung von Hänsgen und Spicher die Männer den Frauen überlegen.

Da der Test öffentlich zugänglich ist und zur Vorbereitung auf den in der Schweiz nach wie vor jedes Jahr neu gestalteten Eingangstest für das Medizinstudium empfohlen wird, wurde in der vorliegenden Untersuchung erhoben, ob die Versuchspersonen bereits Vorerfahrung mit diesem Test hatten. Dies war nicht der Fall. Der Test wurde in Form von zwei Teiltests für medizinisch-naturwissenschaftliches Grundverständnis beziehungsweise Diagramme und Tabellen vorgegeben. Für die beiden Teiltests mit je 5 Aufgaben hatten die Versuchspersonen jeweils 12,5 Minuten Zeit.

Raventest

Für die Intelligenzdiagnostik fiel die Wahl auf den Raventest in Form der Advanced Progressive Matrices (Kratzmeier & Horn, 1980). Er ist ein Verfahren zur sprachfreien Erfassung des Intelligenzpotentials und differenziert auch für Menschen mit überdurchschnittlichen kognitiven Fähigkeiten, wie sie in einer Population von Studierenden zu erwarten sind. Die kognitiven Fähigkeiten werden getestet, indem die Items als Speedtest mit Zeitbegrenzung (50 Minuten) vorgegeben werden. Der Raventest wurde ausgewählt, weil er auch international anerkannt und zuverlässig ist. Für deutschsprachige Studierende wurde er von Heller, Kratzmeier und Lengfelder (1998) neu normiert.

Um die Auswertung der Daten im Rahmen unserer Studie zu optimieren, wurde für die vorliegende Untersuchung eine Computerversion des Raventests erstellt. Dadurch ändert sich am Inhalt des Tests nichts, die Items werden aber auf dem Bildschirm präsentiert: Die ausgewählte Lösung muß nicht mehr in Form eines Codes auf einem Lösungsblatt vermerkt werden, sondern wird direkt durch einen Mausklick markiert. Diese Markierung muß anschließend bestätigt werden, damit das nächste Item bearbeitet werden kann. Für diese Version liegen keine Normierungen vor. Es wurden jedoch bereits Untersuchungen zur Vergleichbarkeit von Computerversionen und Papier-und-Bleistift Versionen vorgelegt (Kubinger & Farkas, 1991; Neubauer, Urban & Malle, 1995). In beiden Studien schnitten die

Probanden, die die Computerversion absolviert hatten, signifikant schlechter ab als diejenigen, die die Papier-und-Bleistift-Version bearbeitet hatten. Dies wird zum einen damit erklärt, daß Computer einen Aufforderungscharakter in Richtung schnelles und daher weniger sorgfältiges Arbeiten hätten. Neubauer et al. fanden auch signifikant kürzere Bearbeitungszeiten für die Computerversion. Zum anderen scheint die graphische Darstellung am Computer bei einzelnen Aufgaben die Schwierigkeit zu erhöhen. Da beide Studien bereits zu Beginn der Neunziger Jahre durchgeführt wurden, wäre eine andere Erklärung, daß die Probanden mit dem Medium Computer noch nicht so vertraut waren und sich dies in den Testergebnissen niederschlägt. Neubauer et al. stellten für die Computerversion eine vergleichbare hohe Reliabilität wie für die Papier-und-Bleistift-Version des Tests fest sowie hohe Interkorrelationen zwischen beiden Testversionen. Dies zusammen mit der Tatsache, daß für diese Studie die Intelligenzmessung nicht für den Vergleich der Stichprobe mit anderen, sondern als Kontrollvariable zum Einsatz kommen sollte, ließ die Nachteile einer Computerversion gegenüber den Vorteilen der erleichterten Durchführung und Auswertung annehmbar erscheinen.

8.3 Versuchspersonen

An der Studie nahmen 39 Psychologiestudierende der Universität Basel freiwillig teil. Da der Jahrgang nur 43 Studierende umfaßte, handelt es sich quasi um eine Vollerhebung. Die Versuchspersonen wurden über Aushänge angeworben, in denen sie sich auf Listen paarweise mit einer befreundeten oder bekannten Person desselben Geschlechts eintrugen (siehe dazu Azmitia & Montgomery, 1993; Barbieri & Light, 1992; Underwood, Underwood & Wood, 2000). Dies ermöglichte einerseits die Zusammenstellung verhältnismäßig valider Dyaden. Das heißt, die Versuchspersonen kannten sich bereits vor dem Experiment, weshalb eine Kennenlernphase entfiel. Die Dyaden wurden also nicht artifiziell nur für die vorliegende Studie zusammengestellt. Andererseits konnten so negative Effekte durch gemischtgeschlechtliche Dyaden vermieden werden. Für die Experimentalgruppe der Einzelpersonen wurden paarweise angemeldete Studierende zufällig ausgewählt und einzeln untersucht, so daß auch für diese Experimentalgruppe gesichert war, daß keine Selektion aufgrund mangelnder sozialer Kontakte stattgefunden hatte.

Es nahmen insgesamt 30 weibliche und 9 männliche Studierende teil. Ihr Alter lag zwischen 21 und 40 Jahren ($M = 25.8$, $SD = 5.5$). Aufgrund technischer Ausfälle bei der Datencodierung konnten die Daten einer Dyade und einer Einzelperson nicht verwertet werden. Damit flossen letztendlich die Daten von 27 weiblichen und 9 männlichen Studierenden in die Auswertung ein. Bei der randomisierten Zuweisung der angemeldeten Dyaden auf die Experimentalgruppen (Einzelpersonen oder Dyaden) wurde darauf geachtet, daß die Proportionen bezüglich des Faktors Geschlecht gewahrt blieben. Auf diese Weise basiert die Auswertung auf Daten von 12 Dyaden (9 weibliche, 3 männliche) und 12 Einzelpersonen (9 weibliche, 3 männliche). Beide Gruppen haben ein vergleichbares Alter ($U = 110.5$, $p = .27$).

Alle Versuchspersonen studierten im 4. Semester. Sie hatten an den Veranstaltungen zur Methodenlehre und zur Statistik inklusive Varianzanalyse teilgenommen und die Klausur dazu absolviert. Auf diese Weise wurde ein vergleichbares Vorwissen für alle Teilnehmenden gewährleistet (Shapiro, 2004). Bei der Stichprobe handelt es sich annähernd um eine Vollerhebung des entsprechenden Jahrgangs von insgesamt 45 Studierenden. Die Personen, die nicht am Experiment teilnahmen, gaben Terminschwierigkeiten als Gründe an, so daß ein systematischer Ausfall wohl auszuschließen ist.

Die Teilnehmenden wurden vor dem Experiment darüber unterrichtet, eine leistungsabhängige Bezahlung von 75 bis 100 Schweizer Franken zu erhalten. Um die gemeinsame Motivation zu verstärken (Slavin, 1995), wurde die Bezahlung bei den Dyaden anhand des durchschnittlichen Punktwertes im Nachtest beider Partner ermittelt.

8.4 Versuchsablauf

Die Versuchsanordnung für Dyaden ist in Abbildung 13 dargestellt. Im Versuchsraum waren zwei Computer mit Internetanschluß aufgebaut. Für das eigentliche Experiment saßen bei Dyaden beide Versuchspersonen gemeinsam, bei Einzelpersonen nur die eine Versuchsperson am Versuchscomputer. Für die individuellen Tests wurde bei den Dyaden noch ein zweiter Testcomputer im Raum verwendet.

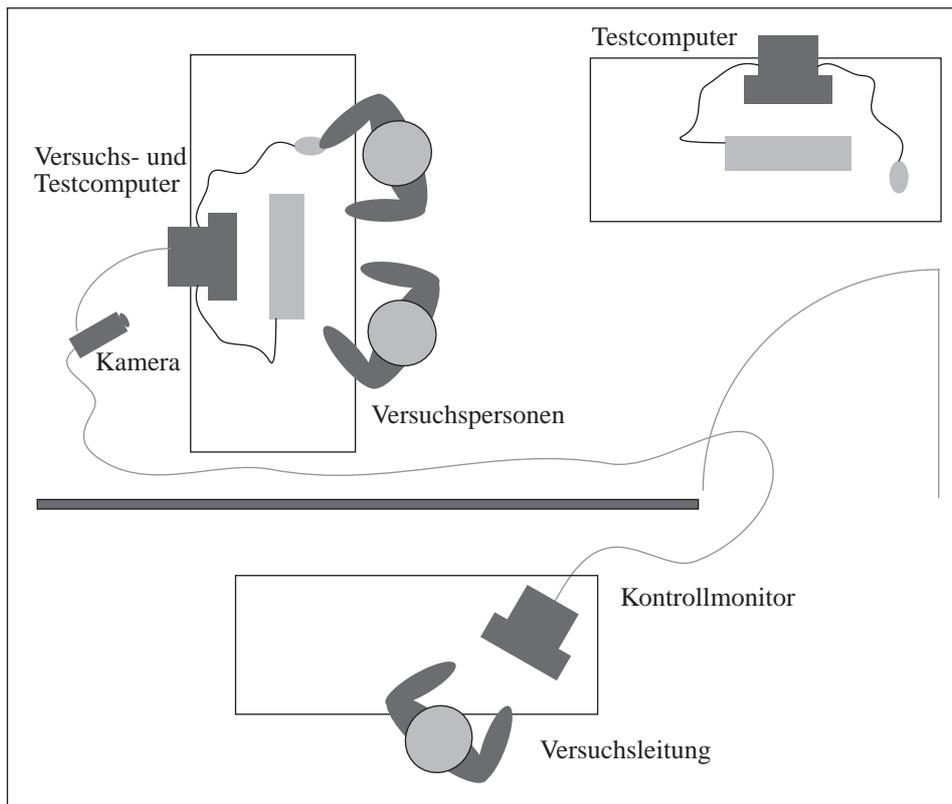


Abbildung 13. Versuchsanordnung.

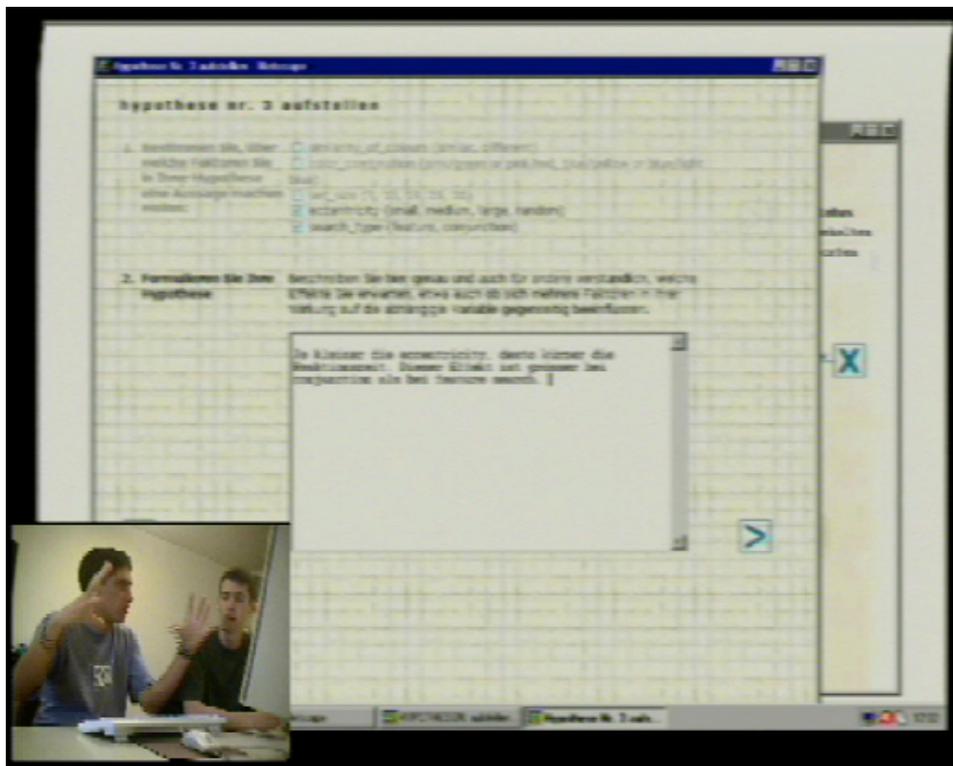


Abbildung 14. Standbild aus einer Videoaufzeichnung.

Hinter dem Experimentalcomputer war eine digitale Videokamera installiert. Ihr Signal wurde zusammen mit dem Bildschirmsignal aufgezeichnet (Abbildung 14). Diese Art der Aufzeichnung ermöglicht später eine genaue Analyse der Interaktion zwischen den Versuchspersonen und dem Lernsystem. Das Videosignal wurde auch auf den Kontrollmonitor für die Versuchsleitung übertragen. Versuchsleiterin oder -leiter saßen während des Experiments in einem Nebenraum und konnten neben dem Videosignal auch die Logfiles von *virtue* in Echtzeit verfolgen.

Die Versuchspersonen trugen sich paarweise mit einer befreundeten oder bekannten Person desselben Geschlechts für einen Termin ein, um am Experiment teilzunehmen. Die angemeldeten Dyaden wurden dann zufällig einer der beiden Versuchsbedingungen, also Einzelpersonen oder Dyaden, zugeteilt. Versuchspersonen in der Einzelbedingung wurden einzeln zum Experiment einbestellt.

Die Versuchsleitung übernahmen drei verschiedene Personen jeweils für einen kompletten Durchgang. Sie führten die Experimente nach einem strengen Ablaufplan durch. Der Ablauf des Experiments ist in Tabelle 3 aufgeführt. Nach

Tabelle 3
Versuchsablauf mit Angaben zur Dauer

Ablauf	Dauer (min.)
Begrüßung und Einführung	5
Fragebogen 1	5
TMS: 2 Untertests	25
Raventest	50
Pause	15
schriftliche Instruktion zur visuellen Suche	10
Vortest	20
mündliche Instruktion zur Kooperation oder zum lauten Denken	5
Wissenschaftliches Entdecken mit <i>virtue</i>	90-120
Nachtest	25
Fragebogen 2	5
Bezahlung und Debriefing	5
Total	260-290

der Begrüßung und einer kurzen Einführung wurden die Versuchspersonen zunächst gebeten zuzustimmen, daß die Experimentalsitzung per Video aufgezeichnet wird. Im ersten Fragebogen wurden dann nach den soziodemographischen Daten Einstellungen zu verschiedenen Aspekten des Lernens erfragt. Die Aufgaben aus den beiden Untertests des Tests für Medizinische Studiengänge wurden in einem Ordner vorgegeben. Die Antworten gaben die Versuchspersonen direkt am Computer ein. Für jeden Untertest hatten die Versuchspersonen 12,5 Minuten Zeit. Den Raventest absolvierten die Versuchspersonen jeweils unter Speedbedingungen innerhalb von 50 Minuten. Alle aufgeführten Tests wurden individuell durchgeführt. Nach diesem ersten Abschnitt der Datenerhebung folgte eine Pause.

Im Anschluß an die Pause erhielten die Versuchspersonen die schriftliche Instruktion zur visuellen Suche in einem Schnellhefter vorgelegt und absolvierten dann individuell den Vortest. Dann wurde für die Dyaden per Los bestimmt, wer von beiden rechts beziehungsweise links vor dem Bildschirm zu sitzen kommt. Für die Dyaden folgte eine Instruktion zur Kooperation. Dort wurden zwei Aspekte betont, einerseits die Abhängigkeit der Bezahlung von der gemeinsamen Leistung im Nachtest und andererseits die Aufforderung, inhaltliche Fragen zu diskutieren, um so den Lernprozeß zu fördern.

Die Einzelpersonen wurden zum lauten Denken instruiert. Dabei wurde herausgestellt, daß sie nur sagen sollten, was ihnen gerade durch den Kopf ging, ohne irgendetwas für die Versuchsleitung zu erklären. Um sich „warmzureden“ erhielten sie in einer Vorübung die folgende Aufgabe: „Zählen Sie auf, wieviele Fenster das Haus hatte, in dem Sie mit Ihren Eltern lebten, als Sie zehn Jahre alt waren.“ (Walach, 1998). Anhand dieser Aufgabe konnte gleichzeitig überprüft werden, ob die Versuchspersonen die Instruktion zum lauten Denken verstanden hatten.

Nach der Anmeldung bei *virtue* erhielten die Versuchspersonen zunächst eine Instruktion zum technischen Umgang mit *virtue*, die sie auch unter der Hilfefunktion später wieder aufrufen können. Dann folgten die Hinweise zu den simulierten Experimenten. Die Versuchspersonen wurden dazu aufgefordert, die in der Instruk-

tion dargestellte Merkmals-Integrations-Theorie und den Einfluß der übrigen Faktoren zu überprüfen, und zwar solange, bis sie aufgrund ihrer simulierten Experimente fundierte Aussagen dazu machen konnten.

Es folgte die Serie von Experimenten in *virtue*. Die Versuchspersonen konnten die Lernumgebung frei explorieren. Außer der oben genannten Anleitung zur Untersuchung der Merkmals-Integrations-Theorie und der weiteren Faktoren waren sie frei in ihrem Vorgehen. Während dieser Zeit wurden die Versuchspersonen und das Bildschirmsignal digital auf Video aufgezeichnet. Außer der computerbasierten Lernumgebung und der schriftlichen Instruktion standen keine weiteren Hilfsmittel zur Verfügung. Dyaden tauschten sich über ihre simulierten Experimente aus, während Einzelpersonen ihr Vorgehen über das laute Denken verbalisierten. Die Versuchsleitung überwachte das Experiment vom Kontrollmonitor im Nebenraum aus und war darüber auch für die Versuchspersonen erreichbar, griff aber nur im Notfall selbst aktiv in die Arbeit der Versuchspersonen im virtuellen Labor ein.

Die Versuchspersonen beendeten ihre Experimentalreihe entweder selbst, wenn sie der Meinung waren, ihre Exploration abgeschlossen zu haben, oder sie wurden nach zwei Stunden durch die Versuchsleitung unterbrochen. Nach Beendigung der Experimentalreihe in *virtue* absolvierten die Versuchspersonen individuell den Nachtest und füllten den Fragebogen zur Evaluation von *virtue* aus. Die Dyaden wurden zusätzlich um eine Bewertung ihrer Interaktion gebeten. Zum Abschluß erfolgte ein Debriefing, in dem die Versuchspersonen über den Zweck der Untersuchung informiert und dazu aufgefordert wurden, ihren Mitstudierenden nichts von dem Experiment zu erzählen. Die Versuchspersonen erhielten nach einer schnellen Auswertung der Nachtests ihre leistungsabhängige Bezahlung.

9

Ergebnisse I: Quantitative Daten aus Tests und Logfiles

In diesem Kapitel werden die Ergebnisse der Untersuchung zum kooperativen wissenschaftlichen Problemlösen mit dem simulierten psychologischen Labor *virtue* berichtet. Zuerst werden die Ergebnisse aus den parallelen Vor- und Nachtests dargestellt. Dann folgt eine prozeßorientierte Analyse der Daten zur Interaktion mit *virtue*, die in den Logfiles festgehalten wurden. Den Abschluß des Kapitels bildet die Darstellung der Ergebnisse aus der Begleitdiagnostik und ihrer Zusammenhänge mit den experimentellen Daten.

9.1 Vor- und Nachtest

Die parallelen Vor- und Nachtests wurden eigens für die vorliegende Untersuchung entwickelt und konnten in Anbetracht des Umfangs der Untersuchung nicht vorher auf ihre Reliabilität hin überprüft werden. An dieser Stelle werden daher einige Daten zur Reliabilität, in diesem Fall zur internen Konsistenz, berichtet, die anhand der Stichprobe aus dem Experiment erhoben wurden.

9.1.1 Reliabilitätsanalyse

Die parallelen Tests sind jeweils in Untertests zu Begriffen aus der Theorie des Gegenstandsbereichs (Untertest Theorie), zu theoretisch vorhergesagten empirischen Zusammenhängen des Gegenstandsbereichs (Untertest Empirie) und zu bereichsübergreifenden Aspekten der Methodologie (Untertest Methodologie) eingeteilt. Da die Untertests jeweils inhaltlich unterschiedliche Konzepte abdecken, wurde ihre Reliabilität getrennt ermittelt. Die Ergebnisse der Reliabilitätsanalyse finden sich in Tabelle 4. Es ist zu beachten, daß die Anzahl der Items zwischen 4 und 7 pro Untertest liegt und daher bei einem einmalig zusammengestellten Test nicht so hoch sein kann, wie bei einem systematisch und schrittweise konstruierten

Untersuchungsinstrument. Zudem kann die Stabilität der Meßwerte bei einer Stichprobengröße von $N = 36$ nicht so hoch sein, wie bei üblicherweise sehr viel größeren Eichstichproben.

Die Reliabilität wurde über Itemanalysen bestimmt. Als Reliabilitätskoeffizient wurde Cronbachs Alpha verwendet. Die korrigierte Trennschärfe und die Item-Schwierigkeit für alle Items sind in Tabelle 4 zusammengefaßt. Für den Untertest Theorie mit 6 Items ergab sich eine Reliabilität von $\alpha = .45$. Der Relia-

Tabelle 4
Reliabilitätsanalyse der Untertests des Vortests

Item	korrigierte Trennschärfe	Schwierigkeit
Untertest Theorie		
1	.38	.75
2	.18	.92
3	.18	.94
4	.13	.75
5	.13	.86
6	.46	.94
Untertest Empirie		
7	-.13	.97
8	.45	.69
9	.41	.64
10	.51	.92
Untertest Methodologie		
11	-.09	.58
12	.26	.69
13	.17	.89
14	.19	.53
15	.09	.61
16a	.21	.31
16b	.19	.28
17a	.36	.56
17b	.30	.61

blitätskoeffizient für den Untertest Empirie mit 4 Items betrug $\alpha = .52$. Für Item 7 stellte sich jedoch eine negative Trennschärfe heraus. Mit einer Lösungswahrscheinlichkeit von 97% zeigte sich für dieses Item auch ein eindeutiger Deckeneffekt. Dieses Item wurde daher eliminiert. Die neu berechnete Reliabilität liegt nun bei $\alpha = .64$. Im Untertest Methodologie wurde Cronbachs Alpha mit $\alpha = .45$ berechnet. Auch hier stellte sich bei einem Item eine negative Trennschärfe heraus. Es handelt sich um das Item 11, welches deshalb aus dem Test ausgeschlossen wurde. Ohne Item 11 ergibt sich für den Untertest Methodologie eine Reliabilität von $\alpha = .48$. Die zu den eliminierten Items aus dem Vortest parallelen Items im Nachtest wurden ebenfalls von den weiteren Berechnungen ausgenommen.

Die jeweilige Reliabilität der drei Untertests ist aus testtheoretischer Sicht als mittelgradig einzustufen. Inhaltlich spiegeln sich in den Alpha-Koeffizienten der Untertests die Homogenität der Stichprobe und die verhältnismäßig geringe Itemschwierigkeit insbesondere im Untertest Theorie wider. Hier starten die Versuchspersonen mit einem hohen Niveau, was indiziert, daß sie die Theorie, wie sie in der Instruktion vorgegeben wurde, gut verstanden haben. Angesichts der kleinen Stichprobe und der eingeschränkten Möglichkeiten, die Tests für das Experiment zu optimieren, kann die Reliabilität als zufriedenstellend betrachtet werden. Sie erlaubt Aussagen über die statistischen Effekte der gemessenen Variablen und darauf aufbauende Schlußfolgerungen. Für weitere Anwendungen wären die Tests allerdings zu optimieren, etwa durch den Austausch spezifischer Items.

Für die zusätzlichen bereichsspezifischen Items im Nachtest zu den Faktoren Exzentrizität, Farbähnlichkeit und Farbkombination läßt sich keine Reliabilität angeben, da sie keine einheitliche Skala bilden. Die Exploration der einzelnen Faktoren ist voneinander unabhängig und die Faktoren werden in unterschiedlichem Detailgrad abgefragt. Die Werte bei den bereichsspezifischen Items dienen also eher einer allgemeinen Orientierung über den Wissenserwerb in diesem Bereich.

9.1.2 Lernerfolg

Der Wissenserwerb beim kooperativen wissenschaftlichen Problemlösen mit *virtue* wurde in parallelen Vor- und Nachtests gemessen. Diese waren in die drei Untertests Theorie, Empirie und Methodologie unterteilt, deren Ergebnisse hier nun im Einzelnen berichtet werden. Darüber hinaus werden die Resultate der zusätzlichen,

bereichsspezifischen Items aus dem Nachtest dargestellt. Über alle Tests hinweg sprechen die Korrelationen der jeweiligen Parallelitems in Vor- und Nachtest in beiden Experimentalgruppen dafür, daß die Rangreihe der Probandinnen und Probanden im Vor- und Nachtest dieselbe ist (siehe Tabelle 5). Das heißt, Personen, die im Vortest überdurchschnittlich waren, sind dies auch im Nachtest; im Vortest schwache Versuchspersonen haben auch im Nachtest im Vergleich zu ihrer Experimentalgruppe Schwierigkeiten.

Tabelle 5
Korrelationen zwischen Parallelitems der Vor- und Nachtests

Item		
Vortest	Nachtest	r
Untertest Theorie		
1	1	.29
2	4	.80
3	6	.15
4	5	.36
5	2	.60
6	3	.70
Untertest Empirie		
8	8	.35
9	9	.32
10	10	.21
Untertest Methodologie		
12	18	.74
13	17	.39
14	19	.39
15	20	.30
16a	21a	.32
16b	21b	.31
17a	22a	.32
17b	22b	.46

Untertest Theorie

Im Untertest Theorie wurde das Verständnis für Begriffe abgefragt, die in der Instruktion zum Gegenstandsbereich vermittelt werden sollten, die bereits vor dem Vortest stattgefunden hatte. Dieser Untertest diente also vor allem dazu zu überprüfen, ob die Versuchspersonen die Instruktion verstanden hatten. Dazu wurde eine zweifaktorielle Varianzanalyse mit Meßwiederholung auf einem Faktor durchgeführt. Als unabhängiger Faktor ging die Experimentalgruppe mit den beiden Stufen Einzelpersonen und Dyaden in die Analyse ein. Bei dem Meßwiederholungsfaktor handelte es sich um die aufsummierten Items aus Vor- und Nachtest.

Die Ergebnisse sind in Abbildung 15 dargestellt. Um eine bessere Vergleichbarkeit zwischen den einzelnen Untertests zu ermöglichen, wurden nicht die absoluten Mittelwerte, sondern die mittleren Prozentwerte abgetragen. Beide Gruppen zeigten bereits im Vortest ein hohes Niveau und erreichten dort im Durchschnitt schon 5.1 von 6 möglichen Punkten. Diese Leistung konnten zum Nachtest hin nur die Dyaden noch leicht steigern. Es ergaben sich jedoch weder für den Unterschied zwischen den beiden Experimentalgruppen ($F(1, 34) = 0.58, p = .45$) noch für die Meßwiederholung bei Vor- und Nachtest ($F(1, 34) = 0.94, p = .34$) und die Interaktion zwischen den beiden Faktoren ($F(1, 34) = 2.33, p = .14$) bedeutsame Effekte. Zusammenfassend fand sich hier im Vortest bei beiden Versuchsgruppen ein vergleichbares, hohes Ausgangsniveau, das sich im Nachtest kaum noch steigerte.

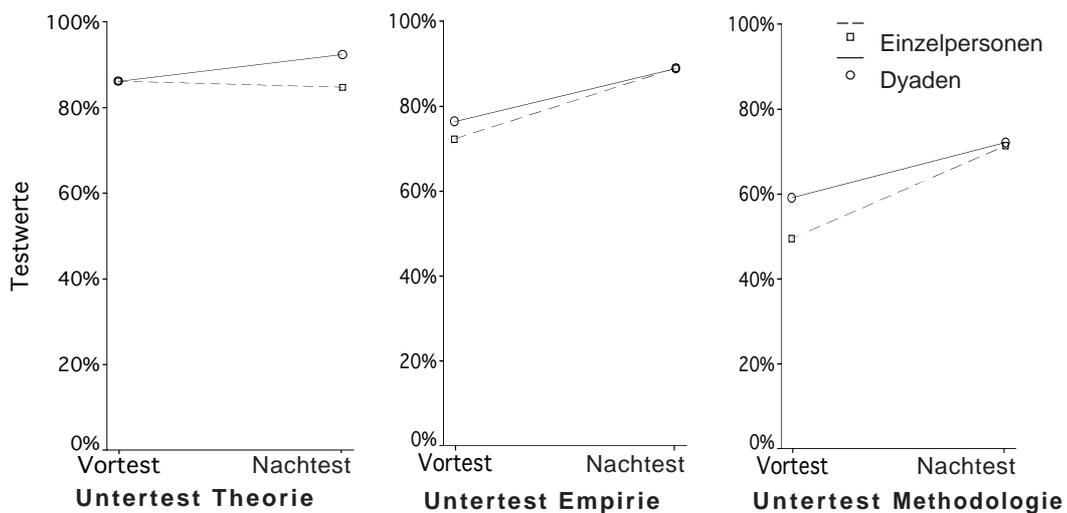


Abbildung 15. Ergebnisse der Untertests Theorie, Empirie und Methodologie dargestellt als relative mittlere Häufigkeiten.

Beide Gruppen sollten also bereits zu Beginn des wissenschaftlichen Problemlösens im simulierten Labor über ein ausreichendes Instruktionsverständnis verfügt haben, um sinnvolle Experimente im Gegenstandsbereich der visuellen Suche planen zu können.

Untertest Empirie

Der Untertest Empirie überprüfte das Verständnis der empirischen Zusammenhänge, wie sie durch die in der Instruktion vermittelte Merkmals-Integrations-Theorie (Treisman & Gelade, 1980) vorhergesagt werden. Durch die Instruktion hatten beide Gruppen bereits ein Vorverständnis über diese Zusammenhänge erworben, welches sich in einer durchschnittlichen Leistung von 75% im Vortest widerspiegelt (Abbildung 15).

Für die Inferenzstatistik wurde eine Varianzanalyse mit Meßwiederholung gerechnet. Als unabhängiger Faktor ging die Experimentalgruppe ein, als Meßwiederholungsfaktor dienten die Daten aus den gemittelten Items des Untertests Empirie in den parallelen Vor- und Nachtests. Zwischen den Gruppen ergab die Varianzanalyse keine bedeutsamen Unterschiede ($F(1, 34) = .05, p = .82$). Beide Gruppen verzeichnen jedoch einen deutlichen Lernzuwachs ($F(1, 34) = 7.82, p = .008$). Bei einer Teststärke von $1 - \beta = .78$ (berechnet für $\alpha = .05$) beträgt die Effektgröße $f = .48$. Nach Cohen (1988) handelt es sich hierbei um einen großen Effekt², das heißt einen durchaus auch für die Praxis relevanten Lerneffekt für den Gegenstandsbereich. Die Interaktion zwischen den beiden Faktoren zeigte dagegen keinen bedeutsamen Effekt ($F(1, 34) = .16, p = .69$). In diesem Untertest konnten demnach keine Unterschiede zwischen Einzelpersonen und Dyaden nachgewiesen werden. Die Ergebnisse lassen keinen Vorteil für Dyaden gegenüber Einzelpersonen beim wissenschaftlichen Problemlösen erkennen. Trotz des hohen Ausgangsniveaus im Vortest trat ein relevanter Lerneffekt zutage.

-
2. Cohen (1988) teilt Effektgrößen für die einzelnen Effekte einer mehrfaktoriellen Varianzanalyse in kleine, mittlere und große Effekte ein. Die Effektgröße berechnet sich aus dem partiellen Etaquadrat mittels folgender Formel:

$$f = \sqrt{\frac{\eta^2}{1 - \eta^2}}. \quad (1)$$

Als Effektgröße für kleine Effekte legt Cohen $f = .10$, für mittlere Effekte $f = .25$ und für große Effekte $f = .48$ fest.

Zusätzliche bereichsspezifische Items im Nachtest

Im Nachtest gab es fünf zusätzliche bereichsspezifische Items, die im Vortest nicht präsentiert wurden, da davon auszugehen war, daß die Versuchspersonen dazu kein Vorwissen hatten, und sich auf diese Weise der Vortest effizienter gestalten ließ. In den fünf Items wurde Wissen zu den drei Faktoren aus der visuellen Suche abgefragt, deren Wirkung nicht mit der Merkmals-Integrations-Theorie (Treisman & Gelade, 1980) erklärt werden kann und die daher nicht Gegenstand der Instruktion waren. Die Items prüften verschiedene Effekte in unterschiedlichem Detailgrad. Deshalb konnten die Items nicht zu einer homogenen Skala zusammengefaßt werden. Dennoch ist der Summenwert informativ, da er ausdrückt, welches Wissen die Versuchspersonen insgesamt zu den genannten drei Faktoren erwarben.

Die Stichprobenkennwerte für die Summe und die einzelnen Items sind in Tabelle 6 dargestellt. Anhand des Summenwertes ist zu erkennen, daß die Versuchspersonen tatsächlich etwas über die abgefragten Faktoren wissen. Es ist davon auszugehen, daß sie dieses Wissen während der Experimente mit *virtue* erworben haben. Die beiden Experimentalgruppen unterscheiden sich statistisch nicht bedeutsam voneinander ($t(34) = -1.151, p = .26$). Betrachtet man die Mittelwerte der einzelnen Items, so wird deutlich, daß sie von unterschiedlicher Schwierigkeit waren.

Tabelle 6

Zusätzliche bereichsspezifische Items im Nachtest

	Einzelpersonen		Dyaden	
	M	SD	M	SD
Item				
11	.92	.29	.83	.38
12	.58	.52	.71	.46
13	.92	.29	.88	.34
14	.58	.52	.58	.50
15	.08	.29	.46	.51
Σ	3.08	.90	3.45	.93

Anmerkung. Bei jedem Item konnte ein Punkt erreicht werden.

Untertest Methodologie

Der Untertest Methodologie fragte die bereichsübergreifenden, methodischen Fertigkeiten ab. Bei diesem Untertest gab es bezüglich des Lernzuwachses, also der Differenz zwischen Nachtest und Vortest, eine Dyade zweier männlicher Teilnehmer mit extremen, negativen Ausreißern. Beide Versuchspersonen erreichten im Vortest bereits hohe Werte für die Methodologie und fielen im Nachtest um 2 Punkte (also um 25%) beziehungsweise sogar um 5 Punkte (63 %) ab. Dies sind die beiden einzigen Fälle, in welchen die Differenz negativ ausfällt. Da es sich um ein und dieselbe Dyade handelt, ist anzunehmen, daß sich hier nach dem langen Experiment ein Motivationsabfall in den Testwerten niederschlägt, in dem sich die beiden Versuchspersonen gegenseitig verstärkt haben. Um eine Verzerrung der Daten zu vermeiden, wurde bei beiden Versuchspersonen die mittlere Differenz zwischen Nachtest und Vortest in der Gruppe der Männer zum Ergebnis des Vortests addiert.

Verglichen mit dem Vortest zur Empirie starteten die Versuchspersonen beim Vortest Methodologie von einem niedrigeren Niveau aus. Sie konnten im Vortest zur Methodologie einen geringeren Anteil an Items richtig beantworten als beim Untertest Empirie. Deshalb fällt der Lernzuwachs bezüglich der Methodologie zum Nachtest hin in beiden Experimentalgruppen etwas größer aus als bei den Untertests zur Empirie (Abbildung 15). Auch hier wurde eine Varianzanalyse mit Meßwiederholung auf dem Faktor Methodologietest berechnet. Als unabhängiger Faktor ging die Experimentalgruppe in die Analyse ein. Für diesen Faktor ergab sich kein bedeutsamer Haupteffekt ($F(1, 34) = 1.56, p = .22$). Obwohl das Ergebnis nicht signifikant ist, ergibt sich nach Cohen (1977) mit $f = .21$ ein kleiner Effekt für ein insgesamt besseres Abschneiden der Dyaden gegenüber den Einzelpersonen, und zwar im Vortest wie im Nachtest. Dies kann jedoch nicht als Hinweis auf einen Vorteil des kooperativen Lernens gewertet werden, da hierfür Dyaden mehr Lernzuwachs verzeichnen müßten als Einzelpersonen, was sich in einem Interaktionseffekt äußern müßte. Für die Meßwiederholung in Vor- und Nachtest zeigte sich ein statistisch bedeutsamer Anstieg ($F(1, 34) = 57.50, p > .001$). Bei einer Teststärke von $1 - \beta = 1.0$ (berechnet für $\alpha = .05$) handelt es sich nach Cohen (1977) mit $f = 1.30$ um einen Effekt mit großer Effektstärke. Die Interaktion zwischen den beiden

Faktoren erwies sich als statistisch unbedeutend ($F(1, 34) = .60, p = .45$). Ähnlich wie im Untertest Empirie konnte im Untertest Methodologie ein auch praktisch bedeutsamer Lernzuwachs beobachtet werden. Beide Experimentalgruppen profitierten gleichermaßen vom wissenschaftlichen Problemlösen im simulierten Labor.

9.2 Logfileanalyse

Im Folgenden werden die Daten aus den Logfiles von *virtue* dargestellt. Dort wurde das Vorgehen der Versuchspersonen im Hypothesen- und Experimenterraum, versehen mit Zeitstempeln, aufgezeichnet. Über die kognitiven Werkzeuge konnten die Versuchspersonen auch Rückschlüsse über die Gültigkeit von Hypothesen angesichts der Ergebnisse aus den simulierten Experimenten eingeben. Im Unterschied zu den Daten aus den Vor- und Nachtests handelt es sich hierbei um Daten je Untersuchungseinheit, also je Einzelperson oder Dyade, für die beiden Versuchspersonen innerhalb einer Dyade werden also keine getrennten Ergebnisse berichtet.

9.2.1 Allgemeines Vorgehen im Hypothesen- und Experimenterraum

Betrachtet man die Hypothesen und Experimentalreihen der beiden Versuchsgruppen, so unterscheiden sie sich auf den ersten Blick kaum (siehe Tabelle 7). Beide Experimentalgruppen arbeiteten gut neunzig Minuten in *virtue*. Es fanden sich keinerlei korrelative Zusammenhänge zwischen den Ergebnissen der Untertests und der Zeit für die Experimente in *virtue*. In beiden Experimentalgruppen wurden zwischen 6 und 7 Hypothesen aufgestellt und je 7 bis 9 Experimente durchgeführt. Die Zahlen zeigen auch, daß fast zu jedem Experiment eine Hypothese aufgestellt wurde, die Studierenden bis auf wenige Ausnahmen also mehr oder weniger hypothesengeleitet voringen.

Tabelle 7

Kennwerte zu den Versuchsreihen in virtue

	Einzelpersonen ($n = 12$)		Dyaden ($n = 12$)	
	$M (SD)$		$M (SD)$	$t(df)$ p
Zeit in <i>virtue</i> (in min.)	99.16 (17.02)		91.83 (18.64)	1.00 (22) .33
Anzahl der Hypothesen	7.17 (4.28)		5.92 (1.83)	0.93 (22) .36
Anzahl der Experimente	9.00 (4.86)		7.25 (2.53)	1.11 (22) .67

Ein wichtiger Schritt bei der Bearbeitung der vorgegebenen Aufgabe, dem Testen der Merkmals-Integrations-Theorie (Treisman & Gelade, 1980), ist die Durchführung eines Experiments mit dem Versuchsplan *Suchtyp X Größe der Anordnung*. Dieses sogenannte kritische Experiment ist das einzige, mit dem die Gültigkeit der Theorie in *virtue* überprüft werden kann. Während dieses Experiment von acht Einzelpersonen und neun Dyaden und damit in beiden Experimentalgruppen fast gleich häufig durchgeführt wurde, unterscheidet sich dessen Position in der Experimentalreihe signifikant. Bei Dyaden steht das kritische Experiment durchschnittlich an Position 2,4 und bei Einzelpersonen an Position 7,1 in der Experimentalreihe ($t(8.1) = 2.4$, $p = .02$; aufgrund signifikant unterschiedlicher Varianzen im Levene-Test wurde mit korrigierten Freiheitsgraden gerechnet. Der p -Wert bezieht sich auf eine einseitige Testung). Hierbei handelt es sich nach Cohen (1977) mit $d = 1.19$ um einen großen Effekt³. Dyaden überprüfen die Theorie bereits im zweiten oder dritten Experiment, während Einzelpersonen sich damit Zeit lassen bis zu ihrem siebten Experiment. Dies könnte ein Hinweis darauf sein, daß Dyaden stärker theoriegeleitet vorgehen. Für Dyaden scheint die Überprüfung der Theorie in jedem Fall einen deutlich höheren Stellenwert zu haben als für Einzelpersonen.

9.2.2 Formulierung der Hypothesen

Die Genauigkeit des Vorgehens beim wissenschaftlichen Problemlösen manifestiert sich auch darin, wie die Hypothesen formuliert wurden. Hier lassen sich zweierlei Kriterien differenzieren: erstens die Spezifität einer Hypothese, also die Unterscheidung in gerichtete und ungerichtete Hypothesen, und zweitens die sprachliche Qualität einer Hypothese, kategorisiert in sauber oder unsauber formulierte. Je ein Beispiel für eine sauber und eine unsauber formulierte Hypothese fin-

-
3. Die Berechnung der Effektstärke erfolgte nach Cohen (1977) über die folgende Formel:

$$d = \frac{M_A - M_B}{s} \quad (2)$$

Für die Populationsmittelwerte wurden nach Bortz und Döring (1995) als Schätzung die Zellenmittelwerte der Stichprobe eingesetzt. Die Streuung wurde nach folgender Formel aus den Streuungen der beiden Stichproben geschätzt:

$$s = \sqrt{\frac{s_A^2 + s_B^2}{2}} \quad (3)$$

det sich in Tabelle 8. Für die Berechnung der Interraterreliabilität stuften zwei unabhängige Raterinnen die gesamten Daten ein. Die Interraterreliabilität für die Kategorisierung der Hypothesen in gerichtete und ungerichtete betrug $K = .97$, die prozentuale Übereinstimmung 98%. Für die Kategorisierung der Hypothesen in sauber und unsauber formulierte wurde eine Interraterreliabilität von $K = .77$ und eine prozentuale Übereinstimmung von 93% berechnet. Beide Kategorisierungen sind also hinreichend reliabel.

Tabelle 8

Beispiele für sauber und unsauber formulierte Hypothesen von Einzelperson E114.

Qualität der Formulierung	Hypothesentext
sauber	je unterschiedlicher die farben sind, desto kürzer sind die reaktionszeiten
unsauber	nach verschiedenen farben mit variierten anzahl von set size wird ein signifikanter unterschied stattfinden wird, ob es feature oder conjunction

Die Häufigkeiten für die jeweiligen Qualitätskategorien zur Einstufung der Hypothesen enthält Tabelle 9. Zwischen den jeweiligen Kategorien besteht eine Abhängigkeit, ohne daß es sich um eine Meßwiederholung im eigentlichen Sinne handelt. Daher wurde in diesem Fall nur für jeweils eine Kategorie ein statistischer Test gerechnet, dessen Ergebnis vor allem eine orientierende Funktion hat. Für die Spezifität der Hypothesen ergibt sich für ungerichtete Hypothesen kein statistisch bedeutsamer Unterschied ($t(22) = .97, p = .34$). Dies ändert sich auch nicht, wenn Prozentwerte betrachtet werden. Für die Variable sprachliche Qualität wurden die

Tabelle 9

Qualität der formulierten Hypothesen in absoluten Häufigkeiten

	Spezifität der Hypothesen		Sprachliche Qualität der Hypothesen	
	gerichtet	ungerichtet	sauber	unsauber
Einzelpersonen	51	32	58	25
Dyaden	47	19	62	4
Total	98	51	120	29

unsauber formulierten Hypothesen statistisch ausgewertet. Es zeigt sich, daß Dyaden bei der Formulierung ihrer Hypothesen sorgfältiger vorgehen als Einzelpersonen ($t(12.79) = 2.56, p = .01$, Freiheitsgrade korrigiert nach signifikantem Levene-Test).

9.2.3 Systematik der Experimentalreihe

Für eine nähere Beschreibung der Systematik beim experimentellen Vorgehen wurden verschiedene Indikatoren identifiziert. Bereits erwähnt wurde, daß fast zu jedem Experiment eine Hypothese formuliert wurde (siehe auch Tabelle 6). Bei 3 Einzelpersonen und 4 Dyaden findet sich darüber hinaus zumindest zeitweilig die Strategie, mehrere Hypothesen nacheinander zu formulieren, bevor sie diese dann am Stück in einer Reihe von Experimenten überprüfen. Diese Strategie kommt immerhin bei einem guten Viertel der Versuchspersonen zum Einsatz. Experimente ganz ohne Hypothese werden hingegen überhaupt nicht durchgeführt.

Ein zentraler Faktor für die Interpretierbarkeit einer Experimentalreihe ist die Konstanthaltung nicht variiertter Faktoren auf derselben Stufe über mehrere Experimente hinweg. Dies wird auch als VOTAT-Strategie für „vary one thing at a time“ bezeichnet (Tschirgi, 1980). Für die Auswertung wurden diejenigen Übergänge zwischen den Experimenten gezählt, in denen die nicht variierten Faktoren nicht konstant gehalten wurden. Dyaden gingen in diesem Fall konsequenter vor als Einzelpersonen. Während Dyaden im Durchschnitt in 0,9 Fällen pro Experimentalreihe ihr Design nicht konstant hielten, fanden sich bei Einzelpersonen durchschnittlich 2,4 solcher Fälle ($t(22) = 3.0, p = .004$, einseitig getestet). Nach Cohen (1977) handelt es sich hierbei mit $d = 1.20$ um einen großen Effekt. Dyaden gehen also bei der Planung ihrer Versuchsdesigns sorgfältiger vor als Einzelpersonen.

Als ein Sonderfall der Konstanthaltung von Variablen wurde beobachtet, daß einige Versuchspersonen den Faktor Suchtyp systematisch auf der Stufe Verbindungssuche konstanthielten. Dies könnte als ein Indiz dafür interpretiert werden, daß während des Experimentierens die Bedeutung des Faktors Suchtyp erkannt wurde, zumindest daß auf dieser Faktorstufe andere Variablen stärker variieren als auf der Stufe Merkmalssuche. Jeder Datensatz wurde danach eingestuft, ob der Faktor nach einem kritischen Experiment mit Variation des Faktors Suchtyp über-

haupt konstantgehalten wurde, ob er über alle Experimente hinweg systematisch variiert wurde oder ob er auf der Stufe Merkmals- bzw. Verbindungssuche konstantgehalten wurde. Hier spiegelt sich in leichter Tendenz der generelle Befund zur Konstanthaltung wider, nämlich daß Dyaden sorgfältiger bei der Versuchsplannung vorgehen und sich zusätzlich etwas stärker am Faktor Suchtyp orientieren (siehe Tabelle 10). Ein Chi-Quadratstest ergab allerdings keine Signifikanzen ($\chi^2(3, N = 24) = 4.14, p = .25$).

Tabelle 10
Konstanthaltung des Faktors Suchtyp

	Keine Konstanthaltung	Konstanthaltung auf Merkmals-suche	Suchtyp systematisch experimentell variiert	Konstanthaltung auf Verbindungs-suche
Einzelpersonen ($n = 12$)	3	3	1	5
Dyaden ($n = 12$)	0	4	3	5

Komplexer werdende Versuchspläne stellen ein weiteres Merkmal für die Systematik einer Versuchsreihe dar. In *virtue* sind zwar maximal zweifaktorielle Versuchspläne möglich. Wird ein zweifaktorielles Experiment jedoch mit derselben Faktorkombination wiederholt, indem nur einer der nicht variierten Faktoren auf einer anderen Stufe konstant gehalten wird, so kann man von einem pseudo-dreifaktoriellen Experiment sprechen. Eine Ausnutzung immer komplexerer Versuchspläne läßt sich bei vier Einzelpersonen und zwei Dyaden beobachten. Eine Hinwendung zu einfacheren Versuchsplänen vollziehen eine Einzelperson und zwei Dyaden. Diese Versuchspersonen waren eventuell damit überfordert, die Ergebnisse aus komplexeren Experimenten zu interpretieren. Ein Chi-Quadrat-Test konnte keine bedeutsamen Unterschiede zwischen den Gruppen aufzeigen ($\chi^2(2, N = 24) = .26, p = .58$).

Weitere Muster einer systematischen Entwicklung der Versuchsreihe sind das einfaktorielle Testen zweier Variablen, bevor sie im direkt darauffolgenden zweifaktoriellen Experiment kombiniert werden, sowie das blockweise Testen eines Faktors in mindestens drei aufeinanderfolgenden Experimenten. Die Mittelwerte und die Ergebnisse der t-Tests finden sich in Tabelle 11. Bei den Zyklen

zunächst ein-, dann zweifaktorieller Variablen- testung handelt es sich um ein eher selten auftretendes Muster, das sich bei zwei Einzelpersonen und vier Dyaden findet. Blöcke von Versuchen zu einem Faktor führen acht Dyaden und acht Einzelpersonen durch, jeweils fünf von ihnen sogar zwei oder mehr Experimentalblöcke.

Tabelle 11

Kennwerte für die systematische Entwicklung einer Versuchsreihe in absoluten Häufigkeiten

	Einzelpersonen ($n = 12$)		Dyaden ($n = 12$)	
	M (SD)		M (SD)	
Zyklen ein-, dann zweifaktorieller Variablen- testung	0.25 (0.62)		0.58 (1.00)	$t(22) = -0.98$, $p = .36$
Experimentalblöcke	1.50 (1.57)		1.50 (1.51)	$t(22) = 0.00$, $p = 1.00$

9.2.4 Abdeckung des Hypothesen- und Experimenterraums

Über das Verständnis der Versuchspersonen vom Gegenstandsbereich und ihr methodisches Vorgehen bei der Suche im Hypothesen- und Experimenterraum geben die aufgestellten Hypothesen und die zu den einzelnen Faktoren durchgeführten Experimente Auskunft. Fast alle Dyaden und Einzelpersonen untersuchten jeden Faktor mindestens in einer Hypothese und in einem Experiment. Die Ausnahmen bildeten zwei Einzelpersonen und eine Dyade, die einen Faktor unberücksichtigt ließen, sowie eine Einzelperson, die sogar zwei Faktoren ungeprüft ließ. Hierbei handelte es sich jeweils um Faktoren, deren Wirkung nicht in der Theorie beschrieben worden war (Farbkombination, Farbähnlichkeit und Exzentrizität). Von 10 möglichen Interaktionen untersuchten die Versuchspersonen in beiden Gruppen etwa 3,5 in ihren Hypothesen und 4 in ihren Experimenten und unterscheiden sich darin jeweils nicht von einander ($t(22) = .472$, $p = .64$ beziehungsweise $t(22) = .215$, $p = .83$). Einen detaillierteren Überblick über die Befunde zu den einzelnen Faktoren ermöglichen die Abbildungen 16 und 17.

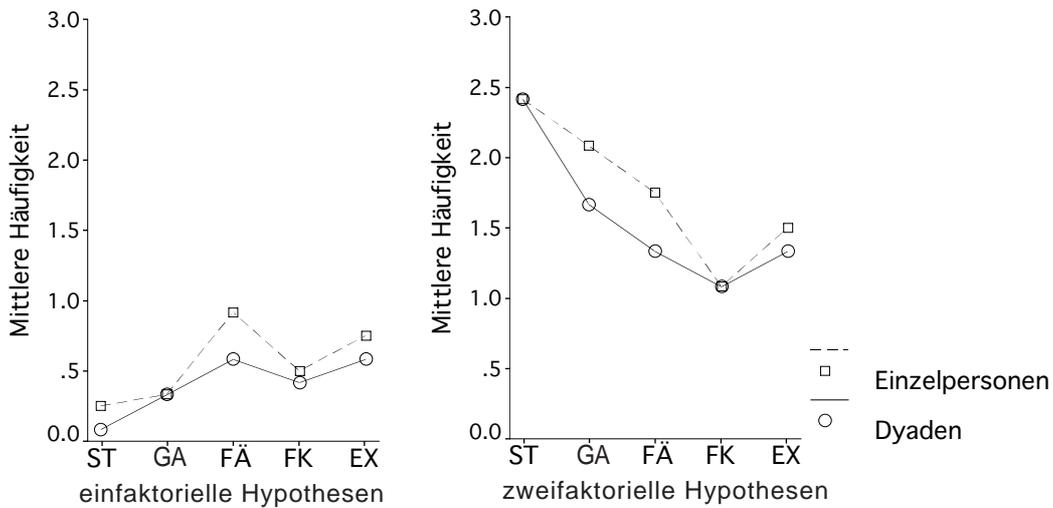


Abbildung 16. Von den Versuchspersonen ($N = 24$, 12 Dyaden und 12 Einzelpersonen) in *virtue* formulierte ein- und zweifaktorielle Hypothesen nach Faktoren. Zweifaktorielle Hypothesen wurden für jeden Faktor getrennt, also insgesamt zweimal gewertet. ST = Suchtyp; GA = Größe der Anordnung; FÄ = Farbähnlichkeit; FK = Farbkombination; EX = Exzentrizität.

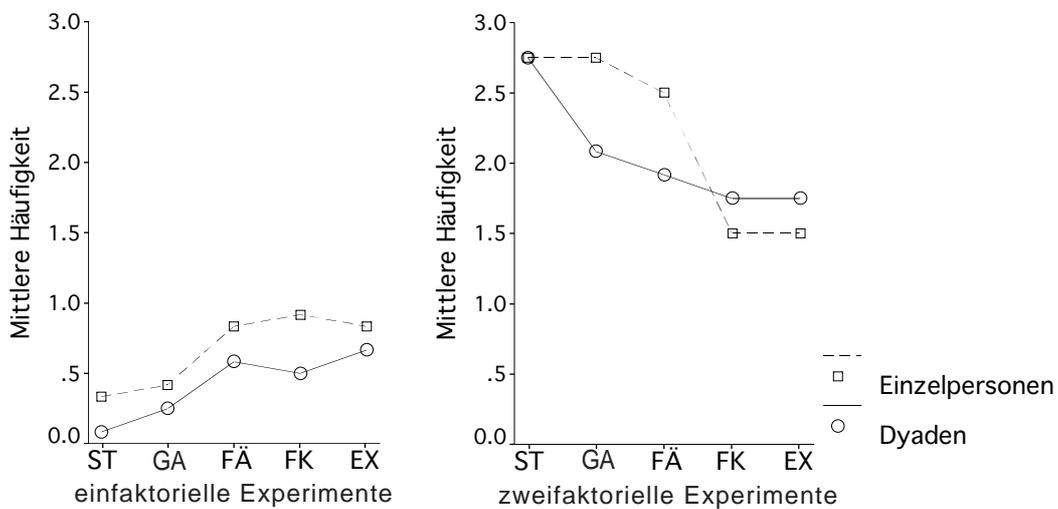


Abbildung 17. Von den Versuchspersonen ($N = 24$, 12 Dyaden und 12 Einzelpersonen) in *virtue* durchgeführte ein- und zweifaktorielle Experimente nach Faktoren. Zweifaktorielle Experimente wurden für jeden Faktor getrennt, also insgesamt zweimal gewertet. ST = Suchtyp; GA = Größe der Anordnung; FÄ = Farbähnlichkeit; FK = Farbkombination; EX = Exzentrizität.

Die deskriptiven Daten zeigen deutlich unterschiedliche Tendenzen für die fünf Faktoren des Gegenstandsbereiches. Da es sich um abhängige Daten aus einer Meßwiederholung handelt, die Voraussetzungen für eine Varianzanalyse aber nicht

erfüllt sind und wegen der ungleichen Verteilungen auch keine geeigneten non-parametrischen Verfahren zur Verfügung stehen (Vorberg & Blanken, 1999), mußte auf eine inferenzstatistische Auswertung verzichtet werden.

Für die beiden Faktoren Suchtyp und Größe der Anordnung stellten Einzelpersonen wie Dyaden kaum einfaktorielle Hypothesen auf und führten auch fast keine einfaktoriellen Experimente dazu durch. Zu den Faktoren Farbähnlichkeit, Farbkombination und Exzentrizität führten sie etwas mehr einfaktorielle Experimente durch und formulierten abgesehen von einem leichten Einbruch beim Faktor Farbkombination auch entsprechend viele einfaktorielle Hypothesen.

Für zweifaktorielle Hypothesen und Experimente ergibt sich im Vergleich zu den einfaktoriellen ein komplementäres Muster. Beide Experimentalgruppen untersuchten den Faktor Suchtyp am ausführlichsten, gefolgt von den Faktoren Größe der Anordnung, Farbähnlichkeit, Exzentrizität und Farbkombination. Ähnlich wie bei den einfaktoriellen Experimenten wurden auch hier für den Faktor Farbkombination im Vergleich zu den anderen Faktoren etwas weniger Hypothesen formuliert.

Betrachtet man die Korrektheit der Vorhersagen, so fällt wieder ein Unterschied zwischen ein- und zweifaktoriellen Hypothesen sowie zwischen den einzelnen Faktoren auf (Abbildung 18). Auch hier erübrigt sich aus den oben genannten Gründen eine inferenzstatistische Auswertung. Da sich die Muster in beiden Gruppen sehr ähneln, wurden die Gruppen der besseren Übersichtlichkeit wegen in einer Abbildung zusammengefaßt. Insgesamt werden in beiden Gruppen fast genauso viele falsche wie korrekte Vorhersagen formuliert (Anteil korrekter Vorhersagen für Dyaden $M = 54\%$, für Einzelpersonen $M = 50\%$). Bei den einfaktoriellen Hypothesen fällt auf, daß vor allem für den Faktor Farbähnlichkeit korrekte Vorhersagen überwiegen, sich dieses Verhältnis bei den zweifaktoriellen Vorhersagen jedoch umkehrt. Dies gilt in etwas abgeschwächter Form auch für den Faktor Exzentrizität. Sowohl für den Faktor Suchtyp als auch für den Faktor Größe der Anordnung fallen korrekte zweifaktorielle Vorhersagen leichter als für die übrigen Faktoren. Dies könnte auf das Vorwissen über diese beiden Faktoren aus der Instruktion zurückzuführen sein. Daß zu diesen beiden Faktoren auch falsche Vorhersagen getroffen werden, ist einerseits auf das noch nicht ganz vollständige Ver-

ständnis der Theorie zurückzuführen und andererseits auf Vorhersagen in Kombination mit den drei nicht in der Instruktion beschriebenen Faktoren. Der Faktor Farbkombination wird fast nur im Zusammenhang mit anderen Faktoren für Vorhersagen herangezogen. Insgesamt werden zu diesem Faktor die wenigsten Hypothesen überhaupt formuliert. Diese sind häufig korrekt, hier wird die Unwirksamkeit dieses Faktors vorhergesagt.

Der Vergleich der beiden Experimentalgruppen bezüglich der Häufigkeiten von Hypothesen und Experimenten spiegelt die bereits aus akkumulierten Daten bekannten Verhältnisse wider (siehe Tabelle 7): Es finden sich bei Einzelpersonen in fast allen Fällen etwas mehr Hypothesen und Experimente als bei Dyaden. Eine bemerkenswerte Ausnahme bilden die zweifaktoriellen Hypothesen und Experimente zum Faktor Suchtyp. Hier liegen Dyaden und Einzelpersonen gleich auf, das heißt, im Verhältnis zur Gesamtzahl der Experimente ist der Anteil der Experimente mit dem Faktor Suchtyp bei den Dyaden (M = 65%) höher als bei den Einzelpersonen (M = 50%). Dies ist ein weiterer Hinweis darauf, daß die Bedeutung des Faktors Suchtyp von Dyaden eher erkannt wurde als von Einzelpersonen.

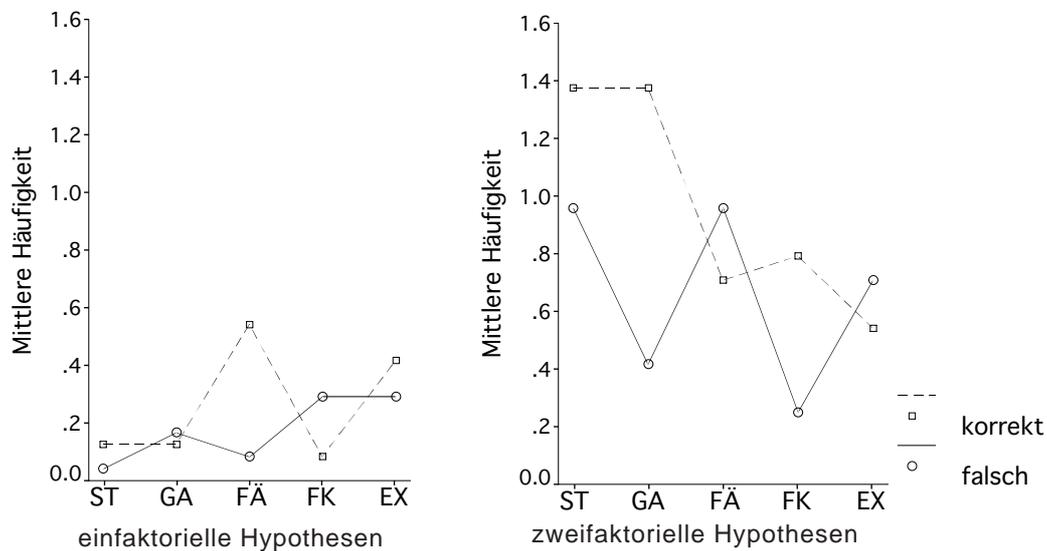


Abbildung 18. Korrektheit in *virtue* formulierter ein- und zweifaktorieller Vorhersagen nach Faktoren, zusammengefaßt für beide Experimentalgruppen ($N = 24$). Zweifaktorielle Hypothesen wurden für jeden Faktor getrennt, also insgesamt zweimal gewertet. ST = Suchtyp; GA = Größe der Anordnung; FÄ = Farbähnlichkeit; FK = Farbkombination; EX = Exzentrizität.

Das komplementäre Muster bei ein- und zweifaktoriellen Untersuchungen könnte darauf beruhen, daß die Wirkung der Faktoren Suchtyp und Größe der Anordnung in der Instruktion bereits theoretisch beschrieben wird, die Wirkung der restlichen Faktoren jedoch nicht. Die Versuchspersonen könnten daran interessiert gewesen sein, die Wirkung dieser Faktoren separat zu überprüfen, um leichter zu interpretierende Ergebnisse zu erhalten. Dafür spricht auch, daß solche einfaktoriellen Untersuchungen nicht nur vor zweifaktoriellen mit denselben Faktoren durchgeführt wurden, sondern auch ohne weitere zweifaktorielle Untersuchung. Daß der Faktor Suchtyp am häufigsten in zweifaktoriellen Untersuchungen variiert wurde, spricht dafür, daß zumindest einige Versuchspersonen dessen zentrale Rolle für den Gegenstandsbereich erkannt hatten.

Der Faktor Farbkombination nimmt eine Sonderrolle ein, es finden sich weniger einfaktorielle und zweifaktorielle Hypothesen als für die übrigen Faktoren. Diese Besonderheit könnte einerseits darin begründet liegen, daß es im Gegensatz zu den anderen nicht theoretisch in der Instruktion erklärten Faktoren Farbähnlichkeit und Exzentrizität nicht so leicht fällt, eine intuitive Hypothese hierzu zu formulieren. Zum anderen ist dies derjenige Faktor, der tatsächlich keinen Einfluß auf die Reaktionszeit bei der visuellen Suche hat.

9.2.5 Ergebnisinterpretation und Bewertung von Hypothesen

Die Interpretation empirischer Ergebnisse und daraufhin die Bewertung von Hypothesen ist ein wesentlicher Schritt bei der Beurteilung der Gültigkeit einer Theorie. Dieser Schritt fällt Anfängerinnen und Anfängern besonders schwer. Bei der Bewertung von Hypothesen können vier Fälle unterschieden werden (Klahr & Dunbar, 1988): Korrekte Entscheidungen werden gefällt, (1) wenn eine Hypothese durch die empirischen Daten gestützt wird und die Hypothese beibehalten wird; und (2) wenn die empirische Evidenz der Hypothese widerspricht und die Hypothese verworfen wird. Um falsche Entscheidungen handelt es sich, (3) wenn die empirischen Ergebnisse der Hypothese entgegengesetzt sind, die Hypothese aber dennoch beibehalten wird (Bestätigungsfehler); und (4) wenn die Ergebnisse der Hypothese entsprechen, dies aber nicht erkannt wird und die Hypothese trotzdem verworfen wird. Die Häufigkeiten der vier Fälle sind in Tabelle 12 abgetragen.

Tabelle 12

Bewertung von Hypothesen: absolute Häufigkeiten

	Beibehaltung, bestätigende Daten	Verwerfung, wider- sprechende Daten	Beibehaltung, wider- sprechende Daten	Verwerfung, bestätigende Daten	Keine Bewertung der Hypothese
Einzelpersonen	41	24	13	4	2
Dyaden	38	24	3	1	2
Total	79	48	16	5	4

Insgesamt bewerten die Versuchspersonen die meisten Hypothesen korrekt. Für die beiden Fälle korrekter Bewertung (1) und (2) ergeben sich augenscheinlich keine großen Gruppenunterschiede. Dyaden haben aber eine wesentlich geringere Neigung zum *confirmation bias* (3) als Einzelpersonen. Hierfür wurde ein U-Test nach Mann-Whitney berechnet, da die Daten die Voraussetzungen für einen t-Test nicht erfüllen. Es ergab sich bei $U = 35.50$, $p = .02$ ein bedeutsamer Unterschied zwischen den beiden Gruppen. Daß Dyaden zu weniger Fehlentscheidungen neigen als Einzelpersonen zeigt sich auch für den Fall, daß Hypothesen trotz bestätigender Daten verworfen werden (4). Hier läßt sich angesichts der geringen Fallzahl jedoch keine aussagekräftige Statistik mehr berechnen. Dieselben statistischen Effekte ergeben sich auch, wenn statt mit absoluten mit relativen Häufigkeiten gerechnet wird. Dyaden gehen also bei der Bewertung von Hypothesen sorgfältiger vor, sie treffen weniger Fehlentscheidungen.

9.3 Soziodemographische Daten, Fragebögen und Intelligenzdiagnostik

Um die Vergleichbarkeit der beiden Stichproben zu kontrollieren, wurden im Rahmen des Experiments verschiedene Daten in Form von Tests und Fragebögen erhoben. Bei den Tests handelt es sich im einzelnen um den Raventest (APM) und die beiden Untertests „Medizinisch-naturwissenschaftliches Grundverständnis“ und „Diagramme und Tabellen“ aus dem Test für medizinische Studiengänge. Die Fragebögen bestanden aus den Skalen „Need for Cognition“, „Standfestigkeit gegenüber Gruppen“, Items zur Plan- und Zielorientierung sowie Fragen zu Einstellung

und Erfahrung beim kooperativen Lernen und Arbeiten am Computer. Eine weitere Einstufung erlauben die Noten aus den kurz vor dem Experiment absolvierten Prüfungen zum zweiten Propädeutikum.

9.3.1 Vergleich der Experimentalgruppen

Für die Tests und Fragebögen sowie für die Prüfungsnote in Statistik und Methodenlehre ergab sich kein Unterschied zwischen den beiden Experimentalgruppen (Tabelle 13) mit einer Ausnahme, der Zielorientierung. Da die Zielorientierung jedoch, wie sich gleich zeigen wird, nicht mit den Ergebnissen der Wissenstests korreliert, ist dieser Befund zu vernachlässigen. Die Gruppen sind also im Hinblick auf Vorwissen, allgemeine Intelligenz und Handlungsstile als äquivalent zu betrachten.

Tabelle 13

Diagnostische Kennwerte der Experimentalgruppen im Vergleich

	Einzelpersonen (<i>n</i> = 12)		Dyaden (<i>n</i> = 24)		<i>t</i> (34)	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Alter	24.5	5.41	26.4	5.6	-0.92	.36
TMS: med.-nat.wiss. Grundverst. (max 5 P.)	2.58	1.44	3.00	1.18	-0.90	.38
TMS: Diagramme und Tabellen (max 5 P.)	2.33	.98	2.15	.90	0.55	.58
Raventest Rohwerte APM I + II	36.1	5.47	37.8	5.15	-0.94	.35
Need for Cognition	4.42	.50	4.68	.58	-1.35	.18
Standfestigkeit gegenüber Gruppen (1-5)	2.11	1.14	2.78	1.25	-1.55	.13
Einstellung zum kooperativen Lernen (1-5)	5.23	1.03	4.70	1.32	1.22	.23
Zielorientierung	3.56	.56	3.01	.66	2.43	.02
Planorientierung	3.23	.51	3.21	.50	0.09	.93
Note Statistik/Methodenlehre	4.21	1.03	4.38	1.27	-0.39	.70
Computerliteracy (0-1)	.74	.17	.76	.15	0.44	.66
Bewertung von <i>virtue</i>	4.64	1.20	4.46	1.13	-0.50	.62

Anmerkung. Den Variablen, bei denen nichts anderes angegeben ist, lagen mit Ausnahme des Raventests und der Statistiknote Items mit einer Ratingskala von 1 bis 7 zugrunde. In der Schweiz wird eine Notenskala mit 6 als bester und 1 als schlechtester Note verwendet. Eine Prüfung ist bestanden, wenn mindestens die Note 4 erreicht wurde.

9.3.2 Zusammenhänge von Diagnostik und Lernerfolg

Um zu bestimmen, inwiefern die Kennwerte aus den genannten Tests mit dem Wissenserwerb zusammenhängen, wurden die Korrelationen bestimmt zwischen den Testwerten und den einzelnen Untertests des Vortests sowie dem eigentlichen Lernzuwachs, also der Differenz zwischen Nachtest- und Vortestwerten für die jeweiligen Untertests. Nur in vier Fällen fanden sich bedeutsame Korrelationen: Die Skala aus dem TMS zum wissenschaftlichen Textverstehen korreliert mit $r = .36$, $p = .003$ mit dem Untertest Theorie des Vortests und mit $r = .41$, $p = .01$ mit dem Untertest Empirie des Vortests; die Skala aus dem TMS zu Diagrammen und Tabellen korreliert mit $r = .43$, $p = .009$ mit dem Untertest Theorie des Vortests; und der Raventest korreliert mit $r = .33$, $p = .05$ mit dem Untertest Theorie des Vortests. Die Bedeutung dieser signifikanten Ergebnisse für die Skalen aus dem TMS und den Raventest ist an der Vielzahl der durchgeführten statistischen Tests zu relativieren, und diese Resultate beziehen sich auch nur auf den Vortest und nicht auf den Lernzuwachs. Während also geringe Einflüsse des Vorverständnisses auf die Ausgangsleistung festgestellt werden konnten, zeigte sich keinerlei Zusammenhang zwischen den Daten der Testdiagnostik und dem gemessenen Lernzuwachs.

9.3.3 Frauen und Männer

Für eine Analyse des Faktors Geschlecht wurden beide Experimentalgruppen zusammengefaßt, da die Teilstichprobe der Männer sonst für eine Analyse zu klein ausgefallen wäre und sich in den Vor- und Nachtests keine gravierenden Unterschiede zwischen den Experimentalgruppen ergeben hatten. Zweifaktorielle Varianzanalysen mit Meßwiederholung auf dem Faktor Vor- und Nachtest und dem unabhängigen Faktor Geschlecht ergaben für die Untertests Theorie und Empirie abgesehen von dem bekannten Lerneffekt beim Untertest Empirie keine bedeutsamen Haupt- und Interaktionseffekte.

Die Daten des Untertests Methodologie erfüllen die Voraussetzungen für eine Varianzanalyse mit dem Faktor Geschlecht aufgrund zu unterschiedlicher Varianzen nicht. Für diesen Untertest findet sich in der Gruppe der Männer im Vortest ($M = 4.87$, $SD = 2.48$) und im Nachtest ($M = 5.99$, $SD = 2.12$) trotz der Korrektur der beiden negativen Ausreißer eine deutlich größere Streuung als bei den

Frauen mit $M = 4.34$, $SD = 1.25$ im Vortest und $M = 6.01$ und $SD = 1.11$. Aus diesem Grund wurde mit den einzelnen Vor- und Nachtestwerten sowie den Differenzwerten von Vor- und Nachtest jeweils ein Mann-Whitney U-Test gerechnet. Die beiden Gruppen unterschieden sich weder im Vortest ($U = 109.5$, $p = .67$) oder Nachtest ($U = 19.0$, $p = .67$) noch in den Differenzwerten ($U = 90.5$, $p = .26$).

Auch für die in den Fragebögen erhobenen Variablen ergaben die jeweiligen t-Tests keine bedeutsamen Unterschiede. Zusammenfassend zeigte sich bei den Männern im Untertest Methodologie eine bedeutend größere Streuung der Leistung als bei den Frauen, ansonsten ließen sich jedoch keine Unterschiede feststellen.

9.3.4 Nachbefragung der Versuchspersonen

Die Computererfahrung der Versuchspersonen wurde auf einer sehr globalen Ebene erhoben. Es wurde allgemein nach der Bedienung von Computern gefragt und die Zeit abgefragt, die sie in der Regel am Computer verbringen. Bei diesen Items schätzten die Versuchspersonen ihre eigene Computererfahrung relativ hoch ein (siehe die Mittelwerte in Tabelle 13).

Für die Einschätzung des Experiments wurden zwei Summenwerte gebildet. Zum einen wurden die Items zur Beurteilung der Lernumgebung *virtue*, zum anderen die Items der Dyaden zur Beurteilung ihrer Kooperation zusammengefaßt. Die Versuchspersonen bewerteten *virtue* etwas oberhalb des Skalenmittelwertes, also leicht positiv (Tabelle 13). Es fanden sich keine weiteren statistischen Zusammenhänge für diesen Parameter. In den freien Kommentaren erwähnten sie am häufigsten, daß das Experiment interessant gewesen sei und daß sie es als lang und anstrengend empfanden. Dies mag ein Grund für die neutrale Bewertung von *virtue* sein. Schwierigkeiten in der Bedienung von *virtue* wurden weder von den Versuchsleiterinnen und -leitern beobachtet, noch von den Versuchspersonen in der Nachbefragung erwähnt.

Beide Gruppen berichteten eine insgesamt positive Einstellung zum kooperativen Lernen, die Einzelpersonen sogar noch mehr als die Dyaden, auch wenn dieser Unterschied statistisch nicht signifikant ist. In der Nachbefragung wurden die Dyaden um eine Einschätzung ihrer Kooperation während des wissenschaftlichen Problemlösens mit *virtue* gebeten. Diese fiel bei Items, die auf einer Ratingskala

von 1 bis 7 einzustufen waren, ausgesprochen positiv aus ($M = 5.32$, $SD = .89$). Zudem wurde zwischen der Einstellung zum kooperativen Lernen und der Bewertung der Kooperation im Experiment eine bedeutsame Korrelation festgestellt ($r = .57$, $p < .01$). Da sich keine weiteren Zusammenhänge mit den beiden Korrelationsparametern ergaben, können keine Schlußfolgerungen über die Ursache dieses Zusammenhangs gezogen werden. Die Einstellung zum kooperativen Lernen könnte sich auf die Wahrnehmung der experimentellen Kooperation ausgewirkt oder auch direkt die soziale Interaktion beeinflußt haben. Auch frühere Erfahrungen mit dem kooperativen Lernen könnten sich auf die Einstellung dazu und die Interaktion im Experiment auswirken.

10

Ergebnisse II: Qualitative Analyse ausgewählter Einzelfälle

Die anhand der oben berichteten quantitativen Daten gewonnenen Aussagen sollen in diesem Kapitel anhand einer qualitativen Analyse von Protokollen zweier Dyaden und zweier Einzelpersonen vertieft werden. Die Auswahl der Einzelfälle erfolgte aufgrund verschiedener Erwägungen: Die Fälle sollen demonstrieren, wie wissenschaftliches Lernen mit *virtue* möglich ist und welche Schwierigkeiten dabei auftreten können. Beide Settings, also Lernen in Dyaden und als Einzelperson sollten ebenso berücksichtigt werden, wie verschiedene Konstellationen, was das individuelle Vorwissen anbelangt.

Die qualitative Analyse orientiert sich an den Arbeiten von Roschelle (1992) und Teasley (1995). Während die quantitativen Daten eine durchschnittsbezogene Analyse erlauben, soll die qualitative Analyse Aussagen über den Ablauf des wissenschaftlichen Lernens und Problemlösens mit *virtue* ermöglichen. Hier wird eine prozeßorientierte, beschreibende Datenanalyse angestrebt. Sie umfaßt die Beschreibung der experimentellen Zyklen von der Hypothesenformulierung über die Versuchsplanung bis zur Auswertung der Daten und fragt nach der Entwicklung von Zyklus zu Zyklus. Zudem wird die inhaltliche Argumentation der Versuchspersonen betrachtet, etwa woran sie ihr Vorgehen orientieren und inwiefern sie Vorwissen bei ihren Entscheidungen einbeziehen. Bei Dyaden ist auch die interaktionelle Komponente zu berücksichtigen, beispielsweise welche Rollen die Versuchspersonen im Dialog einnehmen, wer dabei welche Inhalte einbringt oder wer das experimentelle Vorgehen bestimmt.

Die Auswahl von je zwei Einzelpersonen und Dyaden erfolgte anhand qualitativer Kriterien so, daß eine möglichst große Bandbreite an unterschiedlichen Verhaltensweisen beim Lernen mit *virtue* exemplarisch dargestellt werden kann. Die

Tabelle 14

Ergebnisse der für die qualitative Analyse ausgewählten Versuchspersonen

	Theorie (max. 6)		Empirie (max. 3)		Methodologie (max. 8)	
	Vortest	Nachtest	Vortest	Nachtest	Vortest	Nachtest
Einzelpersonen						
<i>M (SD)</i>	5.2 (1.4)	5.1 (1.1)	2.2 (.7)	2.7 (.7)	4.0 (1.4)	5.7 (1.6)
Ella	6	6	2	3	4.75	7.75
Emelie	6	4	3	3	3.5	6.5
Dyaden						
<i>M (SD)</i>	5.2 (.8)	5.5 (.7)	2.3 (1.1)	2.7 (.8)	4.7 (1.7)	5.8 (1.1)
Dana (1)	6	6	3	3	6.75	7.75
Delia (1)	4	5	0	3	3.75	5.75
Dido (2)	5	6	3	3	3.75	5.75
Dora (2)	6	6	2	1	4.5	7.5

Anmerkung. Als erste Zeile für jede Gruppe sind die Gruppenmittelwerte und in Klammern die Standardabweichungen aufgeführt. Die eingeklammerten Zahlen hinter den Namen der Versuchspersonen aus den Dyaden kennzeichnen die Zugehörigkeit zu Dyade 1 bzw. Dyade 2.

Ergebnisse der ausgewählten Versuchspersonen für die einzelnen Untertests finden sich in Tabelle 14. Einzelperson 1⁴, hier Ella genannt, wurde wegen ihrer guten Leistung im Experiment ausgewählt. Sie zeigt sowohl im Untertest Empirie als auch im Untertest Methodologie einen Leistungszuwachs vom Vor- zum Nachtest und erreicht im Nachtest bis auf einen Viertelpunkt die volle Punktzahl. Einzelperson 2, Emelie, kann zwar im Untertest Methodologie zum Nachtest hin zulegen und schneidet dort sogar überdurchschnittlich ab, reicht jedoch nicht an das Niveau von Ella heran und verliert im Nachtest auch noch zwei Punkte beim Untertest Theorie.

Bei der Auswahl der Dyaden wurde auf die Konstellation innerhalb der Paare, insbesondere das im Vortest gemessene Vorwissen geachtet, weil damit unterschiedliche Möglichkeiten für Erklärungen und Nachfragen gegeben sind. In Dyade 1 legt Dana bereits einen sehr guten Vortest ab und kann sich zum Nachtest nur noch in Methodologie steigern, so daß sie dann fast volle Punktzahl erreicht.

4. Der besseren Lesbarkeit wegen bekommen alle Versuchspersonen hier Namen, die der Dyaden beginnen mit D, die der Einzelpersonen mit E.

Ihre Partnerin Delia schneidet im Vortest relativ schwach ab, verzeichnet im Nachtest aber einen beachtlichen Wissenszuwachs in allen Untertests. Bei Dyade 2 ist das Ausgangsniveau der beiden Teilnehmerinnen Dido und Dora in etwa dasselbe. Während Dido sich in den Untertests Theorie und Methodologie um insgesamt drei Punkte verbessert, verliert Dora im Untertest Empirie einen Punkt, erreicht aber im Untertest Methodologie eine größere Steigerung als Dido.

Die Transkripte wurden zur Analyse in sogenannte Schritte oder Sprechakte gegliedert. Die Schritte wurden durchnummeriert. In ihrem Ausmaß ähneln die Schritte der grammatikalischen Einheit des Satzes. Während in der Grammatik jedoch formale Eigenschaften beschrieben werden, stehen bei der Analyse verbaler Daten die funktionalen, pragmatischen Eigenschaften einer Äußerung, also der Zweck, den ein Sprecher mit seiner Äußerung verfolgt, im Vordergrund (Searle, 1971). Ein Schritt umfaßt genau eine solche Sinneinheit, kann jedoch ein Wort oder einen ganzen Satz enthalten. Betrachtet man die Anzahl der Schritte in allen Transkripten, so sind diejenigen der Dyaden ($M = 1456$, $SD = 392$) erwartungsgemäß umfangreicher als diejenigen der Einzelpersonen ($M = 618$, $SD = 175$; nach Métrailler, 2002). Dyaden äußern etwas mehr als doppelt so viele Schritte wie Einzelpersonen, so daß der Redeanteil eines Dyadenmitglieds in etwa demjenigen einer Versuchsperson in der Einzelbedingung entspricht. Im Einzelnen beträgt die Länge des Transkripts von Ella 726 Schritte, das von Emelie 507 Schritte. Dana und Delia produzieren gemeinsam 873 Schritte, während Dido und Dora auf 1522 Schritte kommen.

Im Folgenden wird die qualitative Analyse zu den ausgewählten Einzelpersonen und Dyaden dargestellt. Die Abfolge der experimentellen Zyklen ist jeweils in einer Tabelle zusammengefaßt (siehe Tabellen 15 bis 18). Mit experimentellen Zyklen ist eine Hypothese und das dazugehörige Experiment gemeint. Werden mehrere Hypothesen nacheinander aufgestellt oder zu einer Hypothese mehrere Experimente durchgeführt, so wird dies gesondert betrachtet. Den Abschluß des Kapitels bildet eine Auswertung der gesondert kategorisierten Metaaussagen.

10.1 Ella: Systematische Exploration des Experimenterraums

Einen Überblick über den Ablauf von Ellas Experimentalreihe gibt Tabelle 15. Nachdem sie sich einen Überblick über die verschiedenen Funktionen von *virtue* verschafft hat, öffnet sie das Hypothesenscratchpad für ihre erste Hypothese.

Ella: Zyklen 1 bis 4

Ella testet zunächst den Faktor Größe der Anordnung in der Erwartung, daß die Reaktionszeit mit zunehmender Größe zunehme. Die Wahl des Faktors für die erste Hypothese begründet sie nicht, sie entspricht jedoch einem ersten Test der vorgegebenen Theorie zur visuellen Suche. Der Suchtyp bleibt in dieser Hypothese und im dazugehörigen Experiment unberücksichtigt, wird nicht diskutiert und auf der Stufe Merkmalsuche konstantgehalten. Das Ergebnis überrascht (Transkript 1):

69 jetzt kann ich da outputs haben, und dann möchte ich gerne alles (klickt
alle drei an)
70 Mittelwert ist, mm
71 Höher, was?
72 macht gar keinen Unterschied
73 die Reaktionszeit ist praktisch gleich
74 Und zwar ist sie sogar noch kleiner, wenn, wenn wir viele Objekte haben
75 Das ist komisch
76 Em
77 Keine Signifikanz, weil sie praktisch gleich ist
78 Em
79 Also eigentlich
80 Müsste man diese Hypothese verwerfen
81 Weil die Reaktionszeit ist für nur sieben Objekte kleiner als für
fünfundzwanzig
82 Okay
83 (liest nochmals ihre Hypothese)
84 muss ich verwerfen
85 die Begründung
86 Ist
87 em
88 (schreibt Hypothesenbewertung) „die sieben Objekte zu (...) undzwanzig“
89 (schliesst das Resultatfenster, 1. Experiment)
90 Restzeit
91 Gehen wir noch einmal eine Hypothese aufstellen
92 Dann nehmen wir
93 Set size möchte ich trotzdem noch einmal nehmen

Transkript 1. Ella wundert sich über das Ergebnis ihres ersten Experiments.

Ella bemerkt sofort, daß das Ergebnis nicht ihren Erwartungen entspricht und äußert dies in der Metaaussage „das ist komisch“ (Schritt 75). Sie betrachtet alle

Tabelle 15
Experimentelle Zyklen von Einzelperson 1: Ella

Zyklus	Hypothese Text	Versuchsplan und Ergebnisse																	
		Bewertung	Suchtyp	Farbähnlichkeit	Größe der Anordnung	Farbkom- bination	Exzen- trizität	Merkmals- Verbind.s.	ähnlich versch.	7	13	19	25	31	pink...	pink...	gering mittel	Exzen- trizität	
1	Je mehr Items im Suchraum vorhanden sind, desto länger dauert es, das Zieltarget zu finden.	V	O	O	x	x	O	O	x	O	O								O
2	Wenn die Objekte verschiedene Farbe haben, dauert die Suche nach dem Zielobjekt weniger lang, als wenn die Objekte ähnliche Farbe haben.	B	O	x	O	O	O	O											O
3	Je mehr Objekte vorhanden sind, desto länger dauert die Suche bei conjunction search.	B	X	O	X	X	O	O											O
4	Je mehr Objekte, desto länger dauert conjunction search. (Alle mögl. Anzahlen von Objekten)	B	X	O	X	X	O	O	X	X	X	X	X	X	X	X	X	X	O
5	Die Farbkombination hat eine Auswirkung auf die Reaktionszeiten.	V	O	O	O	O	O	O											O
6	Farkombination bei den beiden Sucharten eine Rolle spielt.	B	x	x	O	O	x	x											O
7	Der Abstand des Zielreiz vom Fixationspunkt hat eine Auswirkung auf Reaktionszeit.	V	O	O	O	O	O	O											O
8	Eccentricity und feature search: Abstand des Zielreizes vom Fixationspunkt hat Auswirkung.	V	O	O	O	O	O	O											O
9	Anzahl Objekte und Abstand vom Fixationspunkt wirken sich aus.	B	O	O	X	X	O	O	X	X	X	X	X	X	X	X	X	X	O
10	Auswirkung von Farbähnlichkeit und -kombination unterschiedlich.	B	O	x	O	O	x	x											O
11 a	Bei ähnl. Farben, viel Objekten, großer eccentricity und conjunction search = hohe RZ.	ungeprüft																	
b	Farbkombination und Abstand wirken sich aus.	b	O	O	O	O	O	O											O
12	Farbkomb. + Eccentricity wirken sich bei feature-suche aus.	V	O	O	O	O	O	O											O
13	Art der Suche und Farbähnlichkeit haben Einfluss.	B	x	x	O	O	x	x											O
14	Farbkomb. und Searchtype spielen Rolle.	V	x	x	O	O	O	O											O
15	Farbkomb. wirkt sich bei zunehmender Anzahl Objekte aus.	B	O	O	X	X	O	O	X	X	X	X	X	X	X	X	X	X	O
16	Farbkomb. und Set size haben bei feature untersch. Einfluss.	b?	O	O	X	X	O	O	X	X	X	X	X	X	X	X	X	X	O

Anmerkung. B = korrekte Beibehaltung der Hypothese bei bestätigenden Daten; b = falsche Beibehaltung der Hypothese bei widersprechenden Daten; V = korrekte Verwerfung der Hypothese bei widersprechenden Daten; v = falsche Verwerfung der Hypothese bei bestätigenden Daten. O = Konstanzhaltung auf dieser Stufe; x = Variation eines Faktors auf dieser Stufe ohne signifikanten Effekt; x = signifikanter Haupteffekt; X = signifikanter Haupteffekt und signifikante Interaktion.

drei Arten der Ergebnisdarstellung, deskriptive Werte, Inferenzstatistik und Interaktionsplot und verwirft korrekterweise ihre Hypothese (Schritte 79 bis 81). Ella verfolgt jedoch den Grund dafür, daß ihre Erwartungen nicht eingetroffen sind, nicht weiter, sondern geht einfach zum nächsten experimentellen Zyklus über. Sie äußert, daß sie den Faktor Größe der Anordnung nochmals untersuchen möchte, bezieht ihn jedoch nicht in ihren nächsten experimentellen Zyklus ein.

Ella kommt auf ihre Verwunderung über ihre Ergebnisse vom Anfang in Zyklus 3 zurück (Transkript 2). Sie liest nochmals in der Theorie nach und erinnert sich an den Unterschied zwischen Merkmals- und Verbindungssuche. Ella macht in Schritt 141 eine Metaaussage. Sie äußert sich erstaunt darüber, daß die Größe der Anordnung keine Auswirkung auf die Reaktionszeit hat und sucht nach einem Grund dafür. Diesen leitet sie in den Schritten 145 bis 147 ab. Dies beeinflußt ihre weitere Planung. Sie möchte nun die Verbindungssuche näher untersuchen und entscheidet sich schließlich dafür, das kritische Experiment durchzuführen.

139 gut, dann gehen wir in die Theorie schauen
 140 em
 141 das ist einfach schon, mit der Ziel, mit der Anzahl von Distraktoren, das hat mich ein bisschen erstaunt
 142 aber ich mache natürlich, aha, ja, ich mache natürlich immer, eh
 143 nn
 144 noch einmal
 145 was ist das, diese Merkmalssuche
 146 dort spielen ja
 147 die Distraktoren keine Rolle
 148 aber Ähnlichkeit der Farben spielt eine Rolle
 149 und jetzt könnte ich ja diese Verbindungssuche machen, weil dort spielt eben die Anzahl, em
 150 eine (?) Rolle

Transkript 2. Ella erinnert sich an die Auswirkungen von Verbindungs- und Merkmalssuche.

So führt Ella im dritten und vierten Zyklus jeweils das kritische Experiment mit den Faktoren Suchtyp und Größe der Anordnung durch. Ab dem dritten Zyklus schaut Ella, wenn sie die Ergebnisse ihrer Experimente aufruft, immer zuerst den Interaktionsplot an, bevor sie sich der Inferenzstatistik zuwendet. Dadurch ist sie immer über die Richtung ihrer Effekte informiert (Transkript 3). Bei der Inferenzstatistik achtet Ella auch auf Interaktionen und versucht, diese zu interpretieren.

294 Was ist das für eine Interaktion?
295 Dass sich je nach Anzahl Objekten die Art der Suche signifikant auswirkt
296 Mh
297 Und dass die feature Suche immer schneller geht, für search type

Transkript 3. Ella interpretiert die statistische Interaktion des Experiments aus Zyklus 3.

Auch den vierten Zyklus beginnt Ella mit einem Blick in die Theorie. Sie wiederholt das kritische Experiment, möchte eigentlich ein dreifaktorielles Experiment durchführen und bekommt eine Warnung, daß in *virtue* nur zweifaktorielle Experimente durchführbar sind. So ändert sie gegenüber dem Design des letzten Experiments die Stufe, auf der sie den Faktor Farbähnlichkeit konstant hält, von ähnlich auf unterschiedlich. Damit führt sie zusammen mit dem Experiment in Zyklus 3 quasi ein dreifaktorielles Experiment durch (im Folgenden werden solchermaßen zusammengehörende Experimente als pseudo-dreifaktoriell bezeichnet), vergleicht die Ergebnisse aus den Zyklen 3 und 4 jedoch nicht und gewinnt somit auch keine neuen Erkenntnisse über den Faktor Farbähnlichkeit.

Da in *virtue* nur zweifaktorielle Experimente möglich sind, können Wechselwirkungen zwischen drei Faktoren nicht überprüft werden. Einen Ausweg bietet die zweimalige Durchführung desselben Experiments, an dem nur die Konstanthaltung eines einzigen Faktors verändert wird. Dabei können allerdings Interaktionen mit dem über die Konstanthaltung variierten Faktor nicht überprüft werden.

Ella erinnert sich in Zyklus 3 an ihre Verwunderung über die Ergebnisse in Zyklus 1. Dies zahlt sich aus, handelt es sich doch beim Faktor Größe der Anordnung um einen der beiden in der Theorie beschriebenen Faktoren, zu dem sie ein auf den ersten Blick widersprüchliches Ergebnis erhalten hatte.

Die Ergebnisse läßt Ella sich in allen drei möglichen Formaten anzeigen und weiß diese auch zu nutzen. Das kritische Experiment führt Ella bereits in Zyklus 3 durch. Dies wirkt sich deutlich auf ihre weiteren Experimente aus. Sie bezieht den Faktor Suchtyp ab diesem Zyklus meist in variiertes Form oder in pseudo-dreifaktoriellen Experimenten ein. In den drei Experimenten, in denen sie den Suchtyp konstant hält, tut sie dies auf der Stufe der Verbindungssuche während sie in den ersten beiden Experimenten für diesen Zweck die Verbindungssuche ausgewählt hatte.

Ella: Zyklus 5

Im Vorfeld des fünften Zyklus sucht Ella in der Theorie nach Informationen zum Faktor Farbähnlichkeit, findet nichts und erinnert sich an die Testung dieses Faktors in Zyklus 2. Sie läßt ihre zunächst formulierte Hypothese dazu ungetestet und wendet sich dem Faktor Farbkombination zu. Warum sie sich gerade für diesen Faktor entscheidet, begründet sie nicht. Zu diesem Faktor formuliert Ella zum ersten Mal eine ungerichtete Hypothese. Da sie keinen Effekt entdecken kann, verwirft Ella ihre Hypothese korrekterweise.

Ihr Verhalten korrespondiert hier mit dem Ergebnis der kumulierten Daten, daß die Lernenden gerade zum Faktor Farbkombination selten gerichtete Hypothesen formulieren. Diese ungerichtete Hypothese hat entscheidende Auswirkungen auf alle folgenden Hypothesen. Von diesem Punkt an formuliert Ella nämlich nur noch eine gerichtete Hypothese in Zyklus 11, die sie nicht experimentell prüft. Sie postuliert lediglich in zwei Fällen einen Unterschied für verschiedene Faktorstufen und beschränkt ihre Vorhersagen ansonsten darauf, daß sich die Faktoren unspezifisch auswirken.

Ella: Zyklus 6

Bei der Versuchsplanung in Zyklus 6 verwechselt Ella die Faktoren Farbkombination und Farbähnlichkeit. In Fortsetzung von Zyklus 5 formuliert sie eine Hypothese zu den Faktoren Farbkombination und Suchtyp, variiert aber bei der Versuchsplanung den Faktor Farbähnlichkeit. Bei der Betrachtung der Ergebnisse wundert sich Ella, daß ihr Farbfaktor nun doch einen signifikanten Haupteffekt ergibt und erinnert sich an Zyklus 2, als sie die Farbähnlichkeit erstmals überprüft hatte. Ella schaut sich die Ergebnisse der beiden Experimente nochmals genau an und vergleicht (Transkript 4). Die Verwechslung von Farbkombination und Farbähnlichkeit klärt Ella nur implizit auf, indem sie sich in der Bewertung der Hypothese statt auf die Farbkombination auf die Farbähnlichkeit bezieht. Damit bestätigt sie ihre Hypothese jedoch fälschlicherweise.

Ella schaut in diesem Zyklus Ergebnisse im Archiv von *virtue* nach. Dies ist hervorzuheben, da Versuchspersonen nur in den seltensten Fällen alte Experimente im Archiv anschauen. Es fördert die Vertiefung des erworbenen Wissens über den Gegenstandsbereich. Ella unterläuft in diesem Zyklus eine Verwechslung von

396 die feature hat wieder weniger lange und jetzt
 397 hat
 398 similarity of colors eben doch eine Auswirkung
 399 und zwar, wenn man da bei beiden
 400 komisch
 401 bei beiden ist, wenn es ähnliche Farben sind, geht es länger
 402 bei beiden
 403 aber bei der, bei der feature ist es nicht so ausgeprägt
 404 und zwar bei ähnlichen Farben ist es bei der siebeneinundachtzig und
 sechs(?)
 405 und da ist es, der Unterschied zwar etwas gleich, he? Conjunction
 406 keine Interaktion
 407 aber
 408 was ist denn das vorher? (schaut nochmals das alte Experiment an)
 409 blue/yellow, set size
 410 was haben wir da gehabt, set size?
 411 Neunzehn?
 412 Das ist das Experiment sechs (schaut nochmals das Design nach)
 413 Neunzehn
 414 Ist genau gleich
 415 Wie da (öffnet Experiment 2)
 416 Da ist es einfach mit dreizehn
 417 Aber nur mit feature
 418 Okay, und da hat es beides
 419 (schreibt Hypothesenbewertung) „signifikantes Ergebnis für die Farbkombi-
 Farb, für Farbähnlichkeit“
 420 dann behalten wir es bei

Transkript 4. Ella vergleicht die Ergebnisse der Experimente aus den Zyklen 6 und 2.

Farbkombination und Farbähnlichkeit. Obwohl sie eine Hypothese zur Farbkombi-
 nation aufstellt, führt sie ein Experiment zur Farbähnlichkeit durch und interpretiert
 ihre Hypothese auch im Sinne der Farbähnlichkeit. Eigentlich kann sie mit
 ihrem Experiment gar keine Aussage über diese Hypothese machen und kommt
 deshalb, als sie die Hypothese bestätigt, zu einer Fehlentscheidung.

Ella: Zyklen 7 bis 9

Die Entscheidung, im Zyklus 7 die Exzentrizität zu überprüfen, begründet Ella
 wieder nicht ausdrücklich. Sie verfolgt die Frage der Farbkombination zunächst
 nicht weiter. Stattdessen konzentriert sie sich nun auf denjenigen Faktor, den sie
 bisher noch überhaupt nicht beachtet hat, die Exzentrizität. Ella erhält in ihrem ein-
 faktoriellen Experiment ein nicht-signifikantes Ergebnis und plant sofort einen
 weiteren Zyklus, in dem sie den Faktor Suchtyp auf Merkmalssuche setzt, nach-
 dem sie ihn zunächst auf Verbindungssuche konstantgehalten hatte. Sie kommt
 wieder auf ein nicht-signifikantes Ergebnis und plant ein weiteres Experiment mit
 den Faktoren Exzentrizität X Größe der Anordnung.

In Zyklus 9 möchte Ella wirklich wissen, was es mit der Exzentrizität auf sich hat und plant ein großes Experiment mit 15 Zellen. Sie ist sich der Größe bewußt und verweist auf die experimentelle Ökonomie mit den folgenden Worten (Transkript 5):

488 das braucht jetzt viel Leute
 489 ui
 490 wieder mit dreißig?
 491 Ja, nehmen wir zwanzig
 492 Ich weiß nicht, ob sich das lohnt

Transkript 5. Ella stellt Überlegungen zur experimentellen Ökonomie an.

Ella schaut bei der Abfrage der Ergebnisse nicht nur auf Signifikanzen, sondern interpretiert ihre Ergebnisse auch inhaltlich (Transkript 6). Ausgehend vom Interaktionsplot versucht sie, die Haupteffekte und die Interaktion detailliert für einzelne Stufen der jeweiligen Faktoren zu verstehen. Die Planung eines weiteren Experiments zur Exzentrizität in Zyklus 9 nach zwei nicht signifikanten Resultaten in den Zyklen 7 und 8 verdeutlicht Ellas Konstanz. Auch nach zwei Experimenten ohne signifikantes Ergebnis für den Faktor Exzentrizität möchte sie noch nicht hinnehmen, daß dieser Faktor keine Bedeutung für die visuelle Suche haben soll, und untersucht ihn erneut unter veränderten Bedingungen. In Zyklus 9 bekommt sie dann die Bestätigung für ihre Vermutung über die Bedeutung der Exzentrizität.

510 ist das wohl der Haupteffekt von eccentricity?
 511 sehr
 512 und set size haben wir eigentlich schon gewußt
 513 und (?) mit der Interaktion
 514 und zwar
 515 daß sich die Anzahl
 516 was ist denn da die Interaktion?
 517 Dass sich
 518 Dass zum Beispiel
 519 Wenn sie mittelnah sind, dann finden sie sie sogar weniger, haben sie länger
 520 Als wenn sie ganz weit weg sind, das ist noch
 521 Aha, es kommt auf die Anzahl Objekte an
 522 Und erst, wenn's neunzehn und mehr sind, brauchen sie nachher wirklich länger
 523 Und bei sieben und dreizehn
 524 Geht's am längsten, wenn sie
 525 mittelweit weg sind

Transkript 6. Ella interpretiert Haupteffekte und Interaktion des Experiments in Zyklus 9.

Ella: Zyklen 10 bis 12

Nun kommt Ella nochmals auf den Faktor Farbkombination zurück. An dieser Stelle wird deutlich, daß sie die Verwechslung von Farbkombination und Farbähnlichkeit für sich ausgeräumt hat. Beide Farbfaktoren kombiniert sie in Zyklus 10 in einem zweifaktoriellen Experiment. Als sie wieder keinen Effekt für die Farbkombination findet, schlägt sie nochmals in der Theorie nach und formuliert für Zyklus 11 zunächst eine sehr spezifische Hypothese, die sich jedoch nicht experimentell prüfen läßt. Sie besinnt sich, überprüft ihr bisheriges Wissen (Transkript 7) und entscheidet sich dafür, den Faktor Farbkombination noch näher zu betrachten.

534 jetzt wissen wir einmal
 535 wirklich weit weg und viel, dann geht es einfach lange
 [...]

580 Was könnte ich denn noch überprüfen?
 581 Dass
 582 Ich gar nicht so viel variieren kann
 583 das mit diesen Farben ist ja schon noch nicht ganz klar
 584 dann nehmen wir diese color combination

Transkript 7. Ella überprüft ihr bisher erworbenes Wissen und plant ihr weiteres Vorgehen.

In diesem Zyklus unterläuft Ella erneut ein Fehler bei der Hypothesenbeurteilung. Obwohl sie für die Farbkombination keinen Effekt erhält, bestätigt sie ihre Hypothese. In der Begründung ihrer Entscheidung äußert sie zwar, daß die Farbkombination „unwichtig“ sei, zieht aber für die Gesamtbewertung keine Konsequenz daraus und begeht damit einen Bestätigungsfehler. Ella repliziert dieses Experiment in Zyklus 12 für die Merkmalsuche und führt damit zusammen mit dem vorangegangenen Experiment wieder ein pseudo-dreifaktorielles Experiment durch. Sie wundert sich, daß sie hier wieder keine Effekte vorfindet, verwirft ihre Hypothese dieses Mal jedoch korrekt.

Ella: Zyklen 13 bis 16

Nach zwölf Zyklen fällt es Ella schwer, noch den Überblick zu wahren, und so kommt es, daß sie unabsichtlich das Experiment Farbähnlichkeit X Suchtyp repliziert, ohne im Archiv nachzuschauen. In den restlichen Zyklen geht es ihr darum, ihre Erkenntnisse zu festigen (Transkript 8). Planungsfragen und die Zusammenfassung der bereits erworbenen Kenntnisse stehen im Vordergrund. Die Kommentierung der einzelnen Experimente wird immer knapper.

633 was spielt denn bei der feature überhaupt eine Rolle?
 634 Verwerfen
 635 (schreibt Hypothesenbewertung) „für feature Suche, eccentricity und
 Farbkombination unwichtig“
 636 was ist denn für die feat? (schliesst Resultatfenster, 12. Experiment)
 637 was ist denn für die feature überhaupt wichtig?
 638 (liest nochmals Instruktionen)
 639 set size hat keine Auswirkung
 640 eccentricity hat keine Auswirkung
 641 dann nehmen wir halt die und schauen, und schauen, ob sie
 642 das habe ich auch schon gemacht
 643 mm
 644 ich weiß nicht, was ich noch überprüfen könnte

Transkript 8. Ella rekapituliert ihre bisherigen Ergebnisse.

Nachdem Ella den Faktor Farbkombination als einzigen mit allen anderen Faktoren in zweifaktoriellen Experimenten überprüft hat, entscheidet sie ohne weiteren Kommentar, ihre Experimentalreihe an dieser Stelle abzuschließen. Die gründliche Untersuchung dieses Faktors ist auf die unklaren, nicht signifikanten Ergebnisse zurückzuführen. Diese motivieren Ella zu einer kritischeren Überprüfung als die signifikanten Ergebnisse zu den anderen Faktoren.

Ella: Zusammenfassung

Ella kommt zum kritischen Experiment in den Zyklen 3 und 4 aufgrund einer Metaaussage. Sie wundert sich, daß der von ihr erwartete Effekt des Faktors Größe der Anordnung nicht eintritt, und sucht nach Gründen dafür. Dies führt sie zu dem weiteren Plan, nun die Verbindungssuche zu untersuchen und zunächst das kritische Experiment durchzuführen. Ella zieht also Konsequenzen aus ihrer Metaaussage.

Auffällig ist, daß Ella vor Durchführung des kritischen Experimentes in den Zyklen 3 und 4 den Faktor Suchtyp auf der Stufe Merkmalssuche konstanthält. Danach wählt sie für die Konstanthaltung immer zuerst die Stufe Verbindungssuche. Die Stufe Merkmalssuche wählt sie nur noch für pseudo-zwei- bzw. pseudo-dreifaktorielle Experimente (Zyklen 7 & 8, 11 & 12, 15 & 16), wobei sie dann immer die Verbindungssuche zuerst untersucht. Auch was den Faktor Größe der Anordnung anbelangt, wird Ella nach dem kritischen Experiment systematischer. Während sie vorher nur zwei Stufen in ihre Versuchspläne einbezieht und bei der Konstanthaltung nicht den mittleren Wert wählt, prüft sie in den Experimenten

danach immer alle Faktorstufen oder hält auf dem mittleren Wert die Größe der Anordnung konstant. Den Faktor Farbkombination hält sie durchgehend auf der Stufe blau/gelb konstant, die Farbähnlichkeit auf der Stufe verschieden. Hier wird eine ausgesprochen systematische Vorgehensweise evident, die einen Vergleich der experimentellen Ergebnisse über mehrere Experimente hinweg überhaupt erst ermöglicht.

Die Reihenfolge, in der Ella die Faktoren Farbähnlichkeit, Farbkombination und Exzentrizität überprüft, die jeweils nicht in der Theorie erwähnt werden, entspricht derjenigen, in der diese Faktoren im Hypothesenscratchpad aufgeführt sind. Hierbei handelt es sich im Sinne von Simon, Langley und Bradshaw (1981) um eine schwache Heuristik, orientiert an den Oberflächenmerkmalen der Lernumgebung und nicht an inhaltlichen Aspekten der Fragestellung.

Ella bringt bereits ein beträchtliches Vorwissen für die Experimente mit und macht davon Gebrauch, indem sie sich etwa bei der Interpretation die Ergebnisse immer in allen drei Formaten ausgeben läßt und diese Formate auch in ihre Schlußfolgerungen einbezieht. An lernförderlichem Verhalten fallen bei Ella mehrere Strategien auf. Nach den ersten Experimenten schaut sie jeweils in der Theorie nach. Sie erinnert sich an die Ergebnisse früherer Experimente und greift bei Unsicherheiten auf das Archiv zurück. Sie geht auf unerwartete Ergebnisse ein und verfolgt diese weiter. Und Ella zeigt eine hohe Zielstrebigkeit beim Planen ihrer Experimente. Auch wenn sie die erwarteten Effekte nicht gleich findet, strengt sie neue Experimente zu den betreffenden Faktoren an. Der Faktor Farbkombination, dessen Daten ihre Vorhersagen am meisten enttäuscht, wird am gründlichsten untersucht.

10.2 Emelie: Lernen trotz aller Fehler

Der Ablauf von Emelies Experimentalreihe ist aus Tabelle 16 ersichtlich. Sie beginnt ihre Arbeit mit *virtue* mit einer systematischen Exploration der Lernumgebung. Sie öffnet alle Fenster, untersucht deren Funktionalität und beschließt dann, eine erste Hypothese aufzustellen. Dabei wird deutlich, daß sie die Operationalisierung der Fragestellung zur visuellen Suche noch nicht verstanden hat (Transkript 9).

Tabelle 16

Experimentelle Zyklen von Einzelperson 2: Emelie

Zyklus	Hypothese Text	Versuchsplan und Ergebnisse						
		Bewer- tung	Such- typ	Farbähn- lichkeit	Größe der Anordnung	Farbkombi- nation	Exzen- trizität	
1	Der Zielreiz wird schneller gefunden, wenn sehr unterschiedliche Farben benutzt werden, als wenn ähnliche Farben benutzt werden.	(B)	O	X X	O	31	X X	O
2	Je weiter der Zielreiz vom Zentrum des Displays weg ist, desto länger ist die Reaktionszeit.	(B)	O	x x	O	13	O	x x
3	Zusätzlich nochmalige Überprüfung der Hypothese aus 1	b	X X	O	X X X X X	25	O	O
4	Verbindungssuche wird von dem set-size beeinflusst, während die merkmalssuche kaum oder nicht von dem set-size beeinflusst wird.	B	X X	O	X X X X X	16	O	O
5	die Verbindungssuche hängt von dem set-size und von der eccentricity ab.	b	X X	O	X X X X X	7	O	O
6	Je ähnlicher sich die Farben sind und je mehr Objekte auf dem Display erscheinen, desto grösser ist die Reaktionszeit.	v	O	O	x x x x x	31	x x	O
7	Je grösser der set-size ist und je ähnlicher die Farben sind, desto grösser ist die Reaktionszeit.	B	O	X X	X X X X X	13	O	O
8	Je weiter der Zielreiz vom Zentrum entfernt (eccentricity) ist und je grösser der set-size ist, desto grösser die Reaktionszeit.	B	O	O	X X X X X	16	O	X X X

Anmerkung. B = korrekte Beibehaltung der Hypothese bei bestätigenden Daten; b = falsche Beibehaltung der Hypothese bei widersprechenden Daten; V = korrekte Verwerfung der Hypothese bei widersprechenden Daten; v = falsche Verwerfung der Hypothese bei bestätigenden Daten. Eingeklammerte Hypothesenbewertungen zeigen an, daß zwar die inferenzstatistischen Ergebnisse die Hypothese belegen, die Interpretation im Transkript jedoch gegenteilig war. O = Konstanthaltung auf dieser Stufe; x = Variation eines Faktors auf dieser Stufe ohne signifikanten Effekt; X = signifikanter Haupteffekt; X = signifikanter Haupteffekt und signifikanter Haupteffekt und signifikante Interaktion. Replikationen werden grau unterlegt.

Emelie: Zyklus 1:

Emelie läßt sich die Operationalisierung der Fragestellung nochmals vom Versuchsleiter erklären und ist dann in der Lage, ihre erste Hypothese aufzustellen (Transkript 9). Als Faktor für ihre erste Hypothese wählt Emelie die Farbähnlichkeit. Ihre Hypothese ist einfaktoriell, gerichtet und sauber formuliert. Bei der Versuchsplanung nimmt sie den Faktor Farbkombination mit auf, über den sie keine Hypothese formuliert hat. Während sie festlegt, auf welchen Stufen die nicht variierten Faktoren konstant gehalten werden sollen, bemerkt sie dezidiert, daß sie den Faktor Suchtyp erst später untersuchen möchte (Schritt 85: „Auf search type gehe ich jetzt noch nicht ein, es hat sicher nachher noch einen Einfluß, aber jetzt will ich einfach einmal das wissen.“).

-
- 15 Also, ich muss zuerst eine Hypothese aufstellen, aber über was muss ich jetzt diese Hypothese aufstellen
- 16 Muss ich jetzt das durchführen, wie es in dieser Theorie ist, wie jemand auf dem Fussballfeld gefunden werden kann?
- 17 Da bin ich aber nicht sicher, hm
- 18 (geht Versuchsleiter (VL) holen)
- 19 Emelie (?) nicht gewusst, über was ich suchen muss. ist das nicht gegeben? muss ich irgend mir etwas einfallen lassen, oder wie?
- 20 VL du hast vorher die Theorie gelesen?
- 21 Emelie muss ich, also, ob jetzt jemand, wie jemand im Fussballfeld gefunden werden kann. Muss ich jetzt das suchen, oder?
- 22 VL nein, du musst jetzt einfach, es geht um die visuelle Suche
- 23 Emelie eben, ja, aber
- 24 VL und
- 25 Emelie ich habe ja gar nichts, was ich suchen muss
- 26 VL du hast Möglichkeiten, über die visuelle Suche etwas zu erfahren
- 27 Emelie ja
- 28 VL also, die Theorien, die es gibt
- 29 Emelie (gleichzeitig) genau, die habe ich gelesen
- 30 VL (gleichzeitig) zu überprüfen. Und für das musst du dir überlegen, was kann sein an der visuellen Suche, kannst du Hypothesen aufstellen, kannst du entsprechende Experimente dazu designen, durchführen, bekommst Resultate dafür und am Schluss solltest du dann wissen, was bei der visuellen Suche abgeht.
- 31 Emelie also ich muss einfach, em, die visuelle Suche bestimmen zu versuchen
- 32 VL herausfinden
- 33 Emelie was es ist
- 34 VL was für Einflüsse dass es dort gibt
- 35 Emelie mh, okay
-

Transkript 9. Emelie reflektiert die Aufgabenstellung.

Von den Ergebnissen ihres ersten Experiments läßt sich Emelie nur den Interaktionsplot anzeigen. Emelie stützt ihre Beurteilung auf rein qualitative Kriterien und kommt zu dem Schluß, daß weder der Faktor Farbkombination noch der Faktor Farbähnlichkeit einen großen Einfluß haben (Transkript 10). Dennoch entscheidet sie sich dafür, ihre Hypothese beizubehalten, begründet diese Entscheidung jedoch nicht. Die Beibehaltung der Hypothese entspricht damit der statistischen Evidenz und wird in Tabelle 16 daher auch als korrekt bestätigt aufgeführt. Um den Widerspruch zu Emelies verbal geäußelter Interpretation auszudrücken, wurde diese Beurteilung jedoch in Klammern gesetzt.

112 Also, da sehe ich jetzt, dass blue/ yellow or blue, also blau/ gelb und
 blau/ hellblau, also blau/ gelb ist ein bisschen besser als grün und pink,
 aber nur minim, und da, bei den unterschiedlichen, sind sie dann gleich
 113 Wieso sind die jetzt da, wieso gibt es keinen grösseren Unterschied?
 114 Ähnlichkeit der Farben, wieso gibt es zwischen pink/ grün und pink/ rot
 nicht einen grösseren Unterschied? Hm?
 115 Das heisst ja eigentlich, dass das keinen so grossen Unterschied macht, ob
 jetzt die Farben sich deutlich, deutlich variieren oder nicht
 116 Hm?
 117 Okay, dann muss ich halt noch etwas anderes hineinnehmen
 118 Einmal schauen, wie es dann aussieht
 119 Jetzt wollen wir einmal weiter
 120 (erscheint Warnschild, dass sie das Experiment bewerten muss) okay
 121 (Warnschild erscheint wieder) ich habe jetzt ja meine Hypothese bewertet
 122 mh
 123 also, ich behalte sie einmal bei, ich führe sie noch ein bisschen aus

Transkript 10. Emelie interpretiert die Ergebnisse aus Zyklus 1.

Emelie betrachtet bei der Interpretation der Ergebnisse von Zyklus 1 nur den Interaktionsplot. Dies setzt sich im weiteren Verlauf fort. Über die gesamte Experimentalreihe hinweg läßt Emelie die Interferenzstatistik außen vor, auch die dazugehörigen statistischen Begriffe diskutiert sie nicht. Ab Experiment 3 bezieht sie die deskriptive Statistik in ihre Betrachtung mit ein. Aussagen über die statistische Bedeutsamkeit der von ihr untersuchten Effekte scheint sie nicht zu vermissen.

Daß Emelie die Inferenzstatistik außer acht läßt, hat Folgen für die Bewertung der Hypothesen. Emelie kann statistisch bedeutsame Effekte im Interaktionsplot nicht erkennen, setzt sich dann jedoch über ihre qualitative Einschätzung der Effekte hinweg, indem sie eine Art Bestätigungsfehler begeht, und kommt schlußendlich durch einen doppelten Fehler zu einer korrekten Schlußfolgerung.

Emelie hat also anscheinend keinen guten Start in ihre Experimentalreihe und, so stellt sich die Frage, warum es dennoch interessant sein könnte, gerade ihren Fall weiterzuverfolgen. Aus den Ergebnissen der Nachtests geht hervor, daß Emelie durch die Arbeit mit *virtue* einen Lernzuwachs erreichen konnte. Daher stellt sich die berechtigte Frage, wie dies unter diesen schwierigen Bedingungen möglich war. Welche Lernprozesse ermöglichten es Emelie dennoch, neues Wissen zu erwerben? Zudem ist es gerade bei computerbasierten Lernsystemen von Bedeutung, in Erfahrung zu bringen, welchen Problemen schwächere Lernende dabei begegnen. In vielen Fällen profitieren gerade stärkere Lernende von solchen Lernsystemen und schwächere haben mit Schwierigkeiten zu kämpfen, so daß sich der Abstand zwischen starken und schwachen Lernenden eher vergrößert. In diesem Sinne ist es von besonderem Interesse, die Schwierigkeiten der schwächeren Lernenden zu analysieren, um Verbesserungsmöglichkeiten für computerbasierte Lernsysteme zu eruieren und gezielte Maßnahmen zur Unterstützung dieses Personenkreises initiieren zu können.

Emelie: Zyklus 2

Im zweiten Zyklus geht Emelie den in Zyklus 1 aufgeworfenen Fragen zunächst nicht nach, sondern wendet sich dem Faktor Exzentrizität zu und formuliert dazu ihre nächste Hypothese, wiederum gerichtet und einfaktoriell. Die Entscheidung für diesen Faktor trifft sie nach ihrem spontanen Interesse und äußert dazu in Schritt 143: „Ich denke, ein wichtiger Punkt ist sicher das eccentricity“.

Bei der Versuchsplanung entschließt sich Emelie dann, ihre erste Hypothese zum Faktor Farbähnlichkeit noch einmal zu testen und begründet dies mit den folgenden Worten: „Also, der Witz ist ja eigentlich, daß alle die miteinander zusammenhängen, darum teste ich die erste auch noch einmal, also, es hängt ja alles miteinander zusammen“ (Schritt 159). Wie in Zyklus 1 führt Emelie auch in diesem Zyklus ein zweifaktorielles Experiment durch, hat aber eigentlich nur einfaktorielle Hypothesen. In diesem Fall liegen immerhin durch Hinzunahme der Hypothese aus dem 1. Zyklus für beide Faktoren Hypothesen vor.

Bei der Auswertung der Ergebnisse zum zweiten Zyklus kommt Emelie zu dem Fehlschluß, daß die Daten denjenigen aus Zyklus 1 ähneln (Transkript 11), aus ihrer Sicht also keine Effekte vorliegen. Sie übersieht dabei zum einen, daß in

209 Jetzt habe ich wieder praktisch die gleichen Daten, das ist ja jetzt irgendwie komisch

210 Das ist jetzt wirklich komisch

211 Also

212 huu

213 Also, da ist die Ähnlichkeit der Farben, da ist es gleich, und da ist es sehr unterschiedlich

214 Und dann ist das, ob sie weit oder

215 Weit vom Zentrum weg liegen oder nahe beim Zentrum bei diesem mittleren Fixationspunkt

216 Und es gibt wirklich nur einen sehr kleinen Unterschied, und so, dass die, ja einen ganz kleinen Unterschied, okay, aber nicht wirklich ein Effekt

217 Hm?

218 Und ob sie gleich oder verschieden sind, ist auch nicht so wichtig, also eigentlich überhaupt nicht

219 Hm?

220 okay

221 okay, will ich das beibehalten? Ja, eigentlich schon, weil, ich denke, dass es schon mit dem etwas zu tun haben muss

222 okay, ich behalte es einmal bei

Transkript 11. Emelie interpretiert die Ergebnisse aus Zyklus 2.

Zyklus 2 auch aus dem Interaktionsplot keine Interaktion ersichtlich ist. Zum anderen ignoriert sie wieder die Inferenzstatistik, aus der hervorgeht, daß in diesem Experiment nur für den Faktor Exzentrizität ein bedeutsamer Haupteffekt vorliegt.

In den Schritten 209 und 210 wundert sie sich in Metaaussagen über ihre Befunde. Sie findet jedoch keine Erklärung für diesen Effekt und wiederholt lediglich die Ergebnisse.

Auch die Hypothese aus Zyklus 2 möchte Emelie beibehalten, obwohl ihre Interpretation der Ergebnisse dies eigentlich nicht zuließe, womit sie erneut einen Bestätigungsfehler begeht. Dennoch entspricht die Beibehaltung der Hypothese den Daten aus der Inferenzstatistik und wäre also korrekt – hätte Emelie diese einbezogen und korrekt interpretiert.

Diese Hypothesenbewertung kann also wieder nur unter Vorbehalt als richtig beibehalten beurteilt werden. Wieder unterläuft ihr ein doppelter Fehler, der schließlich zur korrekten Schlußfolgerung führt. Die Hypothese aus Zyklus 1 behält Emelie weiterhin bei, obwohl sie im zweiten Zyklus in Übereinstimmung mit der Inferenzstatistik aus dem Interaktionsplot herausgelesen hat, daß für den Faktor Farbähnlichkeit kein Effekt vorliegt. Hier unterläuft ihr also noch einmal ein Bestätigungsfehler.

Emelie ist bei der Hypothesenbewertung in einer ähnlichen Situation wie Ella in Transkript 2. Auch Ella hatte in ihren bisherigen Experimenten den Faktor Suchtyp auf Merkmalssuche konstant gehalten und sich darüber gewundert, keine Effekte zu finden. Während Ella jedoch nach einer Erklärung sucht und diese eben in der Konstanthaltung auf Merkmalssuche findet, wiederholt Emelie lediglich die Ergebnisse und kommt nicht zu einer Erklärung dieses Effekts. Ella zieht also mehr Nutzen aus ihrer Metabemerkung als Emelie.

Emelie: Zyklen 3 und 4

Vor Zyklus 3 zieht Emelie zunächst einmal Bilanz und kommt in Schritt 267 zu dem Schluß, daß sie noch nicht sehr weit gekommen sei (Transkript 12). Kurz erwägt sie, nochmals die Theorie nachzulesen, entscheidet sich dagegen und plant nun, im nächsten Experiment die beiden Faktoren zu untersuchen, die sie bisher noch nicht betrachtet hatte, nämlich Suchtyp und Größe der Anordnung. Eine explizite Begründung für die Auswahl dieser Faktoren bleibt Emelie schuldig.

267	Also, so viel weiter bin ich jetzt nicht mit diesen Resultaten, dann machen wir
268	Ich kann nur immer wieder neue Hypothesen machen und hoffen
269	Also, jetzt einmal überlegen
270	Sonst kann ich ja noch einmal die Theorie anschauen. Kann ich eigentlich
271	Nein, jetzt versuchen wir es einmal so
272	Also
273	Mh
274	Okay, was ich jetzt eigentlich immer ausgelassen habe, ist natürlich auch, em, conjunction search. Vielleicht finde ich dort mehr
275	Also, bis jetzt bin ich eigentlich immer auf die Farben gegangen, dass das einen Unterschied macht, Ähnlichkeit der Farben, Kombination der Farben
276	Jetzt mache ich einmal etwas anderes
277	Und zwar set size und search type

Transkript 12. Emelie plant ihr weiteres Vorgehen.

Auf diese Weise plant Emelie das sogenannte kritische Experiment zur Überprüfung der Merkmals-Integrations-Theorie, ohne diese zu erwähnen. So stellt sie auch erstmals eine ungerichtete Hypothese auf, obwohl sie aus der Instruktion Informationen hätte, die ihr eine genauere Vorhersage erlaubten. In der Hypothese ist auch noch davon die Rede, daß der Suchtyp von der Größe der Anordnung beeinflußt werde, womit eine Verwechslung von unabhängiger und abhängiger

321 Jetzt kann ich ja einmal alleine auf die conjunction-Bedingung eingehen
 322 jetzt könnte ich noch schauen, wie sich dort zusätzlich noch diese
 Farbenähnlichkeit auswirkt oder wie sich die Entfernung zum Zentrum
 auswirkt. Also ob es auch einen Einfluss hat, wenn es näher beim Zentrum
 ist, ob man es schneller erkennt
 323 ich glaube, das mache ich einmal

Transkript 13. Emelie plant ihre nächsten Experimente.

Variable nahegelegt wird. Sieht man von diesem Fehler ab, so läßt sich aus der Hypothese ablesen, daß Emelie eine Art Interaktion mit einer differentiellen Wirkung der beiden Stufen des Faktors Suchtyp erwartet.

Die Ergebnisse von Zyklus 3 läßt Emelie sich als deskriptive Statistik und als Interaktionsplot anzeigen. In diesem Fall entsprechen ihre qualitativen Interpretationen auch der Inferenzstatistik. Die Daten bestätigen Emelies Hypothese, und sie behält diese bei.

Nach Zyklus 3 plant Emelie ihre weiteren Experimente im voraus und möchte die anderen Faktoren auf der Stufe der Verbindungssuche untersuchen, nachdem sie den Faktor Suchtyp bisher auf der Stufe Merkmalssuche konstant gehalten hatte (Transkript 13).

Für Zyklus 4 stellt Emelie eine Hypothese zu den Faktoren Suchtyp, Größe der Anordnung und Exzentrizität auf. Wieder geht sie von einer Abhängigkeit der unabhängigen Variablen untereinander aus. Die Hypothese „die Verbindungssuche hängt von dem set-size und von der eccentricity ab.“ ist inhaltlich absurd und zudem noch unspezifischer als diejenige aus Zyklus 3. Bei der Versuchsplanung bemerkt Emelie dann, daß ein dreifaktorielles Design in *virtue* nicht möglich ist und gibt noch einmal denselben Versuchsplan wie in Zyklus 3 ein – ohne dies zunächst zu kommentieren. Erst bei der Interpretation der Ergebnisse bemerkt Emelie, daß sie ihr vorheriges Experiment repliziert hat (Transkript 14). Sie geht daher nicht weiter auf die Ergebnisse ein, vor allem aber bestätigt sie ihre Hypothese, ohne auf den dort erwähnten Faktor Exzentrizität einzugehen und kommt damit zu einer Fehleinschätzung.

Als nächstes nimmt Emelie eine Art Standortbestimmung vor. Sie schaut sich genau die Theorie an und rekapituliert, inwiefern sie diese bereits getestet hat (Transkript 15). Sie faßt ihre Erkenntnisse nochmals in eigene Worten und findet die Theorie durch ihre empirischen Daten bestätigt. Im Anschluß daran plant Eme-

362 Okay, jetzt habe ich einen sehr ähnlichen Effekt
 363 Da ist wieder die Reaktionszeit
 364 Und da habe ich wieder set size
 365 Mm, da habe ich doch jetzt ziemlich das Gleiche wie vorher, he?
 366 aha, okay, also jetzt habe ich eigentlich das Gleiche
 367 also somit denke ich eigentlich einmal, dass das, em, set size und, ist
 eigentlich eines der wichtigsten und search type sind zwei, die zwei, also
 zwei von den sehr wichtigen Merkmalen
 368 sicher beibehalten
 369 okay
 370 (schliesst das Resultatfenster, 4. Experiment)
 371 okay, jetzt mache ich noch einmal das Experiment drei
 372 ah, das ist ja schon so, das Experiment drei, ja
 373 okay, doch es ist ziemlich ähnlich

Transkript 14. Emelie vergleicht die Ergebnisse aus den Zyklen 3 und 4.

lie, in den nächsten experimentellen Zyklen die beiden Farbfaktoren nochmals genauer zu betrachten.

Emelie hat mit Zyklus 3 bereits alle Faktoren der visuellen Suche einmal experimentell überprüft. Damit hat sie den Experimenterraum äußerst effizient abgedeckt. Das kritische Experiment hat sie bereits im dritten Zyklus durchgeführt und dabei etwas über die Bedeutung des Faktors Suchtyp gelernt. Dies wirkt sich auf die Designs ihrer weiteren Experimente aus. Von diesem Punkt an hält sie den Suchtyp immer auf der Stufe Verbindungssuche konstant, wogegen sie vorher die Merkmalssuche favorisiert hatte.

Das kritische Experiment scheint Emelie fast aus Zufall in Zyklus 3 durchgeführt zu haben. Die beiden Faktoren Suchtyp und Größe der Anordnung hatte sie bisher noch nicht getestet. Beim Aufstellen der Hypothese für diesen Zyklus zieht sie kurzfristig in Erwägung, nochmals die Theorie durchzulesen, setzt dies aber

378 jetzt schauen wir die Theorie noch einmal an, vielleicht (öffnet Theorie)
 379 mm, also, das ist mir jetzt eigentlich nicht bestätigt worden, dass bei
 der Merkmalssuche, ah, doch, das ist mir ja bestätigt worden, doch
 380 bei der Merkmalssuche, da sind die Reaktionszeiten eigentlich gering,
 variieren gering, egal, wie viele Objekte dass es sind, ob es nur ein paar
 sind oder mehrere, im Gegensatz bei der Verbindungssuche, also conjunction
 search, da nimmt die Reaktionszeit, die man braucht, um den Zielreiz zu
 finden, mit zunehmenden Objekten zu
 381 das habe ich jetzt eigentlich herausgefunden
 382 aber
 383 das ist mir auch bestätigt worden
 384 aber jetzt das andere, inwiefern diese Farben einen Einfluss haben, das
 ist mir eigentlich kaum bestätigt worden

Transkript 15. Emelie rekapituliert, inwiefern sie die Theorie überprüft hat.

nicht in die Tat um. So ist eventuell auch die inhaltlich merkwürdige Hypothese zu erklären. Auch bei der Interpretation ihrer Ergebnisse diskutiert Emelie die Theorie dann nicht mehr.

Dies geschieht erst, nachdem Emelie unabsichtlich das kritische Experiment in Zyklus 4 repliziert hat. Hier liest sie sich noch einmal gründlich die Theorie durch und vergleicht sie mit ihren empirischen Ergebnissen. Dieser Schritt ist bemerkenswert, weil sonst bei keiner anderen Versuchsperson beobachtet werden konnte, daß die Theorie derart explizit in die Beurteilung der Ergebnisse einbezogen wurde.

Emelie: Zyklen 5 bis 7

Bei der Hypothese in Zyklus 5 setzt Emelie wieder korrekt die Reaktionszeit als abhängige Variable ein. Sie formuliert eine gerichtete Hypothese, macht jedoch keine Aussage über eine Interaktion. Warum sie den Faktor Farbähnlichkeit mit der Größe der Anordnung zusammen prüft, begründet Emelie nicht.

Bei der Versuchsplanung unterläuft Emelie dann eine Verwechslung der Faktoren Farbähnlichkeit und Farbkombination. Statt wie in der Hypothese formuliert die Farbähnlichkeit zu variieren, wählt sie die Farbkombination als unabhängige Variable für ihr Experiment. Ihren Fehler bemerkt sie bei der Interpretation der Ergebnisse in Schritt 419 (Transkript 16). Auch im weiteren Verlauf ist sie sich des Unterschiedes zwischen den beiden Farbfaktoren nicht immer ganz bewußt, wie in Schritt 427 deutlich wird.

Ihre Hypothese verwirft Emelie, ohne dies weiter zu begründen. In diesem Falle wäre eigentlich der Status „weiß nicht“ die korrekte Bewertung gewesen, denn schließlich konnte Emelie ihre Hypothese mit dem Experiment gar nicht überprüfen, da die Faktoren in Hypothese und Versuchsplan nicht übereinstimmten. Emelie macht sich die Ergebnisse ihres tatsächlich durchgeführten Experimentes auch nicht zunutze. Sie erwähnt den Faktor Farbkombination nicht mehr und geht anhand des Interaktionsplots wohl davon aus, daß dieser keinen bedeutsamen Effekt habe, obwohl in der Inferenzstatistik für beide der getesteten Faktoren signifikante Haupteffekte vorliegen.

411 okay, da hätte ich vielleicht einfach nehmen sollen, ob die Farben
unterschiedlich oder ähnlich sind
412 also
413 das ist ein geringer Unterschied eigentlich
[...]
419 mh, ich habe das Falsche genommen, ich hätte sollen, die Farben
unterscheiden sich stark oder nicht, oder sind sehr ähnlich, nehmen
420 okay
[...]
427 also kürzere Reaktionszeiten bei deutlich unterschiedlichen Farben

Transkript 16. Emelie verwechselt die Faktoren Farbähnlichkeit und Farbkombination.

Für Zyklus 6 stellt Emelie inhaltlich nochmals dieselbe Hypothese auf wie für den vorangegangenen Zyklus. Dieses Mal definiert sie auch den dazu passenden Versuchsplan. Bei der Interpretation schenkt sie den beiden Haupteffekten keine Beachtung, sondern konzentriert sich voll auf die Interaktion. Diese beschreibt sie völlig korrekt und sehr detailliert (Transkript 17). Welche Bedeutung in diesem Zusammenhang die Konstanthaltung des Faktors Suchtyp auf der Stufe der Verbindungssuche hat, diskutiert Emelie jedoch nicht. Ihre Erklärung gibt sie ab, als sei das Ergebnis allgemeingültig. Sie zieht keine Verbindung zu den Experimenten zu Beginn ihrer Experimentalreihe, als sie den Faktor Farbähnlichkeit bereits untersucht hatte und zu einer ambivalenten Bewertung hinsichtlich seiner Bedeutung für die visuelle Suche gekommen war.

462 also, ich sehe jetzt, dass am Anfang, bei recht wenigen Objekten, bei
sieben, sind die zwei fast gleich, da spielt es keine Rolle, ob, em, die
Farben ähnlich sind oder verschieden. wenn es aber mehr werden, bei
einunddreissig, gibt es doch einen Unterschied in der Reaktionszeit, und
zwar werden die, die deutlich unterschiedlich sind, durch Farben, und
durch unterschiedliche Farben gekennzeichnet sind, die werden schneller,
also der Testreiz wird dort schneller erkannt, also es gibt somit einen
Unterschied

Transkript 17. Emelie interpretiert die Interaktion zwischen Größe der Anordnung und Farbkombination.

Im letzten Zyklus stellt Emelie nochmals eine gerichtete, spezifische Hypothese zum Faktor Exzentrizität und den dazu passenden Versuchsplan auf. Sie fokussiert ihre Interpretation der Ergebnisse auf die Interaktion, ohne die ebenso vorhandenen Haupteffekte zu berücksichtigen. Sie kommt dennoch zum korrekten Urteil, daß ihre Hypothese beizubehalten ist.

Emelie: Zusammenfassung

Emelies Experimentalreihe läßt sich je nachdem, ob man den Umgang mit den Hypothesen oder den Aufbau der Experimentalreihe betrachtet, unterschiedlich bewerten. Die Interpretation der Ergebnisse im Hinblick auf die Hypothesen nimmt sie vorrangig anhand des Interaktionsplots vor. Die Inferenzstatistik läßt sie gänzlich außen vor. So unterlaufen ihr bei der Bewertung der Hypothesen etliche, teilweise sogar doppelte Fehler, so daß sie letzten Endes wieder zu einer inhaltlich korrekten Beurteilung kommt und diese wahrscheinlich auch als Wissen über den Gegenstandsbereich abspeichert.

Betrachtet man Emelies Suche im Experimenterraum, so wird eine Systematik deutlich, die sich aus ihren verbalen Äußerungen und der Suche im Hypothesenraum nicht ablesen läßt. In den ersten beiden Zyklen untersucht Emelie diejenigen Faktoren, über die die Theorie keine Aussagen macht. Dann testet sie die verbliebenen noch nicht untersuchten Faktoren und kommt damit zum kritischen Experiment, was an dieser Stelle fast schon wie ein Produkt des Zufalls wirkt. Nach drei Zyklen hat sie alle Faktoren bereits einmal überprüft. Im Anschluß daran ändert sie die Konstanthaltung des Faktors Suchtyp von der Stufe Merkmalssuche auf die Stufe Verbindungssuche und führt mit jedem der drei nicht in der Theorie erwähnten Faktoren ein Experiment in Kombination mit dem Faktor Größe der Anordnung durch. Emelie begründet diese Wahl an keiner Stelle. Dennoch ist Emelie mit dieser Versuchsplanung zu sehr aussagekräftigen Experimenten gekommen, hat dafür gesorgt, den Problemraum weiträumig zu durchsuchen, und sich somit gute Bedingungen geschaffen, Wissen über den Gegenstandsbereich zu erwerben.

Die Interpretation der Ergebnisse des kritischen Experiments nimmt Emelie zunächst nur anhand der empirischen Evidenz vor. Erst nachdem sie unabsichtlich das kritische Experiment in Zyklus 4 repliziert hat, schaut sie sich nochmals die Theorie an und vergleicht sie mit ihren empirischen Ergebnissen. Sie validiert die Theorie Schritt für Schritt anhand ihrer Ergebnisse. Dieses Vorgehen konnte bei keiner anderen Versuchsperson beobachtet werden und gab Emelie die Gelegenheit, ihr Wissen über die visuelle Suche zu vertiefen.

10.3 Dana und Delia: Wenn eine gut erklären kann...

Zu Beginn ihrer Arbeit mit *virtue* hat Delia die Kontrolle über Maus und Tastatur und ist damit vollständig für die Interaktion mit dem System zuständig. Dana sitzt links von ihr und kann durch ihre Interaktion mit Delia mitbestimmen, was diese eingibt. Der Ablauf der Experimentalreihe von Dana und Delia ist in Tabelle 17 dargestellt.

Gleich zu Beginn bringt Dana ihr Vorwissen aus dem Studium ein und fragt ihre Partnerin „Erinnert dich das nicht an den Dunbar?“ (Schritt 4). Offensichtlich sieht Dana Parallelen zu den Untersuchungen von Dunbar (1993) und Okada und Simon (1997), die beide aus der Vorlesung für Allgemeine Psychologie kennen, und zieht so eine regionale Analogie zum Experiment, an dem sie selbst teilnehmen (Dunbar, 1995, 1997, 2000a, 2001). Dana schlägt als nächstes vor, die Theorie nochmals anzuschauen, bevor Delia anstößt, eine Hypothese aufzustellen. Wenig später kommt sie nochmals auf das erwähnte Experiment zurück mit der Bemerkung „Also, jetzt ist das Experimentatoren versus Theoretiker, gell?“ (Schritt 44).

Dana und Delia: Zyklus 1

In ihrem ersten Zyklus konzentrieren sich Dana und Delia auf die beiden Farbfaktoren. Ihre erste Hypothese erarbeiten sie gemeinsam (Transkript 18). Ohne darüber zu diskutieren, mit welchen Faktoren sie ihre Untersuchung beginnen wollen, stellt Dana in Schritt 51 ihre Frage zur Farbähnlichkeit. Von Delia bekommt sie zunächst nur eine unspezifische Antwort und formuliert selbst ihre Vermutung zur Wirkung dieses Faktors. Delia steigt in Schritt 58 auf Danas Aussage ein, fällt ihr geradezu ins Wort. Was Dana in abstrakter Form gesagt hat, setzt Delia um in die konkrete Operationalisierung in *virtue*. Dann stellt Delia in Schritt 62 doch noch die Frage nach einer übergeordneten Planung der Versuchsreihe. Dana stimmt zu und schlägt wenig später vor, den Faktor Farbkombination gleich mitzutesten. Auf diese Weise bestimmt sie den Fokus von Zyklus 1. Im weiteren Verlauf versuchen Dana und Delia dann ab Schritt 77, ihre Hypothese für einzelne Farben zu spezifizieren, kommen jedoch nicht zu einer Vorhersage für den Faktor Farbkombination, sondern formulieren eine einfaktorielle Hypothese für den Faktor Farbähnlichkeit.

-
- 51 Dana em, also, was, was würdest du jetzt sagen bei similarity of colors?
 52 Delia dass das einen Einfluss hat?
 53 Dana mh
 54 Delia ja, das denke ich auf jeden Fall
 55 Dana also, wenn sie ähnlich sind
 56 Delia ja
 57 Dana dann hast du wahrscheinlich länger als [...] genau
 58 Delia (gleichzeitig) hast du länger. Du musst das zuerst einmal
 eigentlich erkennen und dann vielleicht gezielter schauen, als wenn
 59 Dana ja
 60 Delia du einfach so schnell schnell rot und grün hast, oder? [...] Findest du auch, denkst du das auch?
 61 Dana jaa
 62 Delia also, wollen wir dort anfangen, oder?
 63 Dana ja, also, wir behaupten einmal
 64 Delia ja
 65 Dana es geht länger, wenn die Farben ähnlich sind
 66 Delia mh (liest laut) „Formulieren Sie Ihre Hypothesen. Beschreiben Sie hier genau, welche Effekte Sie erwarten, etwa auch, ob sich mehrere Faktoren in ihrer Wirkung auf die abhängige Variable gegenseitig beeinflussen“
 67 Dana könnten wir nicht die Farben in einem durch nehmen, weißt du, em, da das zweite, color combination?
 68 Delia das zwei auch noch gleich? Kann man das?
 69 Dana ja, em, weißt du, die Farben in einem zusammennehmen
 70 Delia also, gehen wir back, oder? [...]
 77 Dana also ich würde sagen, meine Hypothese wäre, wenn die Farben ähnlich sind, dann hast du länger
 78 Delia mh
 79 Dana und das wäre dann pink/red
 80 Delia ja genau, ja
 81 Dana und blue/light blue
 82 Delia und light blue auch, ja
 83 Dana dann hast du länger
 84 Delia als wenn's blue/yellow oder pink/green
 85 Dana (gleichzeitig) als wenn's eindeutig ist, ja
-

Transkript 18. Dana und Delia erarbeiten ihre erste Hypothese.

Obwohl die beiden eine Hypothese zum Faktor Farbähnlichkeit aufgestellt haben, definieren sie ihren Versuchsplan zunächst als einfaktorielles Experiment nur mit dem Faktor Farbkombination. Dieser Fehler ist beiden Teilnehmerinnen nicht bewußt. Dana thematisiert aber ihre Schwierigkeiten mit den beiden Farbfaktoren in Schritt 129: „für mich ist es einfach irgendwie das gleiche“. Delia pflichtet ihr bei. Bei der Versuchsplanung hat Delia zunächst Schwierigkeiten mit der Definition der konstantzuhaltenden Faktoren. Sie möchte mehrere Stufen auswählen. So kommt es zu einer Diskussion über die Konstanthaltung allgemein und im speziellen über die Stufen, auf denen konstantgehalten werden soll (Transkript 19).

Tabelle 17

Experimentelle Zyklen von Dyade 1: Dana und Delia

Zyklus	Hypothese	Versuchsplan und Ergebnisse												
		Bewertung	Suchtyp	Farbähnlichkeit	Größe der Anordnung	Farbkombination	Exzentrizität	Text						
		Merkmalsverbindung	ähnlich	versch.	7 15 19 25 31	pink ...	gering mittel groß							
1	die ähnlichkeit der farben erhöht die reaktionszeit.	B	O	x	x	O	O	x	x					
2	die reaktionszeit verändert sich in abhängigkeit der farbkombination , aber nicht von suchtyp .	V	x	x	O	O	O	x	x					
3	merkmalssuche besser als verbindungssuche . beide geringere reaktionszeit mit unterschiedlichen farben .	B	x	x	x	O	O	O						
4	setsize wirkt sich nicht auf merkmalssuche aus, aber auf verbindungssuche .	B	X	X	O	X	X	X	O					
5	mit zunehmender ecc , nimmt die rz zu, aber unterschiedlich je nach farbähnlichkeit	V	O		x	x	O	O	O			x	x	x
		V	O		X	X	O	O	O			X	X	X
6	setsize hat einfluss aber eccentricity nicht	V	O		O	O	x	x	x	O	O	x	x	x
		V	O		O	O	X	X	X	O	O	X	X	X

Anmerkung. B = korrekte Beibehaltung der Hypothese bei bestätigenden Daten; b = falsche Beibehaltung der Hypothese bei widersprechenden Daten; V = korrekte Verwerfung der Hypothese bei widersprechenden Daten; v = falsche Verwerfung der Hypothese bei bestätigenden Daten. O = Konstanthaltung auf dieser Stufe; x = Variation eines Faktors auf dieser Stufe ohne signifikanten Effekt; x = signifikanter Haupteffekt; X = signifikanter Haupteffekt und signifikante Interaktion.

138	Delia	beide, oder?
139	Dana	jaja
140	Delia	so, mh, okay (Warnschild kommt) „one value for every constant factor“ okay
141	Dana	ah, dann, mh, ich würde sagen, ja dann müssen wir die anderen Werte konstant halten, dann machen wir alle mit der gleichen set size
142	Delia	alle andingsen?
143	Dana	ja nein, einfach eines davon
144	Delia	ja, das?
145	Dana	dass alle nachher das gleiche haben
146	Delia	oder einfach eines in der Mitte nehmen?
147	Dana	ja, das
148	Delia	das spielt eigentlich keine Rolle, he?
149	Dana	und dann, warte jetzt, was haben wir noch searchtype, nehmen wir den?
150	Delia	ich glaube, wollen wir feature? Merkmal ist das, Merkmalssuche
151	Dana	mh

Transkript 19. Dana erklärt Delia das Konzept der Konstanthaltung.

Delia ist noch nicht in der Lage, das Konzept der Konstanthaltung in *virtue* umzusetzen (Schritt 138). Dana ist im ersten Moment auch keine Hilfe (Schritt 139), erklärt aber dann am Beispiel des Faktors Suchtyp, was mit Konstanthaltung gemeint ist. Ob sie in Schritt 145 sogar eine Konstanthaltung über mehrere Experimente hinweg plant, ist unklar, wäre jedoch eine mögliche Interpretation ihres Satzes. Sie könnte sich aber auch lediglich auf das aktuelle Experiment und dessen Experimentalgruppen bezogen haben.

Bevor Delia den Versuchsplan endgültig verabschieden kann, zögert Dana bei der Frage, wie der Faktor Farbähnlichkeit konstantzuhalten sei, und schlägt vor, den Faktor Farbähnlichkeit gleich im ersten Experiment mitzutesten. Delia zeigt sich einverstanden. So kommen die beiden zu einem zweifaktoriellen Design, das auch die Hypothese zur Farbähnlichkeit abdeckt. Die ursprüngliche Verwechslung der beiden Faktoren bei der Versuchsplanung wird damit korrigiert, auch wenn die Teilnehmerinnen nicht explizit darüber sprechen. Dana berechnet schließlich noch, wieviele Versuchspersonenstunden ihr Experiment bei entsprechender Gruppengröße brauchen wird.

Für die Ergebnisse (und auch für alle späteren Experimente) wählen Dana und Delia gleich alle drei Darstellungsvarianten: deskriptive Statistik, Inferenzstatistik und Interaktionsplot. Delia erkennt sofort, daß der Effekt, den sie vorhergesagt hat, eingetreten ist, und äußert dies folgendermaßen: „ja, super, hei, genau das,

was vorhin gekommen ist, was ich gesagt habe, also irgendwie mit dem“ (Schritt 188). Dana faßt dies dann in den passenden statistischen Begriff: „also, ich würde sagen, wir haben einen Effekt“ (Schritt 191). Anhand der deskriptiven Daten versuchen beide Partnerinnen kurz, die Ergebnisse zu interpretieren, bevor Dana vorschlägt, doch den Interaktionsplot heranzuziehen. Zunächst wenden sie sich dem Faktor Farbähnlichkeit zu. Als dessen Effekt geklärt ist, initiiert Dana Überlegungen zum Faktor Farbkombination (Schritt 219, Transkript 20) und führt dessen statistischen Effekt wieder auf die Farbähnlichkeit zurück.

217	Dana	also, das heisst, es gibt Unterschiede, em, in der Ähnlichkeit der Farben, das entspricht etwa dem, was wir gesagt haben
218	Delia	mh
219	Dana	und, eh, signifikante Unterschiede in den Farbkombinationen, das würde wieder auf, auf das Ähnlichkeit
220	Delia	mh
221	Dana	zurückgehen oder? schätzungsweise
222	Delia	du, aber ich habe ein bisschen Mühe mit dem, warum ist da, p ist ja gar nirgends angegeben
223	Dana	nein, aber wenn es zwei, zwei Sternchen hat, dann ist es
224	Delia	dann heisst es, aha
225	Dana	auf 0.01 signifikant
226	Delia	das habe ich gar nicht gesehen, aha, ja dann ist es logisch signifikant. das hat mich nämlich irritiert vorher
227	Dana	ja, Sternchen ist immer gut
228	Delia	mh
229	Dana	dann ist es signifikant. Also, komm, wir schauen einmal die Ding an. pink/green oder pink/rot, he? also, pink, ah, das ist aber auch noch interesting. Wenn es ähnliche Farben sind, sind sie mit pink trotzdem schneller
230	Delia	mh, als mit
231	Dana	also mit pink und rot bist du schneller als mit blau und hellblau. Also ich nehme
232	Delia	(gleichzeitig) ja
233	Dana	ich nehme jetzt einmal an
234	Delia	ich nehme das ist weil, weil es irgendwie
235	Dana	dieser schwarze Punkt da
236	Delia	ist untendran
237	Dana	ist pink und rot
238	Delia	ja mit weiß ist similar
239	Dana	und der weiße ist blau und hellblau
240	Delia	und hellblau, genau
241	Dana	und dann fallen sie zusammen
242	Delia	fast, ja
243	Dana	ja, blau
244	Delia	also, pink und rot ist in dem Fall besser als, em, hellblau und blau
245	Dana	ja, also pink ist sowieso

- Fortsetzung -

Transkript 20. Dana und Delia interpretieren den Effekt der Farbkombination.

- Fortsetzung -

246 Delia also minim
 247 Dana besser
 248 Delia und da bei different fast keinen Unterschied, oder? ja, auch, eh
 249 Dana ja, ein Spürchen, ja. also rot ist einfach besser und unterschiedlich ist dann bei blau besser
 250 Delia ist das tiefer ein bisschen als da? ich sehe dort so schlecht
 251 Dana ja, ein Spürchen, ja es geht ganz
 252 Delia also, dann müssen wir jetzt von dem ausgehen oder, weiter, am besten
 253 Dana mh. also
 254 Delia (liest laut) „Hypothese beibehalten werden“ „weiß nicht“
 255 Dana „weiß nicht“ nein, ich würde sagen, „beibehalten“ oder?
 256 Delia mh
 257 Dana also, die Unterschiede sind jedenfalls nicht wahnsinnig
 258 Delia also „submit“ ja, jetzt müssen wir eine weitere aufstellen, oder?
 259 Dana mh
 260 Delia also „hypothesis“, „neue Hypothese“ [...] ja, jetzt würde ich gerne von dem ausgehen, was wir jetzt schon haben, weißt du, nicht irgendwie jetzt, irgendetwas anderes
 261 Dana ja, du könntest, weißt du, color combination, ob das mit den verschiedenen Suchtypen einen Einfluss hat

Transkript 20 (Fortsetzung). Dana und Delia interpretieren den Effekt der Farbkombination.

Hier greift Delia ein und fragt nach der statistischen Relevanz der Effekte. Obwohl davon bereits in der Diskussion um den Faktor Farbähnlichkeit die Rede war, hat sie die Darstellung noch nicht verstanden und sucht nach *p*-Werten (Schritt 222). Dana erklärt ihr, daß die Signifikanzen als Sternchen dargestellt sind.

Dana nimmt ihren Faden wieder auf und wendet sich erneut dem Faktor Farbkombination zu (Schritt 229). Sie erkennt sofort die Richtung des Effekts. Im Folgenden erarbeiten beide Partnerinnen nochmals gemeinsam am Interaktionsplot die genaue Bedeutung des Effekts. Sie arbeiten dabei so eng zusammen, daß sie sich sogar in ihren Sätzen gegenseitig ergänzen (Schritte 235 bis 237), bis Delia in Schritt 244 das Ergebnis auch für sich selbst formuliert und zusammenfaßt.

In Schritt 246 kommt Delia auf die Bedeutung des statistischen Effekts zu sprechen. Sie nennt ihn „minim“, Dana kurz darauf „ein Spürchen“. Bei der Beurteilung der Hypothese äußert sich Delia inhaltlich nicht und läßt Dana bestimmen, die korrekterweise auf Beibehaltung plädiert. Dana und Delia prüfen in ihrem ersten Zyklus die beiden Farbfaktoren, kommen aber in ihrer Hypothese nur zu

einer Vorhersage über den Einfluß der Farbähnlichkeit. Hier korrespondiert der Einzelfall deutlich mit den Ergebnissen der Gruppenstatistik, daß nämlich den Versuchspersonen eine Vorhersage über den Faktor Farbkombination schwerfällt.

Das anfänglich in diesem Zyklus geplante Experiment zum Faktor Farbkombination paßt nicht zur Hypothese über die Farbähnlichkeit. Trotz einfaktorieller Hypothese führen Dana und Delia dann ein zweifaktorielles Experiment durch, und nur diesem Umstand ist es zu verdanken, daß am Ende auch die Hypothese mit diesem Experiment überprüft werden konnte. Dieses Experiment bleibt bei Dana und Delia das einzige zum Faktor Farbkombination. Auch sie stellen keine Hypothese zu diesem Faktor auf, was mit den quantitativen Befunden korrespondiert.

Bei der Versuchsplanung berechnet Dana, wieviele Versuchspersonenstunden ihr Experiment bei entsprechender Gruppengröße brauchen wird. Dies ist außergewöhnlich. In den anderen Protokollen finden sich zwar Überlegungen dazu, sparsam mit den Versuchspersonenstunden umzugehen. Eine reale Berechnung findet bei anderen Versuchspersonen jedoch an keiner Stelle statt.

Delias Stil wird in Transkript 19 deutlich. Sie kommuniziert sehr auf Dana bezogen. Delia formuliert ihre jeweiligen Schritte meist in Form von Fragen und Vorschlägen, so daß Delia zwar das Thema vorgibt, Dana letzten Endes aber das Vorgehen bestimmt. Dana ist in der Rolle der Erklärenden. Sie bringt ihr Vorwissen äußerst kompetent ein, etwa wenn sie das Konzept der Konstanthaltung oder die Darstellung der Ergebnisse erläutert. Delia greift Danas abstrakte Erklärungen auf und setzt diese sogleich fallbezogen für die konkrete Aufgabe um. Sie wendet damit eine hervorragende Strategie zur Konsolidierung ihres neu erworbenen Wissens an (Bielaczyc, Pirolli & Brown, 1994).

Dana und Delia: Zyklus 2

Delia äußert dann einen Plan für das weitere Vorgehen (Schritt 260, Transkript 20). Sie möchte an ihre bisherigen Erkenntnisse anknüpfen. Der konkrete Vorschlag für ein Experiment mit den Faktoren Farbkombination und Suchtyp kommt dann im nächsten Schritt von Dana. Auf die Frage von Delia, wie der Faktor Suchtyp denn in die Hypothese für Zyklus 2 einzubringen sei, äußert Dana die Vermutung, daß dieser Faktor keinen Einfluß habe. Diese Vermutung fließt dann in die Hypothese für Zyklus 2 ein.

301 Dana wollen wir einmal nur Farbe nehmen?
 302 Delia mh. Nehmen wir das blue weg? oder
 303 Dana (gleichzeitig) ja
 304 Delia würdest du eher das blue lassen?
 305 Dana nein, ich würde es mit pink
 306 Delia mh. und sonst?
 307 Dana ja. ich glaube, set size müssen wir dann noch
 308 Delia das müssen wir in der nächsten machen, hm?
 309 Dana ja. ich habe Mühe mit dem similarity
 310 Delia komm, wir machen nur einmal different da
 311 Dana ja
 312 Delia mh. Mh. da musst du ja beide nehmen, oder?
 313 Dana nein, für jede variierte musst (?)

Transkript 21. Dana und Delia planen Experiment 2.

Bei der Versuchsplanung in Zyklus 2 definieren Dana und Delia zunächst ein ihrer Hypothese entsprechendes Experiment mit den Faktoren Farbkombination und Suchtyp. Die genaue Ausarbeitung des Versuchsplans wirkt dann etwas unkoordiniert (Transkript 21). Die zu testende Hypothese wird nicht erwähnt. Stattdessen schlägt Dana zunächst ein einfaktorielles Experiment zur Farbkombination vor (Schritt 301), dann zieht sie den Faktor Größe der Anordnung in Erwägung (Schritt 307). Keine dieser Ideen begründet sie inhaltlich, und so erhebt Delia Einspruch (Schritt 308) und verschiebt die Testung des Faktors Größe der Anordnung auf den nächsten Zyklus. Dana läßt sich darauf ein, so daß schließlich doch das ursprünglich geplante Experiment zur Durchführung kommt.

Bei der Diskussion der Ergebnisse von Zyklus 2 erkennt Dana gleich den Haupteffekt für den Faktor Suchtyp (Schritt 337, Transkript 22). Über den Faktor Farbkombination besteht bei beiden Teilnehmerinnen immer noch Unklarheit (Schritte 339 bis 342). Die Ergebnisse zu diesem Faktor besprechen die beiden gar nicht. In Schritt 343 versucht Dana, ein Fazit zu ziehen, und äußert, der Effekt für den Faktor Suchtyp sei bereits bekannt gewesen. Erst als sie sich die Hypothese im unteren Bereich des Ergebnisfensters noch einmal durchliest, erkennt sie, daß diese genau entgegengesetzt formuliert war, und wundert sich darüber (Schritt 345). Sie zieht daraus die klare Konsequenz, daß die Hypothese zu verwerfen sei.

Über den Faktor Farbkombination sprechen die Teilnehmerinnen an dieser Stelle nicht mehr. So stellen sie auch keinen Bezug zu den Ergebnissen aus Zyklus 1 her, in welchem sie einen Effekt für diesen Faktor vorgefunden hatten, der sich im aktuellen Zyklus nicht zeigte.

337	Dana	also search type unterscheiden sich ganz klar
338	Delia	mh
339	Dana	aber, ah, vielleicht hätten wir trotzdem müssen, weißt du, mit ähnlichen und nicht ähnlichen Farben, also irgendwie, mit diesem pink/green und pink/red kann ich nicht viel anfangen
340	Delia	(gleichzeitig) ich verstehe das auch nicht, warum kannst du das nicht einzeln machen?
341	Dana	das sollte man trennen können
342	Delia	gell! Warum kannst du das nicht trennen? Das ist doch wichtig. hm
343	Dana	also, was haben wir? Ja, ich meine, das haben wir schon gewusst
344	Delia	mh
345	Dana	da brauchst du alle Versuchspersonen. Wir haben. Was? so ein Seich. Haben wir geschrieben, „Suchtypus hat keinen Einfluss“?
346	Delia	stimmt ja, haben wir geschrieben
347	Dana	ja, nein, das
348	Delia	ah, ja, wir haben vor allem
349	Dana	jesses Gott
350	Delia	also, müssen wir eh zurück, oder? wir können sie ja, wir müssen es halt revidieren
351	Dana	ja, ich glaube, die müssen wir schon rev blr, revidieren. Warte, komm, dann müssen wir sie schnell verwerfen

Transkript 22. Dana und Delia verwerfen Hypothese 2.

Dana und Delia: Zyklus 3

Die Hypothese für Zyklus 3 stellt Dana fast im Alleingang auf. Sie beginnt mit Metaaussagen, in denen sie sich fragt, was sie bisher schon weiß, und schlägt vor, die Faktoren Suchtyp und Farbähnlichkeit zu kombinieren, ohne dies weiter zu begründen. Dana berücksichtigt die verschiedenen Faktoren der Reihe nach, und Delia stimmt jeweils zu (Transkript 23, Schritte 363, 367 und 369). Delia vollzieht die Bedeutung nach, während sie die Hypothese im Hypothesenscratchpad schriftlich festhält. Delia profitiert hier davon, daß sie die Eingabe in den Computer übernommen hat, und setzt sich auf diese Weise nochmals mit den Inhalten auseinander, die Dana vorgeschlagen hatte.

Die Versuchsplanung für Experiment 3 schaffen Dana und Delia in wenigen Schritten sehr effizient. Die Interpretation der Ergebnisse geht fast genauso schnell. Zunächst betrachten die beiden den Interaktionsplot. Dana zweifelt daran, ob einer der beiden Effekte auch signifikant ist, und schaut in der Inferenzstatistik nach. Ebenso erwähnt Dana erstmals eine Interaktion, überprüft anhand der Statistik, ob eine solche vorhanden ist, und ist zufrieden, daß sich keine Interaktion ergeben hat und ihre Hypothese somit vollkommen bestätigt wurde.

363	Dana	em, also ich würde sagen, für feature, also, sie sind, die Merkmalssuche sind sie schneller als da, Verbindungssuche
364	Delia	ja
365	Dana	und dann, sind sie wahrscheinlich bei beidem
366	Delia	mit pink
367	Dana	mit unterschiedlichen Farben schneller
368	Delia	ja, als
369	Dana	als mit gleichen
370	Delia	mh [...]
380	Delia	aber bei Dings wäre es beides Mal höher, bei pi, also bei unterschiedlichen Farben, als bei
381	Dana	ja also, tiefer,
382	Delia	(gleichzeitig) also dass
383	Dana	du hast kleinere Reaktionszeit
384	Delia	(gleichzeitig) ja. ja [...] ja, das meine ich, ja, also bessere Leistung, einfach, ja, tiefer von der Reaktionszeit, so sollte es dann aussehen, Interaktionsplot. Also, em (liest laut vor, was sie schreibt) „Merkmalssuche besser als Verbindungssuche“ em, „beide besser mit unterschiedlichen Farben“
385	Dana	ja
386	Delia	also, ja „geringere Reaktionszeit“ ja „mit unterschiedlichen Farben“ versteht man das?

Transkript 23. Dana entwickelt Hypothese 3 fast alleine.

In Zyklus 3 ist Dana fast ausschließlich in der Initiative. Sie erarbeitet die Hypothese und interpretiert die Ergebnisse ganz allein. Delia stimmt ihr bei jedem Schritt zu und hat so die Gelegenheit, Dana als Modell zu beobachten und daran zu lernen. Zusätzlich wird der Lernprozeß bei Delia gefördert, wenn sie Hypothesen, die Dana mündlich formuliert, schriftlich festhält und dabei nachfragt, wenn ihr etwas noch nicht einleuchtet. Die Suche im Problemraum ist eine systematische Fortsetzung der bisherigen Experimente. Die Auswahl der Faktoren erfolgte allerdings ohne Begründung, so daß offen bleibt, ob diese Reihenfolge von den Teilnehmerinnen so geplant war oder sich eher zufällig ergab.

Dana und Delia: Zyklus 4

Auch in Zyklus 4 bestimmt Dana den inhaltlichen Fokus. Wieder regt sie zuerst eine Bilanz des bisherigen Verlaufs an und leitet dies mit einer Metaaussage ein (Schritte 427 bis 433, Transkript 24). Während Delia vorschlägt, als nächstes die Wirkung der Exzentrizität zu überprüfen (Schritt 434), beharrt Dana darauf, nun den Faktor Größe der Anordnung anzugehen (Schritte 437 und 439). Sie erinnert sich an die Theorie und bemerkt, daß diese als Vermutung formuliert war (Schritte

427	Dana	also, was wissen wir jetzt?
428	Delia	also, wir wissen, dass
429	Dana	das mit der Farbe haben wir
430	Delia	das haben wir schon lange
431	Dana	ja
432	Delia	und das mit der Merkmalssuche, das haben wir eigentlich auch schon vorher gewusst
433	Dana	ja, wollen wir einmal, eh bei, also, die Farbe haben wir, glaube ich, durchgetestet, mehr oder weniger
434	Delia	em, ja, close. Wollen wir einmal, em, eccentricity?
435	Dana	ja, ich würde eher noch
436	Delia	set?
437	Dana	(gleichzeitig) ja [...] ja set haben wir auch noch nicht
438	Delia	also, wo muss ich jetzt wieder hinein? „neue Hypothese“
439	Dana	also variieren wir set, und auch, ich würde auch nach search type gerade (?)
440	Delia	immer eins wieder reinnehmen und dann, ja. also
441	Dana	also, das wäre jetzt, bö, das ist an sich auch schon gestanden
442	Delia	ja, das ist wirklich schon gestanden. Ja, vielleicht ist es ja gut, wenn wir
443	Dana	(gleichzeitig) es ist ja, glaube ich, nur eine Vermutung gewesen
444	Delia	mh
445	Dana	also, dann vermuten wir einfach einmal das gleiche

Transkript 24. Dana und Delia stellen eine Hypothese zum kritischen Experiment auf.

441 und 443). Der Einfachheit halber schließt sie sich dieser Vermutung an. Delia ist mit der Navigation von *virtue* beschäftigt und stimmt Danas Vorschlägen ansonsten einfach zu.

Dana formuliert eine Hypothese, in der sie auch die Interaktion der beiden Faktoren Suchtyp und Größe der Anordnung vorhersagt. Im Transkript findet sich auch eine Bemerkung, daß sie mit ansteigender Größe der Anordnung für die Verbindungssuche eine steigende Reaktionszeit erwartet. Dies findet jedoch in die schriftliche Fassung der Hypothese keinen Eingang. Die Versuchsplanung ist ebenso rasch erledigt wie im vorangegangenen Zyklus.

Das Ergebnis von Experiment 4 fällt entspricht den Erwartungen der Teilnehmerinnen. Allerdings fällt Dana auf, daß bei der Merkmalssuche die Reaktionszeit bei zunehmender Größe der Anordnung sogar leicht abfällt (Schritt 487, Transkript 25). Delia sucht den Grund dafür zunächst in der Darstellung (Schritt 488), geht aber darauf ein, als Dana auf ihrer Beobachtung insistiert, und sucht nach einer Erklärung (Schritte 492 und 494). Dana geht noch weiter und möchte zunächst einen statistischen Beleg für die Relevanz ihrer Beobachtung suchen. Es ist ihr jedoch klar, daß *virtue* nicht für solche weitergehenden statistischen Berech-

487	Dana	ja. em. Ja, das fände ich noch toll. Geht das abwärts? Ich glaube schon
488	Delia	gell, das sieht so aus, als würde es abwärts gehen. Ja, wobei, es ist nur wegen dem blöden, weil der Strich irgendwie
489	Dana	ja, aber sonst wäre er ja, sonst hätte er ja keine Eckchen drin,
490	Delia	ja, stimmt, dann es ja
491	Dana	wenn er ganz gerade wäre
492	Delia	also, also das heisst, bei Merkmalssuche, hee, ist das logisch? Mit zunehmender
493	Dana	braucht es sogar weniger Zeit
494	Delia	ja. das finde ich komisch, du musst ja trotzdem zuerst alles absuchen. Oder, oder vielleicht, irgend in der Menge drin siehst du es schneller
495	Dana	ja, jetzt müsste man können irgendwelche Kontrast rechnen, aber das kannst du ja nicht mit dem
496	Delia	das finde ich jetzt komisch
497	Dana	aber eben, sie sind auf jeden Fall immer schneller
498	Delia	mh, das ist
499	Dana	und also die Veränderung kann nicht gross sein
500	Delia	nää
501	Dana	und da ist sie doch schon
502	Delia	ja, dann lassen wir es halt, oder?

Transkript 25. Dana und Delia entdecken abfallende Reaktionszeiten bei der Merkmalssuche.

nungen ausgelegt ist (Schritt 495). Den Bezug zur Hypothese spricht keine der beiden an. Wenn die Beobachtung abnehmender Reaktionszeiten unter der Merkmalssuche relevant sein sollte, dann wäre die aufgestellte Hypothese zu verwerfen und daran haben die Teilnehmerinnen kein Interesse. Dana vermutet stattdessen, daß der Effekt nicht groß sein könne (Schritt 499), und Delia schlägt vor, die Diskussion dabei zu belassen (Schritt 502), was dann auch geschieht.

In Zyklus 4 findet auf Danas Initiative hin das kritische Experiment statt. Die Merkmals-Integrationstheorie diskutieren die Teilnehmerinnen dabei nicht mehr. Stattdessen übernehmen sie direkt die Aussagen der Theorie in ihre Hypothese. Bei der Betrachtung des Interaktionsplots findet Dana eine unerwartete Tendenz, deren statistische Bedeutsamkeit sich in *virtue* nicht bestimmen läßt. Dana kann ihr statistisches Wissen zwar äußern, aber nicht anwenden. Beide stellen keinen Bezug dieses vermuteten Effekts zu ihrer Hypothese her, sondern argumentieren, daß der Effekt wahrscheinlich vernachlässigbar sei. Sie bestätigen ihre Hypothese, wie das den Ergebnissen zufolge auch korrekt ist. Der Problemraum ist mit diesem Experiment bis auf den Faktor Exzentrizität abgedeckt. Auf diesen Faktor konzentrieren sich die Teilnehmerinnen als nächstes.

575	Dana	wahrscheinlich, weißt du was
576	Delia	(gleichzeitig) we, das kann aber nicht sein, dass
577	Dana	wir könnten das, wir könnten das zweimal machen
578	Delia	mh
579	Dana	und eh, nehmen wir eins feature und das nächste Mal conjunction [...]
584	Delia	also, nachher machen wir conjunction (Designfenster geschlossen) eh „experiment“ [...] ou [...] uii, ou, hee, das ist aber komisch. Das heisst, es hat gar keinen Einfluss, wie weit weg dass es vom Zentrum ist
585	Dana	ja, aber, das ist jetzt, eben, das ist, ich meine eben, das ist auch eine Vermutung, em
586	Delia	und, und, die Dings hat auch keinen Einfluss
587	Dana	also da hat gar nichts einen Einfluss
588	Delia	mn. Also wollen wir es jetzt mit
589	Dana	ja, ich meine, wenn, wenn die Grösse, weißt du, die Anzahl Distraktoren ja bei feature auch keinen Einfluss hat
590	Delia	stimmt
591	Dana	dann macht es ja Sinn
592	Delia	ja
593	Dana	dass der Abstand auch keinen Einfluss hat. Aber, irgendwie denke ich, beim anderen würde es dann Einfluss haben, bei der Verbindungssuche
594	Delia	also
595	Dana	was müssen wir jetzt? jetzt müssen wir da, glaube ich, „weiß nicht“, sonst können wir es nachher nicht mehr testen

Transkript 26. Dana schlägt ein pseudo-dreifaktorielles Experiment vor und überlegt eine Interpretationsmöglichkeit für nicht gefundene Effekte.

Dana und Delia: Zyklus 5

In Zyklus 5 prüfen Dana und Delia die Kombination der Faktoren Farbähnlichkeit und Exzentrizität. Ihre Hypothese formulieren sie für die Exzentrizität gerichtet, für die Farbähnlichkeit jedoch, trotz der Erkenntnisse aus den Zyklen 1 und 3, ungerichtet. Nach Vollendung der Hypothese schlägt Dana vor, gleich noch mehr Hypothesen aufzustellen, worauf Delia aber nicht eingeht.

Bei der Versuchsplanung schlägt Dana vor, doch den Faktor Suchtyp in das Design einzubeziehen. Da dreifaktorielle Experimente in *virtue* nicht möglich sind, plant Dana, in einem ersten Experiment den Faktor Suchtyp auf der Stufe Merkmalssuche und in einem Folgeexperiment auf der Stufe Verbindungssuche konstant zu halten (Schritte 575, 577 und 579, Transkript 26). Delia stimmt dem Vorschlag zu (Schritt 584).

Die Ergebnisse von Experiment 5 sind so klar, daß deren Interpretation schnell erledigt ist. Dana überlegt sich allerdings eine mögliche Erklärung dafür, daß Exzentrizität keinen statistisch bedeutsamen Effekt hat (Schritt 589). Sie findet

diese in einer lokalen Analogie: Wenn die Größe der Anordnung bei der Merkmalsuche keinen Einfluß hat, dann wäre es doch folgerichtig, daß dies auch für die Exzentrizität gilt. In Schritt 593 stellt sie dann eine Hypothese für das kommende Experiment mit der Verbindungssuche auf, nämlich daß Exzentrizität unter der Verbindungssuche sehr wohl Auswirkungen haben sollte. Diese Hypothese wird zwar nicht schriftlich festgehalten, ist aber eindeutig formuliert. Um sich die Interpretation der bereits aufgestellten Hypothese offenzuhalten, fordert Dana, diese Hypothese vorübergehend mit „weiß nicht“ einzustufen.

Für Experiment 6 wird keine neue Hypothese aufgestellt. Die Versuchsplanung geht schnell. Alles bleibt gleich wie in Experiment 5, nur für den Faktor Suchtyp wählen die Teilnehmerinnen statt Merkmalssuche nun Verbindungssuche aus. Das Ergebnis ist überraschend. Beide Haupteffekte und die Interaktion sind zwar signifikant, entsprechen aber nicht den Erwartungen aus der schriftlichen Hypothese und auch nicht dem, was Dana in Schritt 593 zur Verbindungssuche gesagt hatte. Dies ersehen Dana und Delia aus dem Interaktionsplot. Die mittlere Exzentrizität verlangsamt die visuelle Suche bei hoher Farbähnlichkeit am meisten. Unter allen anderen Bedingungskombinationen verläuft die visuelle Suche beinahe gleich schnell. Beide können sich diese Effekte nicht erklären. Bei der Interpretation der Hypothese nehmen sie keinen Bezug mehr auf Experiment 5, kommen aber aufgrund der qualitativen Tendenzen der Effekte in Experiment 6 zu der korrekten Schlußfolgerung, daß sie zu verwerfen ist.

Auch in Zyklus 5 bestimmt Dana weitgehend das Vorgehen. Auf der Suche im Problemraum wird nun der letzte bisher noch nicht untersuchte Faktor einbezogen. Nur kurz kommt von Dana der Vorschlag auf, die experimentelle Strategie zu ändern und mehrere Hypothesen hintereinander aufzustellen. Delia nimmt ihn nicht an, und Dana verfolgt ihn auch nicht weiter.

Ein anderer Vorschlag von Dana, nämlich ein pseudo-dreifaktorielles Experiment durchzuführen, ist erfolgreich. Bei der Interpretation der Hypothese wählen Dana und Delia zunächst die Möglichkeit „weiß nicht“, weil sie sich noch keine endgültige Aussage erlauben. Die Interpretation der Ergebnisse von Experiment 6 stellt beide vor eine besondere Herausforderung: Obwohl sich signifikante Effekte ergeben, entsprechen diese nicht den Erwartungen aus der Hypothese. Durch die

817	Delia	„beibehalten“
818	Dana	ah, stimmt jetzt das?
819	Delia	mit dem, was wir gesagt, da steht
820	Dana	(liest) „set size“
821	Delia	(liest) „hat Einfluss“ auf Exzentrizität
822	Dana	„hat Ein“
823	Delia	hö, was ist das?
824	Dana	Exzentrizität auch noch
825	Delia	aber [...] stimmt
826	Dana	ja, set size hat, aber [...] also, set size hat einen Einfluss. Ja, und dann, jaja, machen wir „weiß nicht“ oder machen wir, sind wir ehrlich und verwerfen sie? Weil sonst stimmt es nicht
827	Delia	also, komm, verwerfen wir sie. Wir wollen immer „weiß nicht“
828	Dana	okay (schliesst das Resultatfenster)

Transkript 27. Dana und Delia verwerfen Hypothese 6.

gleichwertige Betrachtung qualitativer und quantitativer Aspekte der Ergebnisse kommen Dana und Delia zur korrekten Bewertung der Hypothese, nämlich daß sie zu verwerfen sei.

Dana und Delia: Zyklus 6

Vor dem Beginn von Zyklus 6 verläßt Delia kurz den Versuchsraum. Bei ihrer Rückkunft hat Dana Tastatur und Maus übernommen, so daß sie nun die Navigation und Eingabe in *virtue* bestimmt. Die zu untersuchenden Faktoren für diesen Zyklus legt Dana schnell fest und formuliert die Hypothese ohne weitergehende Diskussion mit Delia. Auch die Versuchsplanung nimmt Dana rasch, jedoch immer in Absprache mit Delia vor. In den Ergebnissen von Experiment 7 finden sich keine signifikanten Effekte. Wieder schlägt Dana vor, die Hypothese vorläufig mit „weiß nicht“ zu bewerten.

Die Initiative, dasselbe Experiment mit Konstanthaltung auf der Stufe Verbindungssuche zu versuchen, kommt dieses Mal von Delia. Eine neue Hypothese ist dazu nicht notwendig, und der Versuchsplan für Experiment 8 ist damit auch schon festgelegt. Die Ergebnisse dieses Experiments bringen nun doch einen Effekt für Exzentrizität ebenso wie für Größe der Anordnung. Die Diskussion der Ergebnisse ist dann jedoch etwas undurchsichtig. Zunächst schlägt Delia vor, die Hypothese beizubehalten (Schritt 817, Transkript 27). Dana stellt das in Frage (Schritt 818), und so lesen beide nochmals den Text ihrer Hypothese. Da nun doch ein Effekt von Exzentrizität aus den Daten zu ersehen ist, zieht Dana den logisch korrekten Schluß, daß die Hypothese zu verwerfen ist. Nach Zyklus 6 ziehen Dana

und Delia eine Bilanz. Sie lesen nochmals in der Theorie nach und schauen sich die Hypothesen an. Dann rekapitulieren sie die Ergebnisse zu den einzelnen Faktoren und beschließen, daß sie ihre Experimentalreihe beenden wollen.

Dana und Delia: Zusammenfassung

Dana und Delia führen ihr erstes Experiment zum Faktor Suchtyp bereits in Zyklus 2 durch und bemerken dessen Einfluß. Auf diese Weise bestimmt die Bedeutung dieses Faktors die ganze weitere Experimentalreihe. In jedem nachfolgenden Experiment wird dieser Faktor berücksichtigt, indem er entweder variiert wird oder in Form pseudo-dreifaktorieller Experimente mit einbezogen wird.

Jeweils vor den Hypothesen der Zyklen 2, 3, 4 und 5, also in der Mehrzahl der Fälle, ziehen Dana und Delia in Metaaussagen Bilanz darüber, was sie bisher schon wissen und was bereits experimentell untersucht wurde. Dies trägt dazu bei, daß sie ihr Wissen festigen und vertiefen. Zudem bewahren sie auf diese Weise einen guten Überblick und können ihre Experimentalreihe effizient gestalten.

Bei der Versuchsplanung berechnet Dana, wieviele Versuchspersonenstunden ihr Experiment bei entsprechender Gruppengröße brauchen wird. Dies ist außergewöhnlich. In den anderen Protokollen finden sich zwar Überlegungen dazu, sparsam mit den Versuchspersonenstunden umzugehen. Eine reale Berechnung findet bei anderen Versuchspersonen jedoch an keiner Stelle statt.

Delias Stil wird in Transkript 19 deutlich. Sie kommuniziert sehr auf Dana bezogen. Delia formuliert ihre jeweiligen Schritte meist in Form von Fragen und Vorschlägen, so daß Delia zwar das Thema vorgibt, Dana letzten Endes aber das Vorgehen bestimmt. Dana ist in der Rolle der Erklärenden. Sie bringt ihr Vorwissen äußerst kompetent ein, etwa wenn sie das Konzept der Konstanthaltung oder die Darstellung der Ergebnisse erläutert. Delia greift Danas abstrakte Erklärungen auf und setzt diese sogleich fallbezogen für die konkrete Aufgabe um. Sie wendet damit eine hervorragende Strategie zur Konsolidierung ihres neu erworbenen Wissens an (Bielaczyc, Pirolli & Brown, 1994).

10.4 Dido und Dora: Gemeinsam erarbeitetes Wissen

Dido und Dora teilen sich die Arbeit am Bildschirm auf. Während Dora rechts sitzt und die Maus navigiert, übernimmt Dido links die Eingabe über die Tastatur. Der Ablauf ihrer Experimente ist in Tabelle 18 zusammengefaßt.

Noch bevor Dido und Dora ihr erstes Experiment beginnen, kommen sie beim Durchlesen der Instruktion wie Dyade 1 auf das Experiment von Dunbar (1993) und Okada und Simon (1997) zu sprechen (Transkript 28, Schritte 8 bis 10). Auch in dieser Dyade wird die regionale Analogie vom Experiment, an dem die beiden teilnehmen, zu den genannten Untersuchungen gezogen. Dora erinnert sich zusätzlich an die Ergebnisse zur Überlegenheit von Dyaden und hofft bereits auf Erfolg allein aufgrund der Zuordnung zur Experimentalgruppe der Dyaden. Ebenfalls noch vor Beginn ihrer Experimentalreihe findet in den Schritten 15 bis 18 die erste Episode eines rein sozialen, nicht aufgabenbezogenen Dialogs statt. Schließlich lesen sie nochmals die Theorie durch und wenden sich dann ihrer ersten Hypothese zu.

8	Dido	ja. das ist wie das Experiment von dem, Dunbar, weißt du, mit dem
9	Dora	ja genau, das erinnert mich schon sehr daran, ja (lesen weiter)
10	Dora	das ist ein bisschen so wie das, als Dyade, sind wir schon super erfolgreich [...]
15	Dido	probieren geht über studieren und noch ein Snickers (nimmt sich einen Schokoriegel)
16	Dora	ich brauche auch noch eines
17	Dido	das ist so gemein gewesen, (?) Snickers (?) Cola, weil mir schlecht war
18	Dora	ja, zum Glück hast du Cola light gewählt, sonst hätte ich einen Ballonbauch bekommen

Transkript 28. Dido und Dora diskutieren ihr Vorwissen und unterhalten sich nicht aufgabenbezogen.

Dido und Dora: Zyklus 1

Zunächst diskutieren Dido und Dora ihre Vorgehensweise (Transkript 29). Dora beginnt in Schritt 39, indem sie, ohne einen inhaltlichen Zusammenhang zu thematisieren, fragt, welche Faktoren für die Hypothese herangezogen werden sollen. Mit „oben“ und „unten“ bezieht sie sich auf die Auflistung der Faktoren im Hypothesenscratchpad. Dido wechselt die inhaltliche Ebene und stellt die prinzipielle

Zyklus	Hypothese	Versuchsplan und Ergebnisse										
		Text	Bewertung	Suchtyp	Farbähnlichkeit	Größe der Anordnung			Farbkombination		Exzentrizität	
						7	13	19	25	31		pink...
Merkmals.	Verbind.s.	ähnlich	versch.	ähnlich	ähnlich	ähnlich	ähnlich	ähnlich	ähnlich	ähnlich	ähnlich	
1	Die Anzahl der Elemente auf dem Display hat bei einer Merkmalsuche mit unterschiedlich farbigen Objekten keinen Einfluss auf die Reaktionszeiten. Bei sehr unterschiedlichen Farben sind die RZ etwas geringer gegenüber ähnlichen Farben .	b	O	x	x	x	x	x	x	O	O	O
2	Die in Hypothese 1 formulierten Erwartungen bleiben auch bestehen, wenn andere Farben verwendet werden (blau statt Pink, Gelb statt Grün und Hellblau statt red) .	V	O	X	X	X	X	X	O	O	O	O
3	iiiiii		O	O	O	O	O	x	x	*	O	O
4	Verschiedene Farbkombinationen führen zu unterschiedlichen RZ, bei ähnlichen Farben erhöht sich die RZ gegenüber unterschiedlichen Farben .	w	O	X	X	O	O	X	X	O	O	O
5	Mit steigender Anzahl der Elemente auf dem Display steigt auch die RZ.	B	O	O	O	x	x	x	x	O	O	O
6	Die RZ erhöht sich mit der größer werdenden Distanz von der Displaymitte .	B	O	O	O	O	O	O	O	O	x	x
7	Je grösser die set size und die eccentricity , desto höher die RZ.	B	O	O	O	X	X	X	X	O	X	X
8	Bei der Merkmalsuche sind die RZ kürzer als bei der Verbindungssuche , da die Suche parallel und unbewusst abläuft im Ggs zur seriellen Verbindungssuche.	B	x	x	O	O	O	O	O	O	O	O
9	Bei der Verbindungssuche sind die RZ sowohl bei ähnlichen als auch bei unähnlichen Farben länger als bei der merkmalsuche . Die RZ bei der Verbindungssuche mit ähnlichen Farben sind länger als bei unähnlichen Farben .	B	X	X	X	X	O	O	O	O	O	O

Anmerkung. B = korrekte Beibehaltung der Hypothese bei bestätigenden Daten; b = falsche Beibehaltung der Hypothese bei widersprechenden Daten; V = korrekte Verwerfung der Hypothese bei widersprechenden Daten; v = falsche Verwerfung der Hypothese bei bestätigenden Daten; w = weiß nicht. O = Konstanthaltung auf dieser Stufe; x = Variation eines Faktors auf dieser Stufe ohne signifikanten Effekt; x = signifikanter Haupteffekt; X = signifikanter Haupteffekt; X = signifikanter Haupteffekt und signifikante Interaktion. * Dieser Versuchsplan wurde nicht in einem Experiment getestet.

39	Dora	wir können von oben anfangen, oder von unten, oder wir können auch die Mitte wählen
40	Dido	was sollen wir uns denn einmal aussuchen, also, ich meine, wir können ja nicht über alles Aussagen machen
41	Dora	nein, wir können etwas Einfaches nehmen, oder?
42	Dido	ja
43	Dora	nicht, was meinst du? Also nicht alles auf's Mal, sonst können wir das vielleicht dann nicht so, oder, ich weiß nicht
44	Dido	nein, wir müssen, wir müssen das ja untersuchen können
45	Dora	ja
46	Dido	nicht alles auf einmal machen, em, ja, fangen wir einmal oben an

Transkript 29. Dido und Dora diskutieren die Komplexität ihrer ersten Hypothese.

Frage, wieviele Faktoren gleichzeitig in einer Hypothese behandelt werden sollen. Sie schlägt vor, oben in der Liste der Faktoren anzufangen und zuerst eine Vorhersage über den Faktor Farbähnlichkeit zu treffen, womit Dora jeweils einverstanden ist.

Dido und Dora geraten für die nächsten 120 Schritte in einen intensiven Dialog und überlegen, ob sie auch die Faktoren Farbkombination und Suchtyp einbeziehen sollen, einigen sich am Ende aber auf eine Hypothese zu den Faktoren Farbähnlichkeit und Größe der Anordnung. Der Dialog ist wenig kohärent, wie ein beispielhafter Ausschnitt der Schritte 82 bis 91 demonstriert (Transkript 30). Innerhalb dieses Abschnitts passiert es dreimal, daß die beiden gleichzeitig sprechen. Dadurch ist es ihnen fast unmöglich, sich in den gleichzeitig geäußerten Schritten inhaltlich aufeinander zu beziehen. Für die qualitative Analyse lassen sich in dieser Episode keine prägnanten Abschnitte herausfiltern. Insgesamt zeigt Dido mehr Initiative im Dialog, d.h. sie bestimmt dessen inhaltlichen Verlauf. Dora ist zwar

82	Dido	ja, vielleicht sollten wir einfach einmal etwa von der Effekt, von der set size ausgehen, ich glaube, das ist einfacher
83	Dora	(gleichzeitig) set size und similarity, oder?
84	Dido	eh, ja
85	Dora	also weil sie müssen sich ja in etwas unterscheiden, dieses Zielobjekt vom
86	Dido	(gleichzeitig) ja [...] ja
87	Dora	ja, ah
88	Dido	ach nö
89	Dora	gut, dann kann man das jetzt einfach so auf deutsch schreiben, oder nicht?
90	Dido	(gleichzeitig) kannst du schnell tippen? Nicht?
91	Dora	ja

Transkript 30. Dido und Dora bestimmen die Faktoren ihrer ersten Hypothese.

daran beteiligt, die Hypothese wörtlich auszuarbeiten, Dido entscheidet aber jeweils als diejenige, die an der Tastatur sitzt, wie die Hypothese letzten Endes aufgeschrieben wird.

Trotz der geringen Kohärenz des Dialogs gelangen Dido und Dora am Ende dieser Episode zu einer sehr differenzierten Hypothese für Zyklus 1 (Tabelle 18), in der sie für zwei Faktoren eine getrennte Vorhersage treffen und sogar festhalten, auf welcher Stufe der Faktor Suchtyp konstant zu halten sei. Damit haben sie entgegen ihrem Vorsatz, zu Beginn „was Einfaches“ zu nehmen (Schritt 41, Transkript 29), eine komplexe Hypothese formuliert. Nachdem sie die Hypothese für Zyklus 1 abgeschlossen haben, wägen Dido und Dora ihr weiteres Vorgehen ab (Transkript 31). Dora fragt nach der Überprüfung der Hypothese und stellt dann eine weitere Frage unklaren Inhalts, die dann von Dido in Schritt 174 aufgegriffen wird und schließlich von Dora mit Inhalt gefüllt wird. Hier handelt es sich um eine echte Ko-Konstruktion. Die beiden Partnerinnen führen eine Metadiskussion über ihr weiteres Vorgehen, ob sie zunächst mehrere Hypothesen aufstellen und diese anschließend überprüfen sollen oder ob sie ihre aktuelle Hypothese direkt testen sollen. Sie finden Argumente für eine sofortige Hypothesentestung und entschließen sich, zunächst ein Experiment zu planen. Ihr weiteres Vorgehen wollen die beiden dann von den Ergebnissen dieses Experiments abhängig machen.

Die Versuchsplanung erledigen Dido und Dora schnell und ohne weitere inhaltliche Diskussion. Sie setzen einen Versuchsplan um, der sie genau testen lässt, was sie in ihrer Hypothese vorhergesagt haben. Die Bestimmung der Stichproben-

173	Dora	super, okay. Und jetzt müssten wir das prüfen, oder? oder meinst du nächstes Mal
174	Dido	oder wir machen jetzt die nächste, das nächste, eh
175	Dora	Hypothese
176	Dido	also entweder Hypothese
177	Dora	ja
178	Dido	Hypothesen höhö, entweder wir überprüfen das jetzt, was wahrscheinlich am übersichtlichsten ist, bevor wir dann zwanzig Hypothesen haben, die wir alle gar nicht mehr überprüfen können
179	Dora	(gleichzeitig) genau [...] und wir können ja aufgrund von dem, was jetzt gezeigt wird, als Resultat dann auch unsere nächste Hypothese formulieren
180	Dido	ja
181	Dora	mh, dann machen wir close window (schliesst das Hypothesenfenster, Hypothese 1)

Transkript 31. Dido und Dora erwägen, mehrere Hypothesen nacheinander aufzustellen.

229	Dora	Stichprobengrösse
230	Dido	was heisst denn fünfzig Stunden?
231	Dora	also ich nehme an
232	Dido	(gleichzeitig) also heisst das, dass wir insgesamt fünfzig Versuchspersonen zur Verfügung haben oder?
233	Dora	also ich nehme an, dass wir wahnsinnig viele zur Verfügung haben, weil wir hier so wahnsinnig viel Zeit verbringen können, oder?
234	Dido	ja, wie viel Uhr ist es jetzt? mh
235	Dora	was meinst du?
236	Dido	ja
237	Dora	also, ich glaube nicht, dass wir uns darauf gross achten müssen
238	Dido	ja
239	Dora	hätte ich jetzt einmal gedacht, sonst sind wir auch zu perfekt
240	Dido	ja, (?) wir müssen ja mit unserer Genialität nicht gleich hausieren gehen
241	Dora	genau, genau, und Stichprobengrösse dreissig ist nicht so schlecht, wir haben
242	Dido	na?
243	Dora	dreissig in jedem Dings
244	Dido	ja
245	Dora	das ist wahnsinnig viel, das ist, das ist super, das haben wir nie gehabt im Projektseminar

Transkript 32. Dido und Dora wägen die Stichprobengröße für ihr erstes Experiment ab.

größe für das Experiment gibt den beiden Anlaß zum Nachdenken (Transkript 32). Dido wirft die Frage nach der experimentellen Ökonomie auf, wieviele Versuchspersonen für die gesamte Experimentalreihe wohl zur Verfügung stehen (Schritte 230 und 232). Dora argumentiert dann in den Schritten 233 und 235 aufgabenorientiert: Da sie als Versuchspersonen so viel Zeit für das Experimentieren mit *virtue* hätten, müßten auch genügend virtuelle Versuchspersonen zur Verfügung stehen. In Schritt 237 entscheidet Dora dann abschließend, daß sie nicht weiter auf die Versuchspersonenzahlen achten möchte, und in Schritt 239 rechtfertigt sie diese Entscheidung auch noch ironisch mit der Bemerkung, sonst seien sie zu perfekt. Als sich die beiden dann auf eine Stichprobengröße von 30 Versuchspersonen pro Zelle des Versuchsplans festgelegt haben, reflektiert Dora ihre bisherigen praktischen Erfahrungen aus dem Projektseminar und freut sich darüber, daß diese Anzahl verglichen mit den Zellengrößen im Projektseminar beachtlich ist.

Dido und Dora lassen sich die Ergebnisse gleich in allen drei Darstellungsmodi anzeigen. Diese Strategie verfolgen sie auch bei allen weiteren Experimenten, weshalb dies nicht bei jedem experimentellen Zyklus erneut erwähnt wird. Da sich keine signifikanten Effekte ergaben, sind sie zunächst enttäuscht, bis sich Dora

daran erinnert, daß ihre Ausgangshypothese für Zyklus 1 gerade für den Faktor Größe der Anordnung keinen Effekt vorhersagte. Dann wenden sie sich dem zweiten Teil ihrer Hypothese über den Faktor Farbähnlichkeit zu. Dora argumentiert ab Schritt 281 qualitativ und beklagt sich in Schritt 285 über die Skalierung des Interaktionsplots (Transkript 33). Nach ihrer Meinung ist diese zu großzügig, so daß sich kleine Effekte nicht daran ablesen lassen. Dennoch will Dora für den Faktor

281 Dora (gleichzeitig) die wird sehr schnell erkannt
 282 Dido ja
 283 Dora ja, wir haben gesagt, dass wir ein bisschen, em, ja bei similar ein
 284 Dido ein bisschen, ja aber
 285 Dora bisschen, es ist wirklich extrem wenig, aber diese Skala ist auch,
 em
 286 Dido ja (gleichzeitig) (?) verändert
 287 Dora die Abstände sind ja ziemlich gross
 [...]
 293 Dora Ah, das ist unser Dings, „wollen Sie aufgrund der vorliegenden
 Resultate die Hypothese eins beibehalten, verwerfen?“ jetzt dürfen
 wir nicht den gleichen Mist machen, wie die in dem Experiment da,
 wie hat das geheissen?
 294 Dido also, ich meine, unsere Hypothese ist doch
 295 Dora mit dem BigTrak-Roboter
 296 Dido ja
 297 Dora oder was das auch immer war
 298 Dido nein, es war, ich glaube, das war diese
 299 Dora nein, etwas anderes
 300 Dido dämliche Geschichte mit diesen haploiden und diploiden Zellen,
 dieses grauenhafte Experiment
 301 Dora nein, wo war das, wo sie die Hypothese, die bestätigt
 302 Dido das war das
 303 Dora nein, nein, das war doch etwas anderes, oi, jetzt weiß ich nicht,
 was das war
 304 Dido Klahr und Dunbar
 305 Dora meinst du, dort geht es mit den Chromosomen, nein, dort hatten wir
 das nicht, diese Tabelle
 306 Dido ah, doch, du meinst den BigTrak, ja klar
 307 Dora genau
 308 Dido fälschlicherweise verworfen und so
 309 Dora genau
 310 Dido ja klar, du hast Recht
 311 Dora genau, es hat sich nicht bestätigt und sie haben das beibehalten
 oder verworfen und so weiter
 312 Dido ja, aber es hat sich doch bestätigt, oder? Also sagen wir einmal
 mehr oder weniger
 313 Dora (gleichzeitig) (liest) „die Anzahl der Elemente auf dem Display
 hat“
 314 Dido (gleichzeitig) oder „weiß nicht“, also, es hat sich natürlich nicht
 völlig bestätigt, sondern nur teilweise bestätigt [...]

- Fortsetzung -

Transkript 33. Dido und Dora nehmen Bezug auf den Bestätigungsfehler.

- Fortsetzung -

- 314 Dido dann schreiben wir hin „weiß nicht“ und sagen halt, dass sich der obere Teil, also der erste Teil der Hypothese, zwar bestätigt hat, der andere aber nicht
- 315 Dora ja, sind die Reaktionszeiten etwas geringer“
- 316 Dido (gleichzeitig) oder?
- 317 Dora ja, obwohl, sie haben sich eigentlich, etwas geringer, wir haben nicht gesagt, dass sie signifikant unterschiedlich sind, die Reaktionszeiten
- 318 Dido warte
- 319 Dora ich hätte jetzt gesagt, dass sie sich, dass wir sie beibehalten können
- 320 Dido ja
- 321 Dora weil sie sind ein bisschen
- 322 Dido ja, komm, wir lassen die beibehalten, ja
- 323 Dora hm?
- 324 Dido ja, wir behalten sie bei und sagen dann oben noch, okay, es ist zwar nicht ganz so, unten so, also, wir hätten dann schon ein bisschen mehr erwartet, aber
- 325 Dora ja, aber ich denke auch, bei diesem Dings da, bei den Experimenten sonst, sind die, eh, sind die Reaktionszeiten, zum Beispiel bei dieser dual task-Dings, sind es ja auch 0.6 Sekunden
- 326 Dido ja
- 327 Dora und das ist ein Unterschied und dann hätten wir vielleicht auch so etwas, weil es ist ein bisschen oberhalb von jedem Punkt
- 328 Dido ja, gut
- 329 Dora okay
[...]
- 338 Dido ja doch, also, wir haben aber, wir müssen einmal unsere Daten angucken, bevor wir jetzt hier, uns auf einmal auf den Interaktionsplot stützen, ja gut, also, wir erkennen da sind alle nicht
- 339 Dora (gleichzeitig) es gibt keine signifikante
- 340 Dido signifikant, und deshalb
- 341 Dora ja
- 342 Dido ja
- 343 Dora genau
- 344 Dido gut [...] hmm [...] ja, das stimmt halt
- 345 Dora ja, genau, ich glaube, das ist auch
- 346 Dido ja, ist die Frage
- 347 Dora ich glaube, auf das kommt es jetzt auch nicht so, also, ich würde
- 348 Dido also, „die Daten entsprechen unseren Erwartungen“ (schreibt)

Transkript 33 (Fortsetzung). Dido und Dora nehmen Bezug auf den Bestätigungsfehler.

Farbähnlichkeit einen Effekt erkennen (Schritt 285). Sie ist sich der Problematik einer fälschlicherweise bestätigten Hypothese jedoch bewußt und greift auf ihr Vorwissen zu den Experimenten von Klahr und Dunbar (1988) sowie Dunbar (1993) zurück. Gemeinsam mit Dido erarbeitet sie sich bis Schritt 311 das Konzept des Bestätigungsfehlers. Dabei geht es zunächst um eine Rekonstruktion aus dem Gedächtnis anhand der Analogie von Oberflächenmerkmalen der beiden genannten

Studien (Schritte 295, 300 und 305). Zwischendurch nennen Dido und Dora abwechselnd die für die Bewertung ihrer Hypothese relevanten Aspekte der beiden Untersuchungen (Schritte 301, 308 und 311).

In Schritt 312 kommt Dido dann auf die Daten aus dem eigenen Experiment in *virtue* zurück. Hier wird der Austausch zwischen den beiden Teilnehmerinnen ausgesprochen intensiv, sie sprechen wieder über mehrere Schritte gleichzeitig. Dido schwankt zwischen Bestätigen und Verwerfen ihrer Hypothese und formuliert das Dilemma in Schritt 314, nämlich daß die Hypothese zweiteilig sei und nur einer der beiden Teile durch die Daten bestätigt worden sei. Damit wäre die logische Schlußfolgerung aus den Daten eigentlich, die Hypothese zu verwerfen. Dido schlägt jedoch vor, in der Bewertung beide Teilhypothesen getrennt zu behandeln.

Dora verweist auf den Text der Hypothese und geht erst in Schritt 317 auf die Argumentation von Dido ein. Um die Hypothese doch noch bestätigen zu können, greift sie auf qualitative Aspekte des Interaktionsplots zurück und übergeht die Inferenzstatistik mit dem nicht signifikanten Haupteffekt für den Faktor Farbähnlichkeit (Schritte 319, 321, 325 und 327). Hierbei unterläuft ihr in Schritt 317 ein methodischer Fehler, indem sie bemerkt, in der Hypothese zwar eine Tendenz, aber keine statistische Signifikanz vorhergesagt zu haben. Es ist ihr in diesem Augenblick offenbar nicht gegenwärtig, daß eine statistische Signifikanz eine Aussage darüber macht, mit welcher Wahrscheinlichkeit eine Hypothese zutrifft.

Dido signalisiert in Schritt 322, daß sie Doras Argumentation zustimmt, lenkt jedoch in Schritt 338 die Aufmerksamkeit nochmals auf die Inferenzstatistik und damit den quantitativen Aspekt der Dateninterpretation. Hier kommt es in den Schritten 338 bis 340 zu einer gegenseitigen Ergänzung, Dora vervollständigt Didos Äußerung. Beide sind sich an dieser Stelle unschlüssig, wie sie die Inferenzstatistik nun einordnen sollen (Schritte 341 bis 346), und deshalb kann Dora in Schritt 347 vorschlagen, dieser bei der Interpretation keine weitere Beachtung zu schenken. So bestätigen beide schließlich trotz der Diskussion um den Bestätigungsfehler ihre Hypothese und schreiben als Begründung, daß die Daten ihren Erwartungen entsprächen. Der differenzierte Interpretationsvorschlag von Dido aus Schritt 314 findet keinen Eingang in die endgültige Bewertung der Hypothese.

Im ersten Zyklus wird bereits deutlich, daß beide Teilnehmerinnen gleichermaßen zum Dialog beitragen. Während Dido bei der Versuchsplanung bestimmend wirkt, übernimmt Dora bei der Interpretation der Ergebnisse die Initiative. Die Auswahl der Variablen für das erste Experiment erfolgte anhand formaler Kriterien und fiel auf den ersten Faktor in der Liste des Hypothesenscratchpads. Die Umsetzung der Hypothese in einen Versuchsplan machte den beiden keine Schwierigkeiten. Bei der Interpretation der Ergebnisse unterläuft ihnen jedoch ein Fehler: Sie bestätigen fälschlicherweise ihre Hypothese trotz einer Diskussion über den Bestätigungsfehler. Die Ursache ist möglicherweise in zwei Faktoren zu suchen, einerseits in einer ausführlichen Diskussion der qualitativen Interpretation des Interaktionsplots und andererseits in einem Mißverständnis dessen, was eine Hypothese aussagt und wie statistische Signifikanzen darauf zu beziehen sind.

Bereits im ersten Zyklus aktivieren beide Teilnehmerinnen ihr Vorwissen. Sie erinnern sich an Inhalte aus der Vorlesung zum Lernen und Problemlösen, insbesondere an die Studie von Klahr und Dunbar (1988). Und sie diskutieren die Stichprobengröße und vergleichen diese mit ihren Erfahrungen aus dem Projektseminar.

Dido und Dora: Zyklus 2

Nach Abschluß von Zyklus 1 lesen sich die beiden Teilnehmerinnen nochmals die Theorie durch. Sie sprechen währenddessen verschiedene Faktoren an, einigen sich dann aber, das Experiment aus Zyklus 1 zu replizieren und dabei den Faktor Farbkombination auf der Stufe blau/gelb, statt wie vorher rot/grün konstant zu halten. Obwohl sie bei der Interpretation ihrer Hypothese in Zyklus 1 Schwierigkeiten hatten und diese nur teilweise bestätigen konnten, übertragen Dido und Dora ihre Vorhersagen aus Zyklus 1 unverändert in Zyklus 2 und sagen für die veränderte Farbkombination dieselben Effekte voraus.

Mitten in der Ausarbeitung der Hypothese für Zyklus 2 wechselt Dido kurzfristig die Dialogebene und äußert ihre Gedanken zum Experiment, an dem sie gerade teilnimmt (Transkript 34). Dora vollzieht den Wechsel der Dialogebene ebenfalls. Während Dido jedoch von Selbsterklärungen spricht, bezieht sich Dora wieder auf die bereits von den beiden erwähnte Studie von Klahr und Dunbar (1988). Ihnen ist im Gedächtnis geblieben, daß Theoretiker beim Experimentieren bessere Ergebnisse erzielen, weshalb sie wie Theoretiker vorgehen wollen. Ob sie

391	Dido	vielleicht machen die auch ein Experiment so wie das mit den Selbsterklärungen
392	Dora	ja, wir als Dyade sowieso, und ob wir zu den Experimen, Experimentatoren oder zu den Theoretikern gehören
393	Dido	mh, wir gehören zu den Theoretikern
394	Dora	ja, unbedingt, weil, die sind viel besser.

Transkript 34. Dido und Dora klären ihr Vorwissen zu Theoretikern und Experimentatoren nach Klahr und Dunbar (1988).

auch wissen, was genau das Vorgehen der Theoretiker so erfolgreich machte, wird aus ihrem Dialog nicht deutlich. Da sie jedoch vor ihrem ersten Experiment eine Hypothese aufgestellt haben, ist ihr bisheriges Verhalten eher demjenigen der Theoretiker zuzuordnen.

Bei der Interpretation der Ergebnisse will Dido die Hypothese zunächst verwerfen (Schritt 495, Transkript 35). Dora kann Dido jedoch von ihrer methodisch korrekten Interpretation überzeugen, daß die Hypothese in diesem Experiment von den Daten bestätigt wurde (Schritte 506, 508, 510). Sie bedient sich dabei erstmals auch der entsprechenden statistischen Begriffe.

Während die Teilnehmerinnen die Begründung für ihre Interpretation der Hypothese formulieren und dabei nach den Abstufungen des Faktors Farbkombination differenzieren, zieht Dido die Schlußfolgerungen dazu in Zweifel (Schritt 541, Transkript 36). Auch wenn der Faktor Farbkombination über die beiden bisherigen Experimente verändert wird und damit ein pseudo-dreifaktorielles Experiment durchgeführt wird, läßt sich kein direkter Vergleich wie in einem einzigen Experiment herstellen. Dies führt dazu, daß die beiden in der Begründung für

495	Dido	also, die müssen wir verwerfen
496	Dora	(gleichzeitig) was machen wir da? verwerfen [...]
506	Dora	aber, em, eigentlich haben wir da eine, ah, doch, es ist ja ein bisschen oberhalb, das schwarze
507	Dido	ja
508	Dora	das heisst, dass wir eigentlich einen signifikanten Haupteffekt Farbe haben, und wir haben aber auch noch die Interaktion
509	Dido	ja
510	Dora	und der Haupteffekt ist ja trotz der Interaktion interpretierbar, wenn ich mich nicht täusche, du müsstest mich korrigieren, wenn das nicht stimmt. Also, wenn man sich das vorstellt
511	Dido	ja

Transkript 35. Dora erklärt Dido die Interpretation eines signifikanten Haupteffekts.

541	Dido	(gleichzeitig) Scheisse, wir haben einen Fehler gemacht, wir haben ja gar nicht direkt verglichen
542	Dora	wie meinst du direkt?
543	Dido	wir haben ja gar nicht direkt verglichen, rot versus blau, sondern wir haben einen rot genommen und den anderen blau, wir haben voll den Scheiss gemacht
544	Dora	aber man kann das nur in Paketen anwählen, oder?
545	Dido	ja
546	Dora	also, es sind zwei verschiedene
547	Dido	ah, stimmt
548	Dora	zwei verschiedene Seiten in diesem Farbkreis
549	Dido	aber wir hätten ja eigentlich die Ergebnisse untersuchen müssen, wenn man rot benützt und wenn man blau benützt

Transkript 36. Dido bemerkt die begrenzten Interpretationsmöglichkeiten pseudo-dreifaktorieller Experimente.

die Interpretation der Hypothese schließlich folgendes schreiben: „[. . .] und zwar steigen die Reaktionszeiten signifikant an, wenn man andere Farben aus dem Farbspektrum verwendet“. Mit dieser Formulierung ist noch keine Aussage über einen Haupteffekt des Faktors Farbkombination gemacht. Die Überlegungen von Dido beeinflussen jedoch das weitere Vorgehen. Die Teilnehmerinnen beschließen auf Didos Initiative hin, als nächstes den Faktor Farbkombination direkt zu testen.

Dido und Dora: Zyklen 3 und 4

Weil Dido unsicher ist, ob ein Experiment mit Variation des Faktors Farbkombination überhaupt möglich ist (z.B. Schritt 544, Transkript 36), schlägt sie in Schritt 619 vor, dieses Design einfach auszuprobieren (Transkript 37). Dafür ist in ihren Augen keine Hypothese notwendig. Doras Einwurf, sie könnten wenigstens eine kurze Hypothese formulieren (Schritt 622), bleibt unberücksichtigt.

Bei der Versuchsplanung wird deutlich, daß der Faktor Farbkombination experimentell variiert werden kann. So kommt Dido nochmals auf die nicht ausformulierte Hypothese zurück (Schritt 649). Dora schlägt dann vor, einfach so zu tun, als sei die Hypothese bereits formuliert worden. Dido stimmt zu und geht noch weiter, indem sie in Schritt 661 fordert, die Versuchsleiter sollten ihre Hypothese, wahrscheinlich die in Schritt 647 geäußerten Erwartungen, aus dem Video herauslesen.

Nach Abschluß der Versuchsplanung für Zyklus 3 führen Dido und Dora anschließend nicht gleich auch das dazugehörige Experiment durch, sondern formulieren die Hypothese für Zyklus 4. Dazu kommt es, weil Dido nach der Ver-

-
- 619 Dido komm, alles löschen. Wir probieren's mal, wir müssen das Experiment ja nicht durchführen, wir gucken einfach einmal, ob das vom Design überhaupt geht
- 620 Dora ja, okay
- 621 Dido em, wie soll man „die“
- 622 Dora wir können nur etwas ganz Kurzes schreiben und überhaupt nicht überflüssig, eine Hypothese zu formulieren, falls es nicht anzuklicken ist, und wir können dann zurückkehren
- 623 Dido ja
- 624 Dora und die Hypothese ausformulieren
- 625 Dido ja
- 626 Dora was meinst du?
- 627 Dido es sei denn, er lässt das nicht zu, aber wir können's probieren
- 628 Dora ja, wir können irgendetwas da, vielleicht so,
- 629 Dido (schreibt nur iii)
[...]
- 647 Dido jetzt müssen wir da gucken, ob wir uns jetzt auf unterschiedliche Farben konzentrieren, also blue/yellow, pink/green, aber es wäre eigentlich interessant, ob pink/rot und blue und, blau und hellblau sich unterscheiden, finde ich, weil, dass die beiden sich irgendwie unterscheiden, ist klar
- 648 Dora ja, ob das innerhalb von diesem Dings da
- 649 Dido mh [...] ich würde sagen, jetzt gehen wir zurück, formulieren einmal eine Hypothese
- 650 Dora ja, okay, dann
- 651 Dido mm, wir dürfen das ja nicht so zumachen
- 652 Dora nein, nein, dann wählen wir, dann klicken wir sie an, als hätten wir die formu, Hypothese formuliert, und dann gehen wir zu der Hypothese zurück
[...]
- 658 Dora dass man das dann
- 659 Dido jaja
- 660 Dora ausgraben kann, aber ich denke schon, oder? Das kann doch nicht sein
- 661 Dido sonst haben sie's jetzt ja mitgekriegt, was unsere Hypothese war. Also machen wir eher, em similar?
-

Transkript 37. Dido und Dora nehmen zwei Anläufe für ihr nächstes Experiment.

suchsplanung doch noch ihre Hypothese für Zyklus 3 aufschreiben möchte, im Prozeß des Formulierens aber zu einer Vorhersage für ein 2x2 Design kommt, bei dem sie Farbähnlichkeit mit Farbkombination gemeinsam testen möchte. Sie kommt wieder zu einer zweiteiligen Hypothese mit getrennten Vorhersagen für beide Faktoren. Dora ist einverstanden, und so definieren die beiden auch gleich den Versuchsplan für das Experiment in Zyklus 4 und führen dieses durch. Auf diese Weise bleibt Zyklus 3 unvollständig, da das dafür geplante Experiment nie durchgeführt wird.

813	Dido	besteht halt wieder aus zwei Teilen
814	Dora	(gleichzeitig) aus zwei Teilen und eins hätten wir jetzt, hätten
815	Dido	ja
816	Dora	der erste Teil hätte sich jetzt bestätigt, und der zweite Teil
817	Dido	ja
818	Dora	ist wegen der Interaktion
819	Dido	ja
820	Dora	nicht
821	Dido	also, das können wir als „weiß nicht“ sagen, und dann können wir halt sagen, weshalb wir das glauben, also, dass der eine Teil

Transkript 38. Interpretation einer nur teilweise bestätigten Hypothese.

Bei der Auswertung der Ergebnisse von Zyklus 4 sind Dido und Dora wieder mit ähnlichen Daten konfrontiert wie in Zyklus 2. Dido erkennt dies in Schritt 813 (Transkript 38). Dora liefert wie in Zyklus 2 die Interpretation. Schwierigkeiten bereitet ihr die signifikante Interaktion. In diesem Zyklus brauchen die beiden jedoch nicht so lange, um zu einer gemeinsamen Auffassung zu kommen. Sie entscheiden sich, ihre Hypothese mit „weiß nicht“ zu bewerten, und begründen dies wie folgt:

Da die Hypothese aus 2 Teilen besteht, und der 1. Teil bestätigt wurde (je nach Farbkombinationen gibt es unterschiedliche Reaktionszeiten), der 2. Teil aber nicht (Interaktionen), können wir die Hypothese weder beibehalten noch verwerfen. Die Farben wirken sich auf die Reaktionszeit unterschiedlich aus, es findet hier eine Wechselwirkung statt.

Mit dieser unentschiedenen Einstufung der Hypothese kommen Dido und Dora zwar weg von der logisch inkorrekten Interpretation, wie sie sie im ähnlichen Fall in Zyklus 2 abgegeben haben, logisch wäre die Hypothese aber zu verwerfen, da die Teilnehmerinnen für beide Faktoren einen Effekt erwartet hatten und die Resultate nur einen Haupteffekt für den Faktor Farbähnlichkeit ergaben. Auch inhaltlich ist die zitierte Begründung falsch. Gerade für den Faktor Farbkombination, in der Begründung als derjenige mit einem Haupteffekt genannt, fand sich nämlich kein signifikanter Haupteffekt, wohl aber für den Faktor Farbähnlichkeit, dessen statistisch bedeutsamer Effekt in der Begründung nicht erwähnt wird. Die Teilnehmerinnen verwechseln die beiden Teile der Hypothese, weil sie sich in ihrer Diskussion ab Schritt 814 nicht mehr auf die Inhalte beziehen, sondern nur noch

vom „ersten und zweiten Teil“ der Hypothese sprechen. In Schritt 814 wirkt es noch so, als wolle Dora ausdrücken, daß sich ein Teil der Hypothese bestätigt habe („eins“), ohne dabei festzulegen, ob sie damit den ersten oder den zweiten Teil der Hypothese meint. In Schritt 816 unterläuft Dido dann eine Verwechslung. Aus „eins“ wird „der erste Teil“. Und dieser bezieht sich in der Hypothese auf die Farbkombination, statistisch signifikant ist aber der Faktor Farbähnlichkeit. Als die Teilnehmerinnen dann ihre Interpretation schriftlich festhalten, beziehen sie den „ersten und zweiten Teil“ mit den dazugehörigen Feststellungen über signifikante Haupteffekte wieder auf die korrekte Reihenfolge der Faktoren in der Hypothese und kontrollieren nicht mehr, ob das inhaltlich stimmt.

In den Zyklen 3 und 4 zeigt sich, daß die Teilnehmerinnen aus den ersten beiden Zyklen systematisch eine neue Fragestellung ableiten und diese zunächst eher explorativ, dann aber anhand einer ausgearbeiteten Vorhersage überprüfen. In Zyklus 3 scheinen Dido und Dora kurzfristig die Strategie der Experimentatoren nach Klahr und Dunbar (1988) zu adaptieren, wechseln aber mit ihrer Hypothese für Zyklus 4 wieder zur Strategie der Theoretiker. Sie durchsuchen damit den Hypothesenraum und den Problemraum zielgerichtet. Die Initiative bei der Festlegung der zu untersuchenden Faktoren hat in diesen beiden Zyklen wieder Dido. Die Interpretation erarbeiten in diesem Zyklus beide gemeinsam. Bei der Bewertung der Hypothese unterläuft ihnen allerdings eine Verwechslung, die letztlich zu einer falschen Begründung der Interpretation führt.

Dido und Dora: Zyklen 5 bis 7

Der Vorschlag, sich von den Farben abzuwenden und die Faktoren Größe der Anordnung und Exzentrizität zu untersuchen, kommt wieder von Dido (Schritte 863 und 867, Transkript 39). Dora ist damit einverstanden. Dido äußert auch einen weitergehenden Plan, nämlich den Faktor Suchtyp ganz zum Schluß zu untersuchen. Ihre Begründung dafür lautet, daß dieser Faktor besonders schwierig zu testen sei.

Zunächst planen Dido und Dora eine kombinierte Hypothese über die beiden Faktoren Größe der Anordnung und Exzentrizität. Dies schlägt sich darin nieder, daß sie die beiden Faktoren im Hypothesenscratchpad ankreuzen. Beim Ausformulieren hat Dido aber eine andere Idee (Schritt 913, Transkript 40). Sie möchte die

863	Dido	ich finde, wir sollten jetzt einmal von diesen Farben wegkommen, das macht einem ja wahnsinnig
864	Dora	(gleichzeitig) ich finde, ja, genau
865	Dido	na
866	Dora	wir könnten jetzt etwas ganz anderes nehmen
867	Dido	(gleichzeitig) wir könnten jetzt einmal gucken, ob das mit der set size, je mehr und mit der eccentricity zusammenhängt [...]
872	Dora	ja, genau, okay, dann nehmen wir das und das
873	Dido	genau
874	Dora	ja
875	Dido	und dann ganz zum Schluss werden wir uns dann, wenn wir damit fertig sind, dann stürzen wir uns einmal auf die unterschiedlichen Suchtypen, oder? Das wird dann noch kompliziert

Transkript 39. Dido legt die weiteren Experimente fest.

Komplexität der Hypothese reduzieren und schlägt daher vor, für jeden Faktor eine unabhängige Hypothese aufzustellen. Da sich die bisherigen Hypothesen nie auf eine Interaktion bezogen, ist diese Idee nur folgerichtig. Dido braucht allerdings einige Schritte, um Dora von ihrem Vorschlag zu überzeugen (Schritte 913, 915 und 917). Signalisiert Dora ihre Zustimmung sonst immer mit einem eindeutigen „ja“ oder „okay“ (Schritte 872 und 874, Transkript 39), so antwortet sie in diesem Fall zurückhaltender mehrfach mit „mh“ (Schritte 914, 916 und 918, Transkript 40).

913	Dido	ehh!! Ich würde sagen, das sind, em, da machen wir jetzt für jede eine Hypothese, das schreiben wir nicht alles in, in eine, in ein insgesamt, sondern machen drei verschiedene Hypothesen dafür
914	Dora	mh
915	Dido	also, weißt du, es ist einfach übersichtlicher, finde ich
916	Dora	mh
917	Dido	weißt du, dann haben wir nicht wieder, okay, wir müssen jetzt den ersten Teil
918	Dora	mh
919	Dido	der leider noch nicht und
920	Dora	ist richtig, ja
921	Dido	also, wie beschreiben wir denn
922	Dora	okay, schaffst du das? [...]
942	Dora	sollen wir das gerade formulieren, oder sollen wir die auch überprüfen?
943	Dido	stimmt, das ist eigentlich noch gut, wenn wir die noch alle schnell formulieren, dann hat man nicht irgendwie schon wieder irgendwelche Gedächtniswindungen, das ist eigentlich gut, wenn wir die direkt formulieren
944	Dora	ja, wir können das machen

Transkript 40. Dido und Dora formulieren mehrere Hypothesen hintereinander weg.

Von Dora kommt der Vorschlag, noch vor der Überprüfung der Hypothese aus Zyklus 5 weitere Hypothesen aufzustellen (Schritt 942). Dido greift den Vorschlag auf, und so entstehen direkt nacheinander die Vorhersagen für die Zyklen 6 und 7. Die Abfolge von Zyklen wird hier aufgebrochen. Es folgt nicht mehr auf jede Hypothese deren Überprüfung. Stattdessen wird eine Reihe von Hypothesen formuliert, die anschließend in einer Reihe von Experimenten getestet werden. Die schriftliche Form der Hypothesen für die Zyklen 6 und 7 erarbeiten die beiden Teilnehmerinnen zügig innerhalb der nächsten 60 Schritte.

Bei der Versuchsplanung für das Experiment zu Hypothese 5 stellt Dora die Frage, auf welcher Stufe der Faktor Suchtyp nun konstant gehalten werden solle (Schritt 1022, Transkript 41). Nachdem sich die beiden Teilnehmerinnen in Zyklus 1 auf die Stufe Merkmalssuche festgelegt hatten, wird diese Entscheidung nun zum erstenmal wieder überdacht. Dora möchte in Schritt 1022 den Faktor Suchtyp wieder auf der Stufe Merkmalssuche festlegen. Dido erinnert im nächsten Schritt daran, daß darüber keine Aussage gemacht wurde (wie übrigens auch in der Hypothese von Zyklus 4). Sie plädiert dennoch zunächst dafür, auch in diesem Experiment den Faktor Suchtyp auf der Stufe Merkmalssuche konstant zu halten (Schritt 1027). Hier wendet Dora ein, daß sie ähnliche Experimente bereits durch-

1022 Dora und dann ist es hier, em, dieses da, oder?
 1023 Dido darüber haben wir gar keine Aussage gemacht
 1024 Dora stimmt, oi. Aber wir müssen etwas anklicken, oder nicht?
 1025 Dido ja, also wir haben bis jetzt immer Merkmalssuche gemacht
 1026 Dora ja, aber
 1027 Dido ich würde sagen, dann haben wir unsere Ergebnisse zur Merkmalssuche, und dann nachher können wir den ganzen Scheiss noch einmal für die conjunction Suche absuchen
 1028 Dora aber wir haben doch in der ersten Hypothese bereits, die wir formuliert haben
 1029 Dido da haben wir uns immer auf die Merkmalssuche gestützt, bis jetzt
 1030 Dora ja, genau, und wenn wir jetzt auch Merkmalssuche, dann, eh anklicken
 1031 Dido stimmt
 1032 Dora dann wäre das dasselbe
 1033 Dido ja
 1034 Dora was wir schon gemacht haben
 1035 Dido (gleichzeitig) ja, dann machen wir jetzt einmal
 1036 Dora (gleichzeitig) und deshalb machen wir dies
 1037 Dido ja
 1038 Dora conjunction

Transkript 41. Dido und Dora wechseln die Konstanthaltung des Faktors Suchtyp zur Verbindungssuche.

geführt hätten und kann Dido in den Schritten 1028 bis 1034 davon überzeugen, daß ein Wechsel zur Stufe Verbindungssuche sinnvoll wäre. Didos Äußerung in Schritt 1027 zeigt auch, daß ihr die Bedeutung des Faktors Suchtyp bisher nur ansatzweise klar ist, insbesondere bezieht sie Wechselwirkungen mit diesem Faktor nicht in ihre Überlegungen ein.

Die Ergebnisse von Experiment 5 sind so eindeutig, daß Dido und Dora innerhalb von 10 Schritten zu einer korrekten Interpretation kommen. Auch die Versuchsplanung und Interpretation für Experiment 6 gestalten die beiden sehr effizient und kommen zu korrekten Schlußfolgerungen. Die Ergebnisse zu Experiment 7 betrachten beide wieder etwas ausführlicher (Transkript 42). Dabei beziehen sie sowohl die Inferenzstatistik (Schritt 1138) als auch den Interaktionsplot (Schritte 1141, 1143 und 1145) mit ein. In diesen Schritten zeigt Dido auch, daß sie die signifikante Interaktion in der Graphik wahrgenommen hat, auch wenn sie sie nicht mit den inferenzstatistischen Daten in Verbindung bringt. In Schritt 1147 bringt Dido wieder Kritik an der Skalierung des Interaktionsplots vor. Auch hier wird nochmals deutlich, daß sie an dieser Stelle keinen Zusammenhang zwischen den Signifikanzen der Haupteffekte und dem Interaktionsplot sieht, sonst könnte sie nicht davon sprechen, daß sich der Faktor Größe der Anordnung kaum auswirkt.

Mit dem Experiment 7 und dessen Interpretation sind die Zyklen 5 bis 7 abgeschlossen. Die Abfolge, wie sie aus dem bisherigen Vorgehen von Dido und Dora sowie den Versuchspersonen in den anderen analysierten Transkripten einge-

1138	Dora	hier hat es auch viele Sterne
1139	Dido	holala, das schaut aber aus
1140	Dora	oioioioioi, haben wir das
1141	Dido	aber guck mal, das da, steigt bei, bei medium und large steigt das aber tatsächlich noch einmal ganz schön an
1142	Dora	ja, das ist wahr
1143	Dido	ah, guck mal, das ist ja witzig, dass die, em, wenn das wenig
1144	Dora	also die
1145	Dido	Exzentrizität, dann wirken sich auch die, die Dinger hier, die Anzahl, die set size wirkt sich dann fast gar nicht aus, also so ein bisschen halt, obwohl ja gut, das ist alles wieder abhängig von dem
1146	Dora	ja
1147	Dido	von der Skala

Transkript 42. Dido und Dora beziehen qualitative und quantitative Aspekte in die Interpretation ihrer Hypothese ein.

halten wurde, nämlich jede Hypothese gleich anhand eines Experiments zu überprüfen, wird von Dido und Dora hier aufgebrochen. Stattdessen formulieren sie gleich drei Hypothesen im Block und überprüfen dieses Paket anschließend in den entsprechenden Experimenten. Der Faktor Suchtyp erscheint beiden Teilnehmerinnen sehr komplex, so daß sie dessen Untersuchung zurückstellen. Sie behalten ihn jedoch im Auge, denn in Experiment 5 wechseln sie dessen Konstanthaltung von der Stufe Merkmalsuche zur Stufe Verbindungssuche. Obwohl die Teilnehmerinnen Interaktionen immer wieder qualitativ beschreiben, fließt dies nicht in die schriftliche Begründung der Hypothesenbewertung mit ein und findet auch keine Berücksichtigung bei der Formulierung neuer Hypothesen. Hierin manifestiert sich, daß Interaktionen von hoher Komplexität sind und es daher für Novizen besonders schwierig ist, diesen bei der Interpretation der Ergebnisse gerecht zu werden.

Dido und Dora: Zyklen 8 und 9

Die Zyklen 8 und 9 gehen Dido und Dora ähnlich an wie die Zyklen 5 bis 7, indem sie Hypothesen für gleich zwei Experimente aufstellen. Für Hypothese 8 schlägt Dido nun vor, sich dem Faktor Suchtyp in Kombination mit der Farbähnlichkeit zuzuwenden. In Schritt 1179 äußert sie: „[. . .] ja, komm, dann nehmen wir jetzt einmal search type und similarity [. . .]“. Dieser Vorschlag ist im Hinblick auf den Faktor Suchtyp konsequent, nachdem alle anderen Faktoren bereits untersucht wurden. Die Kombination mit dem Faktor Farbähnlichkeit wirkt jedoch etwas willkürlich und wird im weiteren Dialog nicht begründet.

In Hypothese 8 zum Faktor Suchtyp greifen Dido und Dora die Merkmals-Integrations-Theorie auf und formulieren eine entsprechende Hypothese für einen einfaktoriellen Versuchsplan, also ohne den Faktor Größe der Anordnung explizit zu erwähnen. Entgegen der ursprünglichen Absicht, Suchtyp und Farbähnlichkeit zusammen zu überprüfen (Schritt 1215, Transkript 43), zieht Dido beim Aufstellen von Hypothese 9 kurz in Erwägung, diese für das kritische Experiment (Suchtyp X Größe der Anordnung) zu formulieren, also unter Einbeziehen des Faktors Größe der Anordnung (Schritt 1219). Der inhaltliche Bezug wird in diesem Fall nur bei Betrachtung der Videoaufzeichnung deutlich. Dora wendet ein, daß es für diesen Plan nun zu spät sei (Schritt 1220), wobei dieses Argument nicht weiter ausgeführt

1215	Dido	dann machen wir mal. Also machen wir erst mal mit, was sollen wir machen? Mit der Ähnlichkeit und dem
1216	Dora	und den, ja, und
1217	Dido	obwohl, ne
1218	Dora	nein
1219	Dido	lass uns doch lieber erst das machen, obwohl ja das
1220	Dora	für das sind wir zu spät
1221	Dido	ja, jetzt ist es zu spät. Gut, aber man darf ja davon ausgehen, dass es sich hoffentlich bestätigen wird, dass die Verbindungssuche länger dauert, würde ich sagen, dass wir, em, dann können wir ja sagen, bei, ne
1222	Dora	ja, haben wir denn das nicht auch schon?
1223	Dido	wir müssen jetzt eigentlich die Verbindungssuche überprüfen, weil wir bei der Ähnlichkeit der Farben bisher immer die Merkmalssuche genannt, genommen haben

Transkript 43. Dido und Dora entscheiden sich gegen das kritische Experiment.

wird und auch nicht unmittelbar schlüssig erscheint. Dido versucht zu begründen, warum das kritische Experiment auch gar nicht so dringend notwendig wäre (Schritt 1221), indem sie die Ergebnisse von Experiment 8 vorweg nimmt. Dabei läßt sie die Interaktion zwischen Suchtyp und Größe der Anordnung außer acht, kann aber Dora von ihrer Argumentation überzeugen. So formulieren sie Hypothese 9 zu den Faktoren Suchtyp und Farbähnlichkeit. Den Versuchsplan zu Hypothese 8 stellen die beiden sehr rasch auf, fast schon in einer Art Routine, aber dennoch genau zur Hypothese passend.

Bei der Diskussion über die Ergebnisse von Experiment 8 erinnert sich Dido an Inhalte aus der Methodologievorlesung (Schritt 1307, Transkript 44). Sie rekapituliert das Poppersche Falsifikationsprinzip und stellt fest, daß sie in ihrem bisherigen experimentellen Vorgehen bestrebt gewesen seien, ihre Hypothesen zu

1303	Dido	das sieht ja schon mal schick aus
1304	Dora	ja, das ist wahnsinnig schön
1305	Dido	also wenn das nicht bestätigt ist
1306	Dora	hei, das ist ja ein wahnsinniger Anstieg, das sind ja
1307	Dido	(gleichzeitig) wir machen übrigens gerade im Fall den Fehler: wir versuchen nämlich die ganze Zeit, unsere Hypothesen zu bestätigen
1308	Dora	ja, komm [...]
1313	Dido	tolle Ergebnisse, genau wie wir es erwartet haben (schreibt Bewertung, schliesst das Resultatfenster, 8. Hypothese)
1314	Dora	em, haben wir das schon? Nein, he? neun, oder?
1315	Dido	ja. Wir können ja mal versuchen, dass wir vielleicht ein paar falsifizieren können.

Transkript 44. Dido spricht über die Falsifikation von Hypothesen.

bestätigen statt sie zu falsifizieren. Ihr Gedanke wird von Dora jedoch nicht aufgegriffen. Dido schwenkt um auf die Interpretation der Ergebnisse von Experiment 8 (Schritt 1313) und kommt in Schritt 1315 nochmals auf die Falsifikation zurück. Auch hier geht Dora nicht auf dieses Gesprächsthema ein, so daß Dido es schließlich fallen läßt.

Die Ergebnisse von Experiment 8 sind so eindeutig, daß die Teilnehmerinnen ihre Hypothese korrekterweise als bestätigt erachten. Eine inhaltliche Diskussion des Faktors Suchtyp und seines Bezugs zur Merkmals-Integrations-Theorie findet nicht statt. So ist eventuell auch zu erklären, warum es selbst nach diesem Experiment nicht zur Planung eines kritischen Experiments mit den Faktoren Suchtyp und Größe der Anordnung zur Überprüfung der Merkmals-Integrations-Theorie kommt. Eher sehen die beiden bereits das einfaktorielle Experiment als Bestätigung der Merkmals-Integrations-Theorie. Dido und Dora halten sich an die Vorhersage in Hypothese 9 und planen sehr effizient das dazu passende Experiment mit den beiden Faktoren Suchtyp und Farbähnlichkeit. Auch die Daten von Experiment 9 bestätigen die entsprechende Hypothese. Dido und Dora erkennen dies sofort, geben es in *virtue* ein und schließen so Zyklus 9 ab.

Auch in den Zyklen 8 und 9 bestimmt Dido wieder, welche Faktoren überprüft werden. Dido und Dora nehmen die Gelegenheit nicht wahr, das kritische Experiment doch noch als Experiment 9 durchzuführen. Bei der Formulierung von Hypothese 9 zieht Dido dies kurz in Erwägung. Aber die beiden Teilnehmerinnen verwerfen diese Idee schnell. Aus dem Transkript entsteht der Eindruck, sie vermeiden das kritische Experiment, weil sie Bedenken haben, gerade dieses Experiment sei besonders komplex und daher schwierig zu interpretieren und für solche Anstrengungen sei es nun zu spät. Andererseits interpretieren sie das einfaktorielle Experiment zum Faktor Suchtyp bereits als Bestätigung für die Merkmals-Integrations-Theorie.

Nach Zyklus 9 will Dora eine neue Hypothese aufstellen. Für beide Teilnehmerinnen stellt sich die Frage, welches Experiment noch nicht durchgeführt worden ist. Sie schauen jedoch nicht im Archiv nach, sondern lesen auf Didos Vorschlag hin nochmals die Theorie. Im Anschluß diskutieren sie verschiedene Varianten für Experimente, können sich jedoch nicht mehr einigen. Ein Beispiel

1395	Dido	haben wir untersucht, wie das ist, mit dem, em, Unterschied, ah, doch, also Merkmalssuche/Verbindungssuche in Bezug auf die Farbe?
1396	Dora	haben wir, eh, das, em, das haben wir einfach
1397	Dido	ja, mit dem similarity
1398	Dora	genau
1399	Dido	also, wir könnten jetzt noch untersuchen, wie das, ja, das ist Blödsinn
1400	Dora	wir können einfach die andere Farbe jeweils anklicken
1401	Dido	ja
1402	Dora	aber das
1403	Dido	obwohl, das wäre noch interessant mit der, mit der Verbindungssuche, ob das dann, wenn das hier ähnliche Farben sind, ob das dann schwieriger wird, also, wenn das diese blau, also ähnliche Farben, und dann eben diese
1404	Dora	mh

Transkript 45. Dido und Dora rekapitulieren die Suche im Hypothesenraum.

für die Suche im Hypothesenraum findet sich in Transkript 45. Der Rest des Dialogs zwischen Dido und Dora bezieht sich vorrangig auf Themen abseits der Aufgabenstellung. Sie können sich nicht mehr auf eine weitere inhaltliche Fragestellung zur visuellen Suche einigen und werden nach weiteren rund 200 Schritten schließlich von der Versuchsleiterin gefragt, ob sie ihre Experimentalreihe an dieser Stelle beenden möchten, was sie dann auch tun.

Allgemein zeigt sich, daß Dido und Dora im Laufe ihrer Experimente zunehmend über Themen abseits der Aufgabenstellung sprechen, vor allem über Freizeit und soziale Aktivitäten. Dies wurde in der bisherigen Analyse nicht dargestellt, da es keinen inhaltlichen Zusammenhang mit der Experimentalreihe gibt. Solche Dialogabschnitte treten auch in Dyade 1 zwischen Dana und Delia auf, jedoch in deutlich geringerem Umfang. Meist werden diese Themen angesprochen, wenn sich im Ablauf der experimentellen Zyklen eine Zäsur ergibt, etwa zwischen Hypothesenaufstellung und Versuchsplanung. In Dyade 2 ist es in den meisten Fällen Dora, die die Aufmerksamkeit wieder auf die eigentliche Aufgabe lenkt und dafür sorgt, daß die Diskussion inhaltlich kohärent fortgesetzt wird.

Dido und Dora: Zusammenfassung

Bereits im ersten Zyklus diskutieren Dido und Dora ihr Vorwissen über den Bestätigungsfehler, kommen aber letztlich bei der Interpretation ihrer Hypothese genau dazu, ihre Hypothese zu bestätigen, obwohl die empirische Evidenz dagegen spricht. Hier kann man auch von tragem Wissen im Sinne von Gruber, Mandl und

Renkl (2000) sprechen. Die Teilnehmerinnen wissen zwar, daß die Gefahr eines Bestätigungsfehlers gegeben ist, und können ihn benennen, die praktische Umsetzung dieses Wissens stellt beide aber noch vor Schwierigkeiten. Immerhin wurde dieses Vorwissen durch die Simulation aktiviert und in Zusammenhang mit der konkreten Fragestellung bei der Interpretation von Hypothesen gestellt. Eventuell hat sich diese Diskussion positiv auf den weiteren Verlauf der Experimentalreihe ausgewirkt. Ein Bestätigungsfehler wäre auch im nächsten Experiment möglich gewesen, dort kommen die Teilnehmerinnen jedoch korrekt zu der Entscheidung, daß ihre Hypothese zu verwerfen sei.

Im Verlauf der Experimentalreihe läßt sich eine Entwicklung bei Dido und Dora beobachten. Zunächst planen sie einzelne Experimente, später formulieren sie gleich mehrere Hypothesen in einem Block. Die Versuchsplanung macht ihnen von Beginn an wenig Schwierigkeiten und wird im weiteren Verlauf noch zunehmend effizienter. Während in den Transkripten der anderen Einzelfallanalysen die Reihenfolge, in der die einzelnen Faktoren überprüft werden, eher zufällig erscheint, leiten Dido und Dora ihre Hypothesen für die Zyklen 3 und 4 aus einer Fragestellung der ersten beiden Zyklen ab. Danach ziehen sie es vor, ihrer Erwartung nach einfachere Experimente zuerst durchzuführen, und prüfen zum Schluß denjenigen Faktor, der bisher noch nicht betrachtet wurde.

Dido und Dora führen das kritische Experiment zur Überprüfung der Merkmals-Integrations-Theorie nicht durch. Im Laufe des Dialogs vermitteln sie den Eindruck, dieses Experiment erschiene ihnen besonders komplex und schwierig zu interpretieren. Dies läßt sich daraus erschließen, daß sie Experimente mit dem Faktor Suchtyp schon zu Beginn verschieben. Einmal äußert Dido auch explizit die Erwartung, Experimente zum Faktor Suchtyp seien besonders kompliziert. Die letzten beiden Experimente beziehen dann den Faktor Suchtyp mit ein, werden jedoch sehr schnell geplant. Kurz kommt die Idee auf, das kritische Experiment durchzuführen, wird aber verworfen mit drei Argumenten: Es sei zu spät und die Daten der anderen Experimente könnten die Gültigkeit der Merkmals-Integrations-Theorie bereits belegen. Ein einfaktorielles Experiment zum Faktor Suchtyp wird ebenfalls als Bestätigung der zu überprüfenden Theorie herangezogen.

Dido und Dora nehmen während des Experiments von Anfang an unterschiedliche Rollen ein. Dido bestimmt durchgehend, welche Faktoren zu welchem Zeitpunkt untersucht werden, und übernimmt damit die Planung der Versuchsreihe. Sie tut dies in Abstimmung mit Dora, die Initiative liegt jedoch eindeutig bei Dido. Dora zeigt ihre Kenntnisse bei der Interpretation der Ergebnisse. Sie bringt die statistischen Begriffe in die Diskussion ein, bestimmt die inhaltliche Richtung der Interpretation und reflektiert die Bedeutung der Ergebnisse. Aber auch sie stimmt dies immer mit Dido ab, und beide zusammen sorgen dafür, daß immer alle drei Darstellungsmodi der Ergebnisse bei der Interpretation berücksichtigt werden. Zusätzlich übernimmt es Dora, das Gespräch nach Dialogabschnitten zu Themen abseits der Aufgabenstellung immer wieder auf die Aufgabe zu fokussieren.

Ein weiteres Merkmal der Dialoge von Dyade 2 sind die zahlreichen gleichzeitig gesprochenen Schritte. Sie stehen für einen intensiven Austausch, bei dem beide Teilnehmerinnen stark beteiligt sind. Gleichzeitig behindern sie einen kohärenten Dialog. Es ist den Teilnehmerinnen gar nicht möglich, sich aufeinander zu beziehen, weil sie sich gegenseitig nicht ausreden lassen. Dies führt dazu, daß die Teilnehmerinnen häufig mehr Austausch benötigen, um sich inhaltlich verständigen zu können, und erklärt neben den vielen Dialogabschnitten zu Themen abseits der Aufgabenstellung die Länge des Transkripts.

10.5 Metaaussagen

Während sich der Charakter der inhaltlichen Auseinandersetzung mit dem Gegenstandsbereich durch die vorangegangene Analyse plastisch darstellen ließ, finden sich Metabemerkungen in den verbalen Protokollen eher verstreut, so daß sich ein gesonderter Blick darauf lohnt. Metaaussagen wurden für diesen Zweck eigens kodiert und ausgezählt. Als Metaaussage wurde jeder Schritt gezählt, der die Reflexion und Überwachung des Problemlöse- und Lernfortschritts zum Inhalt hatte (Kneser & Plötzner, 2001).

Beispiele für Metaaussagen wurden bereits bei der Analyse der einzelnen Transkripte dargestellt. Sie finden sich für Ella in Transkript 2, Schritt 141, für Emelie in Transkript 11, Schritte 209 und 210, für Dana und Delia in Transkript 24, Schritte 427 bis 433 sowie für Dido und Dora in Transkript 31, Schritte 178 und 179.

Die Häufigkeiten der Metaaussagen sind in Tabelle 19 zusammengestellt. Betrachtet man die relativen Häufigkeiten, so äußern Teilnehmerinnen aus Dyaden ungefähr doppelt so viele Metaaussagen wie Einzelpersonen. Nimmt man die Aussagen für die beiden Partnerinnen einer Dyade zusammen, so ändert sich zwar an den relativen Häufigkeiten kaum etwas. Dennoch verarbeitet eine Dyaden-Teilnehmerin während ihrer Versuchssitzung sowohl die von ihr produzierten als auch die von der Partnerin geäußerten Metaaussagen. Dies sind innerhalb einer Versuchssitzung deutlich mehr Metaaussagen als bei Einzelpersonen.

Tabelle 19
*Absolute und relative Häufigkeiten
der Metaaussagen*

	Häufigkeit der Metaaussagen	
	absolute	relative
Einzelpersonen		
Ella	41	5.6%
Emelie	24	4.7%
Dyaden		
Dana (1)	35	9.6%
Delia (1)	58	13.3%
Dido (2)	89	11.7%
Dora (2)	84	11.4%

Um festzustellen, an welchen Stellen Metaaussagen gehäuft auftreten, wurden die Transkripte in fünf Kategorien unterteilt, die inhaltlich nach der Phase des Experimentierens mit *virtue* unterschieden wurden. Die Kategorien sind: Lesen der Theorie, Hypothesenformulierung, experimentelles Design, Ergebnisinterpretation und eine Restkategorie (Andere), in die vor allem die Gespräche abseits der Aufgabenstellung und die Kommunikation mit der Versuchsleitung fallen. Die relativen Häufigkeiten pro inhaltlicher Kategorie und die Metaaussagen je Kategorie sind in Tabelle 20 aufgeführt.

Bei der allgemeinen Verteilung der Schritte fällt auf, daß in drei von vier Fällen der Schwerpunkt der Dialoge auf der Interpretation der Ergebnisse liegt. Zweitstärkste Kategorie ist die Hypothesenformulierung. Die Ausnahme bildet Dyade 2, in der die Hypothesenformulierung etwas mehr Raum einnimmt als die Interpretation der Ergebnisse. Beide Kategorien zusammen machen in drei von vier Fällen rund 70% der Schritte aus. Nur Emelie mit ihrem hohen Anteil an Schritten in der Kategorie Andere kommt auf einen geringeren Anteil an Schritten in den Kategorien Hypothesenformulierung und Interpretation der Ergebnisse. Der

Tabelle 20

Zusammenhänge, in denen Metaaussagen auftreten

	Theorie	Hypothesen	Design	Interpretation	Andere
Einzelpersonen					
Ella					
% alle Schritte ^a	3.6	28.1	24.7	41.6	2.6
% Metaaussagen ^b	15.4	6.3	1.1	6.2	10.5
Emelie					
% alle Schritte ^a	1.9	25.5	20.1	29.6	22.9
% Metaaussagen ^b	40.0	5.5	1.0	5.3	1.7
Dyaden					
Dana und Delia (1)					
% alle Schritte ^a	1.8	31.1	17.2	36.0	13.9
Dana % Metaaussagen ^b	100.0	8.8	4.0	13.4	5.0
Delia % Metaaussagen ^b	100.0	11.0	10.7	19.7	5.0
Dido und Dora (2)					
% alle Schritte ^a	2.0	37.3	19.6	31.9	8.9
Dido % Metaaussagen ^b	38.7	10.6	12.7	11.5	4.4
Dora % Metaaussagen ^b	64.5	7.7	11.4	12.8	4.4

Anmerkung. ^a Die Prozentwerte in dieser Zeile bezeichnen den Anteil von Schritten dieser inhaltlichen Kategorie an der Gesamtzahl der Schritte (z.B. sind 28.1% von Ellas Schritten der Kategorie Theorie zuzuordnen). ^b Die Prozentwerte in dieser Zeile stellen den Anteil von Metaaussagen an den Aussagen innerhalb dieser inhaltlichen Kategorie dar (z.B. sind 6.3% von Ellas Schritten aus der Kategorie Hypothesen Metaaussagen).

hohe Anteil der Kategorie Andere ist auf die anfänglichen Dialoge mit dem Versuchsleiter zurückzuführen (siehe auch Transkript 9). Das Design nimmt in den Gesprächen nicht so viel Raum ein. Da die Teilnehmerinnen die Theorie bereits vor dem Experimentieren in *virtue* zu lesen bekamen, wird sie während der Experimentalreihe nicht mehr so häufig angeschaut, und dementsprechend wenige verbale Äußerungen finden sich hier.

Betrachtet man die Anteile der Metaaussagen innerhalb der inhaltlichen Kategorien, so zeigt sich in allen Fällen die Größte relative Häufigkeit für die Metaaussagen beim Lesen der Theorie. Bei Dyade 1 ist sogar jede Aussage beim Lesen der Theorie eine Metaaussage. Selbst bei Ella fallen in dieser Kategorie anteilig am häufigsten Metaaussagen, obwohl sie bei ihr nur 15,4% ausmachen. Das bedeutet, daß das Lesen der Theorie die Teilnehmerinnen zur Reflexion anregt beziehungsweise daß die Teilnehmerinnen dann die Theorie konsultieren, wenn sie in einem Reflexionsprozeß begriffen sind.

Während der Hypothesenformulierung und der Ergebnisinterpretation finden sich die zweithöchsten Anteile an Metaaussagen. Bei den Einzelpersonen sind diese Anteile deutlich geringer als bei den Dyaden und in beiden Kategorien fast gleich groß. Bei den Dyaden finden sich während der Interpretation der Ergebnisse etwas mehr Metaaussagen als bei der Hypothesenformulierung. Deutliche Unterschiede zwischen Einzelpersonen und Dyaden finden sich hinsichtlich der Kategorie Design. Während die Versuchsplanung Einzelpersonen kaum zu Metaaussagen anregt, produzieren Dyaden auch in diesem Abschnitt des experimentellen Zyklus etwa 10% an Metaaussagen.

Metaaussagen lassen sich noch weiter danach differenzieren, worauf sie sich beziehen. Chi, Bassok, Lewis, Reimann und Glaser (1989) konnten in ihrer Untersuchung zum Selbsterklärungseffekt feststellen, daß erfolgreiche Problemlöser in ihren Metaaussagen eher Verständnisprobleme identifizieren, wogegen weniger erfolgreiche Problemlöser häufiger Metaaussagen dazu äußern, was sie bereits verstanden haben. In einer kursorischen Analyse der vorliegenden Transkripte ergab sich, daß eine solche Kategorisierung für die Dyaden ungeeignet erscheint. Hier treten Äußerungen über den jeweiligen Wissensstand in anderen Formulierungen auf, etwa wenn eine Teilnehmerin die andere fragt, ob sie einen Aspekt verstanden

hab, oder wenn von einer Teilnehmerin bei sich selbst erkannte Verständnisschwierigkeiten gleich in einer Frage an die Partnerin münden. In Dyaden nehmen auch Aussagen über den Problemlöseprozeß, insbesondere die weitere Planung, mehr Raum ein, wie sich bereits aus der obenstehenden Analyse der Zusammenhänge ergab, in denen Metaaussagen auftreten. Insgesamt wäre ein breiteres Kategoriensystem notwendig, um den unterschiedlichen Metaaussagen von Einzelpersonen und Dyaden gerecht zu werden. Dadurch wird jedoch die Vergleichbarkeit der Daten von Einzelpersonen und Dyaden und in der Konsequenz der Wert der Aussagen einer solchen Analyse für die vorliegende Studie eingeschränkt. Um ein breiteres Kategoriensystem sinnvoll einzusetzen, wäre auch eine größere Datenbasis als die hier analysierten vier Transkripte notwendig; die Gesamtzahl der Metaaussagen in allen analysierten Transkripten wäre zu gering, um auf der Grundlage eines breiteren Kategoriensystems fundierte Aussagen treffen zu können. Auf eine weitere Auswertung der Metaaussagen wurde daher verzichtet.

Die dargestellten Ergebnisse weisen eindeutig in die Richtung, daß bei Dyaden mehr Metaaussagen geäußert werden als bei Einzelpersonen. Ausgehend davon, daß gerade Metaaussagen zur Reflexion und zum Lernprozeß beitragen, läßt sich aus der Verteilung schließen, daß diese gerade beim Lesen der Theorie und bei der Interpretation der Ergebnisse stattfinden. Dennoch dürfen die Ergebnisse nur mit Zurückhaltung interpretiert werden. Es handelt sich hier um eine Einzelfallanalyse mit 6 Versuchspersonen. Eine inferenzstatistische Auswertung wäre hier nicht angezeigt, weshalb weitergehende Schlußfolgerungen nur vorsichtig gezogen werden dürfen.

10.6 Vergleich von Dyaden und Einzelpersonen

Bereits aus dem vorangegangenen Kapitel zu den Ergebnissen aus der Logfileanalyse ergaben sich Hinweise auf Unterschiede zwischen Dyaden und Einzelpersonen. Um ein vollständigeres Bild zu erhalten, sollen diese Unterschiede in die folgende Betrachtung mit einbezogen werden. Drei Perspektiven bieten sich für diese Betrachtung an, diejenige der inhaltlichen Auseinandersetzung während des wissenschaftlichen Problemlösens, diejenige des Vorgehens im simulierten Labor und diejenige der Lernstrategien der Teilnehmenden.

Inhaltliche Breite der Äußerungen

In allen der vier qualitativ analysierten Transkripten zeigte sich, daß die Teilnehmenden ihre Äußerungen recht durchgängig auf die Aufgabe, also das wissenschaftliche Problemlösen, konzentrieren. Bei den Einzelpersonen finden sich ausschließlich Äußerungen, die sich eng an der Aufgabenstellung orientieren. Darüber hinausgehende Inhalte sprechen sie nicht an. Die Transkripte der Dyaden sind inhaltlich deutlich reicher als diejenigen der Einzelpersonen. Nur sie diskutieren etwa die Bedeutung der Konstanthaltung von Variablen. Die Auseinandersetzung mit diesem Thema kann hier als Beispiel dafür gesehen werden, was in den anderen, nicht in die qualitative Analyse einbezogenen Dyaden an Dialogen stattgefunden haben dürfte. Solche Diskussionen bieten dann auch eine direkte Erklärung für die quantitativen Befunde, daß Dyaden ihre Versuchspläne über mehrere experimentelle Zyklen hinweg konsequenter konstant halten, als Einzelpersonen. Eine weitere Diskussion in einer Dyade fand zur statistischen Signifikanz statt. Davon hätte auch Emelie profitieren können, die als Einzelperson in ihrer Experimentalreihe die Inferenzstatistik überhaupt nicht berücksichtigt.

Nur die beiden Dyaden sprechen Inhalte aus anderen Lehrveranstaltungen an, in diesem Fall die Experimente von Klahr und Dunbar (1988) und Okada und Simon (1997), und setzen diese zu ihrer eigenen Teilnahme am Experiment mit *virtue* in Bezug. In Dyade 2 wird ausführlich das Konzept des Bestätigungsfehlers aus Erinnerungen an die Inhalte der genannten Arbeiten rekonstruiert. Auch wenn dies nicht verhindern kann, daß die beiden Teilnehmerinnen in der direkten Folge der Diskussion dennoch einen Bestätigungsfehler begehen, ist davon auszugehen, daß das Konzept des Bestätigungsfehlers elaboriert wurde und das Vorwissen mit den Inhalten der praktischen Übungen zur Methodologie in *virtue* verknüpft wurde. Dies hat sich bei den Teilnehmerinnen dieser Dyade insofern positiv ausgewirkt, als ihnen wenigstens im Laufe ihrer weiteren Experimentalreihe kein Bestätigungsfehler mehr unterläuft. Die qualitative Analyse der Diskussion über den Bestätigungsfehler bietet einen Erklärungsansatz für den quantitativen Befund, daß Dyaden weniger Bestätigungsfehler begehen, insofern als davon auszugehen ist, daß ähnliche Diskussionen auch in anderen Dyaden stattgefunden haben könnten.

Nur die beiden Dyaden machen sich Analogien zunutze. In jedem der beiden analysierten Transkripte von Dyaden finden sich jeweils zwei Analogien. Analogien werden Dunbar (1997) zufolge in erfolgreichen Forschungsgruppen herangezogen, um sowohl technische Probleme bei der Realisierung von Experimenten zu lösen als auch um theoretische Perspektiven weiterzuentwickeln. Analogien sind daher ein eindeutiger Indikator für die Qualität der Äußerungen beim wissenschaftlichen Problemlösen. Auch in diesem Aspekt zeigen sich Dyaden gegenüber Einzelpersonen überlegen.

Sorgfältiges Vorgehen beim wissenschaftlichen Problemlösen

In der quantitativen Analyse fanden sich Indikatoren dafür, daß Dyaden in allen Phasen des experimentellen Zyklus sorgfältiger als Einzelpersonen vorgehen. Die von Dyaden formulierten Hypothesen sind sprachlich sauberer, als diejenigen der Einzelpersonen. Dyaden betreiben eine stringenterer Versuchsplanung als Einzelpersonen, was sich in der systematischeren Konstanthaltung nicht variierteter Faktoren über mehrere Experimente hinweg niederschlägt. Bei der Interpretation der Ergebnisse unterlaufen Dyaden weniger Fehlentscheidungen als Einzelpersonen, insbesondere der Bestätigungsfehler tritt seltener auf.

Für ein sorgfältigeres Vorgehen der Dyaden gibt es auch in den Transkripten Anhaltspunkte. Eine gute Basis für einen Vergleich bietet die Interpretation der Ergebnisse des jeweils ersten Experimentalzyklus. Es handelt sich um die Transkripte 1, 10, 20 und 33. Da Ella als einzige zuerst ein einfaktorielles Experiment durchführt, welches eine geringere Komplexität aufweist als die Experimente der anderen, kann hier auch noch Transkript 4 in die Betrachtung einbezogen werden. Im direkten Vergleich zeigt sich, daß in den Dyaden eine intensivere Auseinandersetzung stattfindet als bei den Einzelpersonen. Es scheint, daß bei den Einzelpersonen das Urteil über die Gültigkeit der Hypothese schnell feststeht. Sowohl Ella als auch Emelie äußern jeweils nur das, was sie schließlich auch in *virtue* als Bewertung der Hypothese eingeben. In beiden Dyaden werden die Ergebnisse intensiv analysiert und alternative Bewertungen für die Hypothese erwogen. Auch bei dieser ausführlichen Diskussion in den Dyaden können Fehler passieren, wie die falsche Hypothesenbewertung von Dido und Dora in Zyklus 1 belegt. Die Wahrscheinlichkeit dafür wird aber, wie auch die quantitativen Daten belegen,

deutlich verringert. Bei der Interpretation der Ergebnisse aus dem ersten Zyklus handelt es sich um ein Beispiel, die Interpretation eindeutiger Ergebnisse verläuft auch bei den Dyaden schneller, wenn sie in ihrer Experimentalreihe fortgeschrittener sind, dies trifft jedoch genauso auf die Einzelpersonen zu. Auch die weiteren Transkripte belegen, daß in den Dyaden ausführlicher abgewogen wird, was für oder gegen eine Entscheidung spricht.

Eine mögliche Erklärung für das sorgfältigere Vorgehen der Dyaden sind die Metaaussagen, die Dyaden häufiger äußern als Einzelpersonen. Metaaussagen sind ein Ausdruck der Reflexion über den Problemlöse- und Lernprozeß. Wenn bei Dyaden also mehr Reflexion stattfindet, erhöht dies beispielsweise auch die Wahrscheinlichkeit, daß mögliche Fehler entdeckt werden, bevor sie in den simulierten Experimenten umgesetzt werden.

Lernstrategien

Neben der vertieften Diskussion von Entscheidungen bieten die Dyaden auch ein breiteres Spektrum an Lernstrategien als Einzelpersonen. In den Protokollen der Dyaden finden sich mehrere Lernstrategien, welche sich beim kooperativen Lernen als erfolgreich erwiesen haben. Die Teilnehmerinnen in beiden Dyaden bringen mehr unterschiedliches Vorwissen ein und profitieren davon gegenseitig. Die laut Vortest schwächere Teilnehmerin aus Dyade 1, Delia, kann auf verschiedene Weise von ihrer Partnerin dazulernen. Sie bekommt auf viele ihrer Fragen von Delia Erklärungen auf einem ihr angemessenen Niveau. Dana dient auch als Modell für die Bewältigung der Aufgabe. Abstrakte Erklärungen setzt Delia direkt in konkrete Schritte beim Problemlösen mit der Simulation um, einerseits verbal, andererseits beim Eingeben in den Computer, welches der Sitzordnung zufolge ihre Aufgabe ist. In den beiden Dyaden nehmen die Teilnehmerinnen unterschiedliche Rollen ein, die sich für Lernprozesse beim kooperativen Problemlösen als hilfreich erwiesen haben. Während sich in Dyade 1 vor allem die Rollen der Erklärenden und der Fragenden/Beobachtenden finden, nimmt eine der Teilnehmerinnen in Dyade 2 eine planende Rolle ein, die andere eher eine überwachende, inhaltlich reflektierende.

Bei den Einzelpersonen finden sich deutlich weniger Lernstrategien als bei den Dyaden. Das Repertoire erstreckt sich hier darauf, die Theorie zu konsultieren, sich frühere Experimente nochmals zu vergegenwärtigen, überraschende Ergebnisse zu hinterfragen und zu reflektieren, sowie eine gewisse Stringenz bei der Planung der Experimentalreihe zu demonstrieren.

Die gemeinsame Betrachtung der quantitativen und qualitativen Ergebnisse der vorliegenden Studie veranschaulicht die Bedeutung, welche die Analyse von Einzelfällen haben kann. Dabei werden die typischen und typisierbaren Beziehungen von Einzelfällen zu allgemeinen Phänomenen und Strukturen deutlich. Über die Illustration und Anreicherung quantitativer Befunde hinaus offenbart sich hier das generative Potential der Einzelfallbetrachtung für die Weiterentwicklung wissenschaftlicher Modelle.

Diskussion

Wissenschaftliches Problemlösen zu vermitteln, ist zentrales Ausbildungsziel jedes Universitätsstudiums. Die Grundlage hierfür bilden komplexe Fertigkeiten in den Bereichen Theoriebildung, Formulierung von Hypothesen, Versuchsplanung, statistische Datenauswertung, Ergebnisinterpretation sowie bei der Integration von Inhalten und Methoden. Um Studierende beim Erwerb dieser Kompetenzen zu unterstützen, wurde *virtue* entwickelt, ein internet-basiertes simuliertes psychologisches Labor. In *virtue* können psychologische Experimente zur visuellen Suche geplant, durchgeführt und ausgewertet werden. Das simulierte Labor wurde mit kognitiven Werkzeugen ausgestattet, um das Erlernen von Fertigkeiten für das wissenschaftliche Problemlösen zu unterstützen und damit den Lernerfolg zu fördern.

In der vorliegenden Arbeit wurde untersucht, inwiefern Novizen beim Experimentieren im simulierten psychologischen Labor *virtue* Fertigkeiten zum wissenschaftlichen Problemlösen umsetzen und neues Wissen in diesem Bereich erwerben. Die Untersuchung zum wissenschaftlichen Problemlösen und Lernen mit *virtue* wurde mit zwei Experimentalgruppen von Psychologiestudierenden im Grundstudium, Einzelpersonen und Dyaden, durchgeführt. In parallelen Vor- und Nachtests zu Theorie, Methodologie und Empirie wurde deren Wissen dazu überprüft.

Die Ergebnisse werden in diesem Kapitel zunächst anhand der aufgestellten Hypothesen diskutiert, gefolgt von einer Einordnung der testdiagnostischen Resultate in den Zusammenhang der Untersuchung. Dann erfolgt eine Einschätzung der Untersuchung im Hinblick auf ihre Validität. Den Abschluß bilden offene Fragen und ein Ausblick auf Forschung und Praxis im Zusammenhang mit den in der Studie aufgeworfenen Fragestellungen.

11.1 Bewertung der Hypothesen

Hypothese 1: Die Experimente in *virtue* führen zum Erwerb von bereichsübergreifenden Fertigkeiten. Dies äußert sich in einer Verbesserung bei den bereichsübergreifenden Items zwischen Vortest und Nachtest und gilt für beide Experimentalgruppen.

Als Grundlage für die Arbeit mit *virtue* wurde den Versuchspersonen eine Einführung in die Theorie der visuellen Suche gegeben. Der Wissensstand dazu war im Untertest Theorie überprüft worden. Beide Experimentalgruppen zeigten bereits im Vortest ihr gleichermaßen hohes Ausgangsniveau, das sie zum Nachtest hin kaum noch steigern konnten. In diesem Untertest fanden sich keine Unterschiede zwischen den beiden Experimentalgruppen. Die notwendigen theoretischen Grundlagen konnten demnach durch die Instruktionmaterialien vermittelt werden. Es ist also davon auszugehen, daß die Versuchspersonen über ein ausreichendes Vorwissen verfügten, um sinnvolle Experimente in *virtue* zu planen.

Der Untertest Methodologie fragte bereichsübergreifende, methodische Fertigkeiten ab. Die Ergebnisse zum Untertest Methodologie zeigten im Prä-Post-Vergleich sowohl für die Gruppe der Dyaden als auch für die Gruppe der Einzelpersonen einen signifikanten Wissenszuwachs an. Es handelte sich hierbei nach Cohen (1977) um einen Effekt mit großer Effektstärke, der auch von praktischer Relevanz ist. Der Anstieg ist in beiden Gruppen vergleichbar, sie unterschieden sich also hinsichtlich des Wissenszuwachses nicht voneinander.

Den Testergebnissen zufolge ist es also möglich, durch Experimentieren mit dem simulierten psychologischen Labor *virtue* methodologische Fertigkeiten zu verbessern. Dies gilt gleichermaßen für Dyaden wie Einzelpersonen. Es ist davon auszugehen, daß dabei sowohl Übungsprozesse eine Rolle spielen als auch neue Einsichten, die die Lernenden während des Experimentierens gewinnen. Obwohl es sich bei den methodologischen Fertigkeiten um hochkomplexes Wissen handelt (Carlsen & Andre, 1992, Njoo & de Jong, 1993), war es möglich, diese zumindest teilweise in einem nicht allzu aufwendigen Papier-und-Bleistift-Test zu erfassen.

Hypothese 2: Durch das Experimentieren im simulierten Labor *virtue* wird auch bereichsspezifisches Wissen erworben. Es wird eine Verbesserung in den bereichsspezifischen Items vom Vortest zum Nachtest erwartet, und zwar für beide Experimentalgruppen, für Einzelpersonen wie für Dyaden.

Im Untertest Empirie wurde bereichsspezifisches Wissen zur visuellen Suche erhoben. Bereits im Vortest war in beiden Gruppen einiges Wissen über die visuelle Suche vorhanden. Dies ist darauf zurückzuführen, daß die Theorie der visuellen Suche, die ja durch die Experimente in *virtue* zu überprüfen war, bereits in der Instruktion dargestellt wurde. Aus dieser Theorie lassen sich Vorhersagen über die empirischen Ergebnisse ableiten. Diese Vorhersagen können nicht als Wissen bezeichnet werden, da sie ja auf der noch zu überprüfenden Theorie basieren. Dennoch lassen sich einige Fragen des Vortests beantworten, wenn die Ableitungen aus der Theorie korrekt vollzogen werden. Im Test kann dann nicht unterschieden werden, ob die Antwort zu einem der Faktoren auf dem noch hypothetischen Wissen aus der Instruktion zur Theorie beruht oder bereits empirisch fundiert ist.

Dennoch konnten sich die Versuchspersonen in beiden Experimentalgruppen beim Untertest Empirie zum Nachtest hin deutlich steigern. Verbesserungen der Lernleistung in dieser Größenordnung finden sich auch bei de Jong, Martin, Zamarro, Esquembre, Swaak und van Joolingen (1999) und bei Swaak und de Jong (2001). Statistisch weist der Lerneffekt beim Untertest Empirie nach Cohen (1977) eine große, also auch praktisch relevante Effektgröße auf. Angesichts der bereits im Vortest erreichten Werte liegt aber der Verdacht nahe, es könnte sich um einen Deckeneffekt handeln, und der Untertest könnte dann nicht ausreichend differenzieren. Eine Abhilfe könnten die im Nachtest zusätzlich erhobenen bereichsspezifischen Items darstellen. Sie wurden im Vortest nicht abgefragt, da dies die Dauer des Versuchs und damit die Belastung der Versuchspersonen über ein erträgliches Maß hinaus erhöht hätte, konnten aber im Nachtest mehrheitlich korrekt beantwortet werden. In einer Folgestudie wären diese Items bereits im Vortest abzufragen. Andere Elemente aus der Diagnostik wie die Skalen zu verschiedenen Einstellungen und Handlungsstilen könnten dafür entfallen, da die darin erhobenen Variablen keinen Zusammenhang mit der Lernleistung aufwiesen.

Hypothese 3: Dyaden profitieren vom wissenschaftlichen Problemlösen mit *virtue* mehr als Einzelpersonen. Sie verbessern sich daher vom Vor- zum Nachtest stärker als Einzelpersonen.

Daß Dyaden beim wissenschaftlichen Problemlösen mit *virtue* einen größeren Lernzuwachs verzeichnen als Einzelpersonen, konnte in der vorliegenden Untersuchung nicht bestätigt werden. Es fand sich im Hinblick darauf kein Unterschied zwischen den beiden Experimentalgruppen, weder hinsichtlich der bereichsübergreifenden methodischen Fertigkeiten noch bezüglich des bereichsspezifischen Wissens zur visuellen Suche.

In der qualitativen Analyse offenbarten sich jedoch inhaltliche, prozessurale Unterschiede zwischen Dyaden und Einzelpersonen. So wurde in den Dyaden deutlich mehr Vorwissen aktiviert und zum Gegenstand der verbalen Äußerungen gemacht. In beiden qualitativ untersuchten Dyaden wurden beispielsweise die Arbeiten von Klahr und Dunbar (1988) sowie Okada und Simon (1997) thematisiert. Da es sich um Inhalte einer Vorlesung handelt, die für den gesamten Jahrgang, aus dem die Stichprobe stammt, verbindlich war, ist davon auszugehen, daß diese Inhalte auch den Einzelpersonen bekannt waren.

In einer der Dyaden wurde im Zusammenhang mit diesem Thema auch der Bestätigungsfehler angesprochen sowie die unterschiedlichen Vorgehensweisen beim wissenschaftlichen Problemlösen (theoriegeleitet oder empirisch-explorativ). Ebenso diskutierten diese Teilnehmerinnen den Bezug ihrer simulierten Experimente zu ihren eigenen praktischen Erfahrungen aus dem Projektseminar. Bei den Dyaden fanden sich auch ausführliche Diskussionen zu Themen wie Konstanthaltung von Variablen und Signifikanzen. Alle diese Inhalte kamen bei den Einzelpersonen nicht zur Sprache. Die Aktivierung dieser Inhalte trägt zur Elaboration beim Wissenserwerb bei und sollte sich längerfristig in einem besseren Lernerfolg niederschlagen. Diese größere Vielfalt der Inhalte deckt sich mit Befunden von Dunbar (1997), daß sich verteilte Diskussionen positiv auf das wissenschaftliche Problemlösen auswirken.

Neben der größeren Fülle an Inhalten bieten die Dyaden auch ein breiteres Spektrum an Lernstrategien. In den Protokollen der Dyaden finden sich mehrere aus der Literatur bekannte Lernstrategien, welche sich beim kooperativen Lernen

als erfolgreich erwiesen haben. Generell bringen die Teilnehmerinnen in beiden Dyaden unterschiedliches Vorwissen ein und profitieren davon gegenseitig (siehe auch Dansereau, 1988; Howe, Tolmie & Rodgers, 1990, 1992). Die laut Vortest schwächere Teilnehmerin aus Dyade 1 kann auf verschiedene Weise von ihrer Partnerin dazulernen. Sie bekommt auf viele ihrer Fragen (vgl. Graesser & Person, 1994) von dieser Erklärungen auf einem ihr angemessenen Niveau (vgl. Webb, 1989) und kann diese als Modell für die Bewältigung der Aufgabe nutzen. Abstrakte Erklärungen setzt sie direkt in konkrete Schritte beim Problemlösen mit der Simulation um, teilweise verbal und teilweise, weil sie für die Eingaben in den Computer verantwortlich ist (vgl. Bielaczyc, Pirolli & Brown, 1994; Salomon & Perkins, 1998). Vergleicht man Dyade 1 und Dyade 2, so wird deutlich, daß die Teilnehmerinnen unterschiedliche Rollen einnehmen, die sich als hilfreich für Lernprozesse beim kooperativen Problemlösen erwiesen haben. Während sich in Dyade 1 vor allem die Rollen der Erklärenden und der Fragenden/Beobachtenden finden, nimmt eine der Teilnehmerinnen in Dyade 2 eine planende Rolle ein, die andere eher eine überwachende, inhaltlich reflektierende (Herrenkohl & Guerra, 1998; Kneser & Plötzner, 2001; Miyake, 1986).

Obwohl sich in Dyade 2 häufig beide Teilnehmerinnen gleichzeitig äußern und es auf diese Weise im Dialog zu Problemen bei der Mikrokoordination kommt, stimmen sich die Partnerinnen inhaltlich aufeinander ab und berücksichtigen den Standpunkt der anderen bei Entscheidungen. Diese inhaltliche Abstimmung findet auch in Dyade 1 statt, hier sogar mit wesentlich geringerem Aufwand als in Dyade 2, weil die Teilnehmerinnen dort ruhiger agieren und sich nicht so häufig gegenseitig ins Wort fallen. Und gerade die inhaltliche Koordination ist ein wichtiger Faktor für den Lernerfolg beim kooperativen Problemlösen (Howe, Tolmie, Duchak-Tanner & Rattray, 2000; Roschelle & Teasley, 1995). In beiden Dyaden sind die Entscheidungswege recht klar durch die von den Teilnehmerinnen jeweils eingenommenen Rollen bestimmt. Sie stimmen ihre Entscheidungen ab, eine Teilnehmerin ist in einer spezifischen Situation jedoch aufgrund ihrer Rolle dominant. Daneben finden sich bei den Dyaden am ehesten die beiden von Laughlin et al.

(1991) genannten Entscheidungsmodelle, bei Uneinigkeit eine gemeinsame neue Hypothese aufzustellen oder argumentativ zu demonstrieren, daß eine Hypothese der anderen überlegen sei.

Das Repertoire der beobachtbaren Lernstrategien bei den Einzelpersonen ist gegenüber den Dyaden deutlich eingeschränkt. Es erstreckt sich darauf, die Theorie zu konsultieren, sich frühere Experimente nochmals zu vergegenwärtigen, überraschende Ergebnisse zu hinterfragen und zu reflektieren sowie eine gewisse Stringenz bei der Planung der Experimentalreihe zu demonstrieren.

Vergleicht man an dieser Stelle die Ergebnisse der quantitativen und qualitativen Analysen, so fällt eine Diskrepanz ins Auge. In den quantitativen Tests konnten keine Unterschiede zwischen Dyaden und Einzelpersonen festgestellt werden. Die qualitativen Analysen offenbaren hingegen Unterschiede sowohl im Hinblick auf den Inhalt als auch auf die Lernstrategien, die für Dyaden einen höheren Lernerfolg nahelegen. Dieser Diskrepanz kann man sich mit verschiedenen Erklärungen nähern. So fanden etwa Guntermann und Tovar (1987), Krause, Stark und Mandl (2004) sowie Slavin (1995) keine Überlegenheit des kooperativen Lernens gegenüber dem Lernen alleine. Nach De Lisi und Golbeck (1999) treten die Vorteile kooperativen Lernens und Problemlösens nur bei geeigneten Aufgaben hervor, das heißt, die Aufgaben müssen hinreichend komplex sein.

Eine hinreichende Komplexität ist für das Problemlösen mit *virtue* wahrscheinlich gegeben, wobei gewisse Einschränkungen bezüglich des Gegenstandsbereichs bereits erwähnt wurden. Die Aufgaben der Vor- und Nachtests sind gegenüber der Lernumgebung *virtue* jedoch von einer geringeren Komplexität. Der Grund dafür liegt in der Handhabbarkeit der Untertests. Sie sollten alle wesentlichen Aspekte der Methodologie und des Gegenstandsbereiches erfassen und gleichzeitig nicht zu lange dauern. Die generelle Schwierigkeit bei der Messung komplexer Fertigkeiten, unabhängig davon, ob sie beim kooperativen Problemlösen erworben wurden oder nicht, betonen verschiedene Autoren (Andriessen & Sandberg, 1999; Carlsen & Andre, 1992; Njoo & de Jong, 1993; Winne & Perry, 2000). Die geringere Komplexität der Untertests bedeutet möglicherweise, daß deren Aufgaben die zusätzliche Qualität dessen, was in den Dialogen der Dyaden an Lernprozessen stattfindet, gar nicht erfassen können (siehe auch Knight & Bohl-

meyer, 1990). Zudem konnten die Untertests auch nicht das weitergehende, bei den Dyaden aktivierte Wissen aufgreifen, etwa über die Experimente von Klahr und Dunbar (1988) oder Okada und Simon (1997).

Hypothese 4: Dyaden zeigen sich in ihrem Vorgehen beim wissenschaftlichen Problemlösen Einzelpersonen überlegen. Dies schlägt sich in einer höheren Qualität der Prozeßdaten bei Dyaden verglichen mit Einzelpersonen nieder.

Für das Vorgehen beim wissenschaftlichen Problemlösen wurden in der vorliegenden Untersuchung verschiedene Indikatoren erhoben. Diese gründen sich darauf, daß auch in der Literatur recht unterschiedliche Aspekte zum wissenschaftlichen Problemlösen betrachtet werden, etwa die Suche im Hypothesen- und Experimenterraum (Klahr & Dunbar, 1988), die Interpretation von Ergebnissen im Hinblick auf Hypothesen (Tschirgi, 1980) oder der Umgang mit Analogien (Dunbar, 1995, 1997, 2000a, 2001). Es ist daher kaum möglich, ein gemeinsames, gültiges Maß für die Qualität wissenschaftlichen Problemlösens zu definieren. Stattdessen bieten die verschiedenen Indikatoren Anhaltspunkte, die in der Zusammenschau eine Einschätzung der Qualität des wissenschaftlichen Problemlösens erlauben.

Betrachtet man das Vorgehen von Dyaden und Einzelpersonen im Problemraum, so zeigen sich keine Unterschiede bezüglich Anzahl und Art der durchgeführten Experimentalzyklen. Beide Experimentalgruppen gehen hypothesengeleitet vor. Experimente ohne Hypothesen werden so gut wie gar nicht durchgeführt. Ein gutes Viertel der Versuchspersonen beider Experimentalgruppen wendet sogar zeitweilig die Strategie an, mehrere Hypothesen am Stück aufzustellen und diese dann in einer Reihe von Experimenten zu überprüfen. Diese Befunde stehen im Gegensatz etwa zur Untersuchung von Okada und Simon (1997), in der sich Gruppen von Versuchspersonen deutlich anhand des unterschiedlichen Verhältnisses von Hypothesen und Experimenten in Theoretiker und Experimentierende differenzieren ließen. Eine Erklärung dafür liegt in der Anlage der Lernumgebung *virtue*. Obwohl in der vorliegenden Untersuchung keine systematische Evaluation der in *virtue* implementierten kognitiven Werkzeuge stattfand, spricht das hypothesengeleitete Vorgehen für deren Aufforderungscharakter. Die Versuchspersonen wurden an keiner Stelle instruiert, wie sie in der Lernumgebung diesbezüglich

vorgehen sollten. Allein das Angebot des Hypothesenscratchpads und die Abfrage, ob und wenn ja welche Hypothese mit dem nächsten Versuchsplan überprüft werden soll, scheinen auszureichen, um die Versuchspersonen dazu zu veranlassen, regelmäßig Hypothesen zu formulieren und eine Reihenfolge innerhalb der experimentellen Zyklen einzuhalten, wie sie Experten wählen würden (Schraagen, 1993). Beides führt wiederum zu erwünschten Übungseffekten.

Ein- und zweifaktorielle Hypothesen und Experimente werden von Dyaden wie von Einzelpersonen komplementär behandelt. Einfaktorielle Hypothesen und Experimente finden sich vor allem für die nicht in der Merkmals-Integrations-Theorie erklärten Faktoren Farbähnlichkeit, Farbkombination und Exzentrizität. Für die beiden in der Merkmals-Integrations-Theorie zentralen Faktoren Suchtyp und Größe der Anordnung finden sich vermehrt zweifaktorielle Hypothesen und Versuchspläne. Das komplementäre Muster bei ein- und zweifaktoriellen Untersuchungen könnte darauf beruhen, daß die Wirkung der Faktoren Suchtyp und Größe der Anordnung in der Instruktion bereits theoretisch beschrieben wird, die Wirkung der restlichen Faktoren jedoch nicht. Die Versuchspersonen könnten daran interessiert gewesen sein, die Wirkung dieser Faktoren separat zu überprüfen, um leichter zu interpretierende Ergebnisse zu erhalten. Hinsichtlich der Abdeckung des Experimenterraums unterscheiden sich beide Experimentalgruppen also nicht, was sich mit Befunden von Okada und Simon (1997) deckt. Replikationen von Experimenten fanden so gut wie nicht statt. So kommt es auch, daß Reformulierungen oder Verzerrungen von Hypothesen im Sinne von Dunbar (1993) nicht beobachtet werden konnten.

Was die Systematik der Experimentalreihe anbelangt, ergaben sich für die Dyaden im Vergleich mit den Einzelpersonen Anzeichen größerer Sorgfalt. Dyaden hielten die Faktoren über mehrere Experimente hinweg konsequenter konstant als Einzelpersonen und gelangten dadurch zu besser interpretierbaren Ergebnissen. Ein bedeutender Unterschied zwischen Einzelpersonen und Dyaden bei der Durchsichtung des Problemraums offenbarte sich hinsichtlich des einzigen Experiments, mit Hilfe dessen die Merkmals-Integrations-Theorie tatsächlich überprüft werden konnte, des sogenannten kritischen Experiments (vergleiche Okada & Simon, 1997). Es wird zwar in beiden Experimentalgruppen gleich häufig durchgeführt,

bei den Dyaden geschieht dies jedoch relativ zu Beginn der Experimentalreihe, bei den Einzelpersonen eher am Ende. Dies könnte ein Hinweis darauf sein, daß Dyaden stärker theoriegeleitet vorgehen. Für Dyaden scheint die Überprüfung der Theorie in jedem Fall einen deutlich höheren Stellenwert zu haben als für Einzelpersonen. Dyaden scheinen bezüglich der Konstanthaltung von Variablen und der Interpretation der Ergebnisse Defizite kompensieren zu können, wie sie Schunn und Anderson (1999) in diesen Bereichen bei Novizen beobachtet haben.

Wie auch die qualitative Analyse ergeben hat, kann die Durchführung des kritischen Experiments, insbesondere die Kenntnis über die Wirkung des Faktors Suchtyp, den weiteren Verlauf der Experimentalreihe maßgeblich beeinflussen. So zeigte sich, daß Versuchspersonen häufig den Faktor Suchtyp auf der Stufe der Merkmalssuche konstanthielten, bevor sie detaillierte Kenntnisse über diesen Faktor hatten, und die Konstanthaltung auf die Stufe Verbindungssuche umstellten, sobald sie etwas über die Wirkung dieses Faktors erfahren hatten.

Auch beim Umgang mit Hypothesen fanden sich Unterschiede zwischen Einzelpersonen und Dyaden. Während sich die Hypothesen der Experimentalgruppen hinsichtlich der Spezifität nicht unterschieden, formulierten Dyaden ihre Hypothesen mit größerer sprachlicher Sorgfalt als Einzelpersonen. Schwierigkeiten bei der Formulierung der Hypothesen offenbarten sich auch in der qualitativen Analyse. Insbesondere eine Einzelperson, Emelie, formulierte mehrere Hypothesen über den Einfluß einer unabhängigen Variablen auf eine andere unabhängige Variable, was weder sinnvoll noch empirisch überprüfbar ist. Wahrscheinlich wollte sie damit eine Wechselwirkung ausdrücken.

In beiden Experimentalgruppen wurden etwas mehr als die Hälfte der aufgestellten Hypothesen durch die experimentellen Ergebnisse gestützt, das heißt die Vorhersagen waren richtig. Dies wurde jedoch bei der Interpretation der Hypothesen ebensowenig immer erkannt, wie die Konstellation, daß eine Hypothese durch die Daten widerlegt wurde (Bestätigungsfehler). Der Bestätigungsfehler trat deutlich häufiger auf als der umgekehrte Fall (Tschirgi, 1980). Einzelpersonen unterliefen bei der Interpretation der Ergebnisse in beiden Fällen mehr logische Fehler.

Dyaden sind also bei der Beurteilung von Hypothesen sorgfältiger und treffen weniger Fehlentscheidungen, was den Befunden von Laughlin, VanderStoep und Hollingshead (1991) entspricht.

Ein weiterer Vorteil für Dyaden beim wissenschaftlichen Problemlösen wurde in der qualitativen Analyse aufgedeckt. Dunbar (1997) betonte die Bedeutung von Analogien für erfolgreiches wissenschaftliches Problemlösen. Während beide Dyaden im Verlauf ihrer Experimentalreihe je zwei Mal auf Analogien zurückgriffen, fanden sich bei den beiden Einzelpersonen keinerlei Anhaltspunkte für den Gebrauch von Analogien.

In den Daten der vorliegenden Untersuchung gibt es verschiedene Indikatoren für eine Überlegenheit von Dyaden beim wissenschaftlichen Problemlösen gegenüber Einzelpersonen: Orientierung an der Theorie, Heranziehen von Analogien, Sorgfalt bei der Formulierung der Hypothesen, der Planung der Experimentalreihe und der Interpretation der Ergebnisse. Daß Dyaden zur Ungenauigkeit oder zum Übergehen von Informationen neigen (MacCoun & Kramer, 1996), bestätigte sich in der vorliegenden Untersuchung nicht. Auch wenn die Daten nicht für jeden Indikator höchsten statistischen Ansprüchen genügen, sprechen sie in ihrer Gesamtheit doch dafür, daß Dyaden beim wissenschaftlichen Problemlösen gegenüber Einzelpersonen Vorteile haben. Ein möglicher Grund dafür liegt in der verteilten Diskussion, wie sie Dunbar (1997) als ein Erfolgsmerkmal von Forschungsgruppen identifizierte, also das Einbringen neuer Perspektiven (vergleiche Levine, Moreland & Wingert, 1996), wie es bei den Dyaden in der qualitativen Analyse durch die reicheren Inhalte der Dialoge demonstriert wurde. Ähnliche Befunde zum kooperativen Lernen stammen von Dansereau (1988), von Howe, Tolmie und Rodgers (1990, 1992) sowie von Plötzner, Fehse, Kneser und Spada (1999). Weitere Gründe für die durch die Daten zum Vorgehen beim wissenschaftlichen Problemlösen nahegelegte Überlegenheit der Dyaden werden im Zusammenhang mit Hypothese 5 diskutiert.

Die Tatsache, daß die Versuchspersonen für mehr als die Hälfte der Experimente korrekte Vorhersagen treffen konnten und selten Revisionen ihrer Hypothesen vornahmen, deutet auf eine Problematik des Gegenstandsbereiches hin. Wie bereits in der Diskussion von Hypothese 2 angeklungen ist, ließen sich die Effekte

für einige Faktoren bei Rückgriff auf Allgemeinverständnis und nach Durchlesen der Instruktion recht gut vorhersagen. Neben komplexeren Testaufgaben hätte eine etwas höhere Komplexität des Gegenstandsbereiches eventuell ebenfalls zu differenzierteren Ergebnissen bei Dyaden und Einzelpersonen beigetragen.

Die doppelte Zielsetzung der vorliegenden Untersuchung hatte zur Wahl dieses Gegenstandsbereiches geführt. Es sollte eine reale psychologische Fragestellung sein, die den Versuchspersonen noch nicht aus dem Studium bekannt ist, aber in kurzer Zeit so zu erklären ist, daß sie sinnvolle Experimente dazu planen können. Dazu durfte die Fragestellung nicht so komplex sein, daß die Versuchspersonen dadurch überfordert sind und somit Lernprozesse unmöglich werden. Gleichzeitig sollte die Fragestellung die Beobachtung wissenschaftlichen Problemlösens erlauben, was eher einen komplexeren Gegenstandsbereich erforderte. Die visuelle Suche erschien für diese unterschiedlichen Anforderungen der geeignete Gegenstandsbereich, der auch den hohen Entwicklungsaufwand eines computerbasierten Lernsystems rechtfertigt. Als Lernsystem hat sich *virtue*, wie die Testergebnisse belegen, durchaus bewährt. Für die Beobachtung wissenschaftlichen Problemlösens wäre gegebenenfalls eine Adaptation des Gegenstandsbereiches in Richtung einer höheren Komplexität notwendig und auch für die visuelle Suche durchaus zu realisieren.

Hypothese 5: Dyaden und Einzelpersonen unterscheiden sich in ihren verbalen Äußerungen beim wissenschaftlichen Problemlösen. Bei Dyaden findet sich eine intensivere verbale Auseinandersetzung mit dem wissenschaftlichen Problemlösen sowohl auf der inhaltlichen Ebene als auch auf der Metaebene.

Die Aussagen zu Hypothese 5 beziehen sich vollständig auf die qualitative Analyse. Generell sind die Transkripte von Dyaden umfangreicher als diejenigen der Einzelpersonen. Da es sich hier um einen Vergleich der natürlichen Dialoge der Dyaden mit den Protokollen lauten Denkens der Einzelpersonen handelt, war ein solcher Unterschied zu erwarten. Ericsson und Simon (1993) gehen davon aus, daß lautes Denken bei geeigneter Instruktion die kognitiven Prozesse einer Person qualitativ nicht verändert, lediglich verlangsamt. Auch wenn dies in der vorliegenden Studie nicht explizit untersucht werden konnte, deckt sich dies mit den Eindrük-

ken, die bei der qualitativen Analyse der Transkripte von den beiden Einzelpersonen gewonnen wurden. Obwohl die Ericsson und Simon die Ähnlichkeit von Protokollen lauten Denkens mit Dialogen betonen, so ist doch offensichtlich, daß etwa die Koordination mit einer anderen Person beim lauten Denken entfällt und damit auch wichtige Indikatoren wie Kohärenz und Initiative im Dialog. Gleichzeitig verursacht allein die Koordination bereits mehr verbalen Austausch, was sich auch in den hier analysierten Transkripten niederschlägt. Da bei den Dyaden nicht substantiell mehr Anteile der Dialoge Themen abseits der Fragestellung gewidmet wurden als bei den Einzelpersonen, ist zunächst festzustellen, daß Dyaden beim wissenschaftlichen Problemlösen mit *virtue* generell mehr themenbezogene Äußerungen tun als Einzelpersonen.

Betrachtet man zunächst die inhaltliche Ebene, so zeigt sich, daß Dyaden Entscheidungen gründlicher abwägen als Einzelpersonen. Dies konnte eindeutig für die Interpretation der Ergebnisse demonstriert werden, zeigt sich aber allgemein für alle Transkripte. Für die Metaaussagen fanden sich ebenfalls Unterschiede zwischen Einzelpersonen und Dyaden. Metaaussagen gelten als Indikatoren für ein reflektiertes Vorgehen (etwa Chi et al., 1989). Dyaden äußern nicht nur absolut gesehen mehr Metaaussagen, sondern haben auch bezogen auf die relative Häufigkeit einen doppelt so hohen Anteil an Metaaussagen wie Einzelpersonen. Dieser höhere Anteil an Metaaussagen kann auch dazu beitragen, daß Dyaden überraschende Ergebnisse mit größerer Wahrscheinlichkeit bemerken und daraus die richtigen Schlußfolgerungen ziehen, was nach Dunbar (1993) einen bedeutsamen Faktor für erfolgreiches wissenschaftliches Problemlösen darstellt. Auch bei komplexen Lernprozessen wirken sich Metaaussagen positiv auf den Lernerfolg aus (Berardi-Coletta, Buyer, Dominowski & Rellinger, 1995; Danse-reau, 1988; Kneser & Plötzner, 2001).

Es handelt sich bei diesen Ergebnissen keinesfalls um ein Artefakt in dem Sinne, daß sich bei Dyaden höhere Gesprächsanteile einfach aufgrund der Tatsache finden, daß sie während ihrer Experimente mehr reden (Teasley, 1995). Vielmehr sprechen sowohl die Zahlen bezüglich der Metaaussagen als auch die qualitativen Unterschiede zwischen Einzelpersonen und Dyaden dafür, daß die verbale Ausein-

andersetzung mit dem wissenschaftlichen Problemlösen innerhalb einer jeden Dyade sowohl auf der inhaltlichen Ebene als auch bezogen auf die Metaaussagen jeweils von höherer Qualität ist als bei den Einzelpersonen.

11.2 Testdiagnostische Kennwerte

Als Begleitdiagnostik wurden Kennwerte zu wissenschaftlichem Problemlösen, Intelligenz, Handlungsstilen, Einstellungen gegenüber Gruppen und kooperativem Lernen und zur Computerliteracy erhoben. Zwischen beiden Experimentalgruppen wurden für keinen der erhobenen Kennwerte bedeutsame Unterschiede festgestellt. Auch bezüglich des Geschlechts der Versuchspersonen ergaben sich keine Unterschiede, weder was die testdiagnostischen Kennwerte anbelangt noch – im Gegensatz zu den Befunden von Koslowski (1996) – bei den Ergebnissen der Untertests zum wissenschaftlichen Problemlösen.

Statistisch bedeutsame Zusammenhänge mit den Ergebnissen der Vortests ergaben sich für die Kennwerte zum wissenschaftlichen Textverstehen, Diagrammen und Tabellen aus dem TMS (Institut für Test- und Begabungsforschung, 1995) sowie dem Raventest beim Untertest Theorie, zusätzlich für das wissenschaftliche Textverstehen auch beim Untertest Empirie, womit sich Zusammenhänge zwischen Vorwissen und Intelligenz bestätigen, wie sie Renkl und Schweizer (2000) postulieren.

Keine Zusammenhänge ergaben sich aber für die testdiagnostischen Kennwerte und den Lernzuwachs. Der Lernzuwachs der Versuchspersonen ist also unabhängig von diesen Kennwerten. Hier könnten sich die von Süß (1999) beschriebenen Meßschwierigkeiten bezüglich der Relationen zwischen Intelligenz und komplexem Problemlösen manifestieren. Selbst die Bedeutung der signifikanten Zusammenhänge für die Skalen aus dem TMS und den Raventest ist an der Vielzahl der durchgeführten statistischen Tests zu relativieren, und diese Resultate beziehen sich auch nur auf den Vortest und nicht auf den Lernzuwachs. Während also geringe Einflüsse des Vorverständnisses auf die Ausgangsleistung festgestellt werden konnten, zeigte sich keinerlei Zusammenhang zwischen den Daten der Testdiagnostik und dem gemessenen Lernzuwachs. Es könnte also angenommen werden, daß Versuchspersonen unabhängig von diesen Kennwerten vom Problemlösen profitieren. Eine andere mögliche Interpretation wäre aber auch, daß es sich

bei den Psychologiestudierenden doch um eine verhältnismäßig homogene Gruppe handelt, weshalb sich die testdiagnostischen Kennwerte nicht im Lernzuwachs niederschlagen.

11.3 Einordnung der Studie in den breiteren Forschungskontext

Die vorliegende Untersuchung ist in einem interdisziplinären Kontext zu sehen. Mit ihrer doppelten Fragestellung nach dem wissenschaftlichen Problemlösen von Einzelpersonen und Dyaden und nach dessen Unterstützung innerhalb einer Lernumgebung berührt sie das Gebiet der Kognitionswissenschaft ebenso wie die kognitive und pädagogische Psychologie sowie die Didaktik. Unter allen diesen Perspektiven stellen sich Fragen nach der Gültigkeit und Tragweite der dargestellten Schlußfolgerungen.

Bei der vorliegenden Untersuchung handelt es sich um eine Laboruntersuchung oder, wie Dunbar (1995, 2000a, 2001) es bezeichnet, um „in vitro“-Forschung im Gegensatz zu Beobachtungen realer Forschungsgruppen, der sogenannten „in vivo“-Forschung. Seine Bedenken gegenüber Laborexperimenten bezogen sich vor allem auf die ökologische Validität, sowohl diejenige der Experimentalaufgaben als auch diejenige der Stichprobe. In der vorliegenden Untersuchung wurden reale Experimente zur visuellen Suche (Reijnen, 2001) als Grundlage für die Experimente in *virtue* herangezogen. Der Datensatz wurde für die Verwendung in *virtue* nicht manipuliert, so daß die Versuchspersonen beispielsweise auch mit statistischen Abweichungen konfrontiert wurden.

Dennoch wurden verschiedene Anpassungen notwendig, zum einen um einer Laboruntersuchung wissenschaftlichen Problemlösens gerecht zu werden, zum anderen um die zweite Zielsetzung der Untersuchung umzusetzen, nämlich die der Vermittlung von Fertigkeiten für das wissenschaftliche Problemlösen. Für die Laboruntersuchung ist es notwendig, daß die Experimente in einem überschaubaren Zeitraum stattfinden können. Daher können viele Fragen der Versuchsplanung nicht realistisch abgebildet werden. So ist etwa die Zahl der variierbaren Faktoren sowie deren Operationalisierung im Voraus festgelegt. Die Formulierung von Hypothesen, das Aufstellen von Versuchsplänen und die Interpretation der Ergebnisse werden in *virtue* durch kognitive Werkzeuge unterstützt (Jonassen, 1991, Thurman, 1993, van Joolingen, 1999). Daß der Einsatz des Hypothesenscratchpads

das Verhalten beim wissenschaftlichen Problemlösen verändert, belegen die Daten von Rügsegger (2002). Auf diese Weise wurde also versucht, mit *virtue* eine was den Gegenstandsbereich anbelangt möglichst realistische Experimentalumgebung zu schaffen und gleichzeitig verschiedene didaktische Werkzeuge einzubauen.

Die vorliegende Untersuchung zielte auf das wissenschaftliche Problemlösen von Novizen ab. Hierfür fand eine Vollerhebung eines Jahrgangs von Psychologiestudierenden der Universität Basel statt. Die Stichprobe kann deshalb als repräsentativ gelten, auch wenn ein größerer Stichprobenumfang sicher wünschenswert gewesen wäre. Untersuchungen mit Stichproben dieser Größe sind in den Bereichen des wissenschaftlichen Problemlösens und auch des computerbasierten Lernens durchaus üblich (beispielsweise Klahr & Dunbar, 1988; Njoo & de Jong, 1993; Shute & Glaser, 1990). Das Vorwissen der Versuchspersonen ist als verhältnismäßig homogen einzustufen (Klahr, 2000), da die Veranstaltungen in Methodologie verpflichtend sind und für die simulierten Experimente mit der visuellen Suche ein Gegenstandsbereich ausgewählt wurde, der nicht Stoff des Curriculums bis zu diesem Semester war.

Sowohl aus sozialpsychologischer Sicht als auch aus der Perspektive der pädagogischen Psychologie ist die Zusammensetzung der Dyaden von Interesse. McGrath (1984) forderte mehr Kleingruppenforschung mit natürlichen Gruppen. Azmitia und Montgomery (1993) zeigten auf, daß befreundete Dyaden erfolgreicher lernen als einander unbekannte Personen; auf die Auswirkungen des Faktors Geschlecht beim kooperativen Lernen verweisen Barbieri und Light (1992) sowie Underwood, Underwood und Wood (2000). Auch Dunbar (1995) betonte, wie wichtig es sei, reale Forschergruppen zu untersuchen, die gemeinsam an einem Projekt arbeiteten. In der vorliegenden Untersuchung wurde diesen Forderungen nach einem realistischen Setting insofern Rechnung getragen, als sich die Versuchspersonen als Paare gleichen Geschlechts für das Experiment anmeldeten.

Die vorliegende Untersuchung benutzte *virtue* als Lernumgebung für wissenschaftliches Problemlösen. Eine Evaluation von *virtue* war nicht beabsichtigt. So konnte zwar ein Lerneffekt in beiden Experimentalgruppen nachgewiesen werden und die Lernumgebung wurde von den Versuchspersonen als positiv bewertet. Unklar bleibt aber angesichts der Komplexität der Untersuchung, worauf dieser im

Einzelnen zurückzuführen ist. Immer wieder wird in diesem Zusammenhang die Frage eines Medienvergleichs aufgeworfen: Inwiefern wäre ein Lernerfolg auch mit traditionellen didaktischen Herangehensweisen zu erzielen gewesen? Schulmeister (1997) berichtet von verschiedenen Studien älteren und jüngeren Datums, die einerseits die Schwierigkeiten bei der Untersuchung solcher Fragestellungen belegen, etwa überhaupt vergleichbare Unterrichtsmedien und Einsatzszenarien zu entwickeln und diese so zu evaluieren, daß die Ergebnisse nicht vom Hawthorne-Effekt geprägt sind. Andererseits konnte er in Meta-Analysen in den meisten Fällen keine Leistungsunterschiede zwischen Teilnehmerinnen und Teilnehmern an computervermittelten gegenüber solchen an herkömmlichen Kursen feststellen. Diese Tendenzen zeigen sich auch bei Untersuchungen, die sich direkt auf das wissenschaftliche Problemlösen beziehen (Shute & Glaser, 1990; Triona & Klahr, 2003).

11.4 Ausblick

In der vorliegenden Untersuchung wurden Daten auf verschiedenen Ebenen erhoben und ausgewertet: Vor- und Nachtests wurden ebenso durchgeführt wie eine Analyse prozeßorientierter Daten aus Logfiles und eine qualitative Analyse ausgewählter Transkripte. Eine so breite Herangehensweise findet sich in den wenigsten Untersuchungen in diesem Bereich. Fragen danach, wie diese verschiedenen Datenebenen gemeinsam zu handhaben wären und wie gegebenenfalls diskrepante Ergebnisse verschiedener Ebenen zu interpretieren seien, stellen sich dann nicht. Deshalb ist bisher auch noch zu wenig bekannt über die Zusammenhänge von Prozeßdaten und Lernerfolg. In der vorliegenden Studie wurden die verschiedenen Datenebenen über qualitative Kriterien miteinander in Bezug gesetzt, dies insbesondere, weil bei den Dyaden eine 1:1-Relation zwischen Prozeßdaten und Testdaten nicht möglich ist. Weitere offene Fragen stellen sich im Zusammenhang mit der Leistungsmessung beim komplexen Problemlösen. Hier fehlt ein systematischer Ansatz gerade für Gegenstandsbereiche, die sich nicht so leicht formalisieren lassen, wie etwa Schach- oder Physikprobleme.

Eine Untersuchung an Studierenden zum Erwerb wissenschaftlicher Fertigkeiten ist aus verschiedenen Gründen von Interesse. Die Universität ist der Ort, an dem die Studierenden ihre Fertigkeiten zum wissenschaftlichen Problemlösen ein-

setzen sollen beziehungsweise erwerben müssen, soweit sie diese noch nicht ausgebildet haben. Zum wissenschaftlichen Problemlösen wurden bereits etliche Studien mit Studierenden durchgeführt (etwa Dunbar, 1993; Klahr & Dunbar, 1988; Okada & Simon, 1997; Schunn & Anderson, 1999), kaum jedoch zum Erwerb wissenschaftlicher Fertigkeiten bei Studierenden (Njoo & de Jong, 1993; Shute & Glaser, 1990). Dennoch steht die Notwendigkeit der Vermittlung solcher Fertigkeiten gerade an der Universität außer Frage (etwa Gruber, Mandl & Renkl, 2000). Hier fehlt eine breitere wissenschaftliche Basis, um die aus dem Wissen über das Vorgehen von Novizen beim wissenschaftlichen Problemlösen abgeleiteten Interventionen zu untermauern und zu evaluieren. In diesem Zusammenhang wären auch Katamnesen und längerfristige Studien zu fordern, die zwar einen beträchtlichen Aufwand verlangen, letzten Endes jedoch notwendig sind, um die Nachhaltigkeit didaktischer Interventionen zu überprüfen. Hierzu ist im Zusammenhang mit dem Erwerb von Fertigkeiten für das wissenschaftlichen Problemlösen noch wenig bekannt.

In der vorliegenden Untersuchung konnten sowohl günstige als auch ungünstige Strategien beim kooperativen wissenschaftlichen Problemlösen und Lernen identifiziert werden. Als besonders kritische Punkte im Experimentalzyklus haben sich das Formulieren der Hypothesen und die Interpretation der Ergebnisse herausgestellt. Hier findet jeweils die meiste inhaltliche Arbeit der Studierenden statt. Aus der qualitativen Analyse ließ sich ableiten, daß Hypothesen unspezifisch formuliert werden, obwohl aus den verbalen Äußerungen deutlich wird, daß spezifische Vermutungen vorliegen. Ein Fehler bestand darin, daß Interaktionen in einer Hypothese als Auswirkung einer unabhängigen Variablen auf eine andere unabhängige Variable dargestellt wurden. Dies wären Ansatzpunkte für spezifische Übungen, die die Aussagekraft der Hypothesen und damit die Stringenz der Experimentalreihe verbessern könnten.

Bei der Interpretation der Ergebnisse fiel in der qualitativen Analyse eine Versuchsperson auf, die keine Inferenzstatistik in ihre Überlegungen einbezog. Auch den übrigen Versuchspersonen schien die Integration inferenzstatistischer Ergebnisse mit der deskriptiven Statistik und der graphischen Darstellung der Ergebnisse nicht immer leicht zu fallen. Selbst eine Diskussion über den Bestäti-

gungsfehler hat eine Dyade nicht davor bewahrt, genau dann diesen doch zu begehen. Hier tut sich eine Lücke zwischen theoretischem Wissen und dessen praktischer Umsetzung auf. Dies sind weitere Aspekte, die in Lehrveranstaltungen zur Methodologie besondere Berücksichtigung finden sollten und als Gegenstand weiterer empirischer Forschung zum Erwerb wissenschaftlicher Fertigkeiten zu sehen sind.

Ein weiterer Bereich didaktischer Fragestellungen eröffnet sich im Hinblick auf diejenigen Lernstrategien, die in einen positiven Zusammenhang mit dem Erfolg beim wissenschaftlichen Problemlösen gebracht wurden. An dieser Stelle sind eine angemessene Berücksichtigung der Theorie, die direkte Umsetzung neu erworbenen Wissens in konkrete Schritte, das Hinterfragen überraschender Ergebnisse und insbesondere Metaaussagen zu nennen. Für Dyaden wären zusätzlich das Stellen sinnvoller Fragen und die Einübung in Kohärenz und spezifische Rollen in Betracht zu ziehen. Ansätze hierfür finden sich unter anderem bei King (1997), Palincsar und Brown (1984) sowie White und Frederiksen (1998). Eine interessante Frage in diesem Zusammenhang wäre auch, inwiefern sich das Heranziehen von Analogien trainieren ließe und ob sich ein solches Training positiv auf das wissenschaftliche Problemlösen auswirken könnte.

Betrachtet man die aufgeführten Maßnahmen zur Unterstützung des wissenschaftlichen Problemlösens und Lernens, so stellt sich die Frage, inwiefern diese Maßnahmen wiederum computerunterstützt angeboten werden könnten, und ob dies gegebenenfalls sogar in *virtue* möglich wäre. Feedback zu den verschiedenen Abschnitten im Experimentalzyklus sowie zur Experimentalreihe wäre an dieser Stelle sicherlich als erstes zu nennen. Bisher bekommen Studierende von *virtue* intrinsisches Feedback (vgl. Laurillard, 1987). Eine positive Wirkung von explizitem Feedback konnten unter anderem Krause, Stark und Mandl (2004) nachweisen. Vorschläge für eine Umsetzung in einer simulierten Umgebung für entdeckendes Lernen stammen von Veermans und van Joolingen (1998). Die in diesen Arbeiten vorgeschlagenen Feedbackmechanismen basieren allerdings auf einer starken Formalisierung des Gegenstandsbereichs. Eine Rückmeldung über

die Qualität frei formulierter Hypothesen wäre mit diesen Mechanismen kaum möglich. Geeignet erscheint für diesen Zweck eher der Ansatz der *Latent Semantic Analysis* (Landauer & Dumais, 1997; Miller, 2004).

Mit *virtue* wurde eine internet-basierte Lernumgebung entwickelt. In der vorliegenden Studie wurde *virtue* jedoch im Sinne einer klassischen Experimentalumgebung genutzt. Ziel der internet-basierten Entwicklung von *virtue* war, die Lernumgebung für eine flexible Nutzung verfügbar zu machen. So kann *virtue* in verschiedenen Szenarien genutzt werden, etwa von zu Hause oder im Rahmen von Lehrveranstaltungen. Diese Flexibilität eröffnet ein Spektrum neuer Fragestellungen, wie sie auch von Sassenberg, Boos, Postmes und Reips (2003) angesprochen werden. Wie gehen Lernende mit unterschiedlichem Vorwissen und mit unterschiedlichen Voraussetzungen mit dieser Lernumgebung um? Wird eine solche Lernumgebung ohne enge Einbettung ins Curriculum angenommen? Inwiefern unterscheidet sich das Nutzerverhalten in verschiedenen Einsatzszenarien? Welche Schlußfolgerungen lassen sich daraus für internet-basierte Lehrveranstaltungen ziehen?

Eine zentrale Frage bei der Entwicklung computer-basierter Lernsysteme betrifft deren Einbettung in das Curriculum (Bachmann, Haefeli & Kindt, 2002). Welche Einsatzszenarien für ein simuliertes psychologisches Labor wie *virtue* sind denkbar? An den Universitäten Basel und Chemnitz wurden bereits verschiedene Szenarien ausprobiert. Bewährt hat sich der Einsatz von *virtue* im Experimentalpraktikum oder Projektseminar. Studierende übten Versuchsplanung und Auswertung und bekamen Anregungen für ihre eigenen Experimente. In einer deskriptiven Auswertung von Fragebögen zeigten die Studierenden eine gesteigerte Motivation für die Durchführung eigener Untersuchungen (A. Brunstein, persönliche Mitteilung, Februar 2003). Eine Erweiterung von *virtue* um andere Gegenstandsbereiche wäre in diesem Zusammenhang sicherlich wünschenswert. Aufgrund der modularen Implementation des Lernsystems ist diese auch technisch ohne größeren Aufwand zu realisieren. Auch im Rahmen einer Vorlesung zur Methodologie wurde *virtue* bereits eingesetzt. Anhand von Leitfragen erarbeiteten die Studierenden mit Hilfe von *virtue* Themen wie interne und externe Validität, Kontrolle von Störva-

riablen, die operationale Definition der Variablen und die Objektivität der simulierten Experimente. Eine Evaluation der dargestellten Einsatzszenarien von *virtue* steht noch aus. Die bisherigen Eindrücke sind jedoch vielversprechend.

Literatur

- Alessi, S. (2000). Building versus using simulations. In J. M. Spector & T. M. Anderson (Eds.), *Integrated and holistic perspectives on learning, instruction and technology* (pp. 175-196). Dordrecht, The Netherlands: Kluwer.
- Anderson, J. R., Greeno, J. G., Reder, L. M. & Simon, H. A. (2000). Perspectives on learning, thinking and activity. *Educational Researcher*, 29 (4), 11-13.
- Anderson, J. R., Reder, L. M. & Simon, H. A. (1996). Situated learning and education. *Educational Researcher*, 25 (4), 5-11.
- Andriessen, J. & Sandberg, J. (1999). Where is education heading and how about AI? *International Journal of AI in Education*, 10, 130-150.
- Aronson, E., Balney, N., Stephan, C., Sikes, J. & Snapp, M. (1978). *The jigsaw classroom*. Beverley Hills, CA: Sage.
- Arrow, H., McGrath, J. E. & Berdahl, J. L. (2000). *Small groups as complex systems: Formation, coordination, development, and adaptation*. Thousand Oaks, CA: Sage.
- Ausubel, D. P., Novak, J. D. & Hanesian, H. (1981). *Psychologie des Unterrichts* (W. Vontin, Übers.) (2. Aufl., Band 2, S. 597-647). Weinheim: Beltz (Originalarbeit publiziert 1978).
- Azmitia, M. & Montgomery, R. (1993). Friendship, transactive dialogues, and the development of scientific reasoning. *Social Development*, 2, 202-221.
- Bachmann, G., Haefeli, O. & Kindt, M. (2002). Campus 2002: Die Virtuelle Hochschule in der Konsolidierungsphase. *Medien in der Wissenschaft*, 18. Münster: Waxmann.
- Baker, M. (1994). A model for negotiation in teaching-learning dialogues. *Journal of Artificial Intelligence in Education*, 5, 199-254.
- Barbieri, M. S. & Light, P. H. (1992). Interaction, gender, and performance on a computer-based problem solving task. *Learning and Instruction*, 2, 199-213.
- Barron, B. (2000). Problem solving in video-based microworlds: Collaborative and individual outcomes of high-achieving sixth-grade students. *Journal of Educational Psychology*, 92, 391-398.
- Baumgartner, P. (1999). Evaluation mediengestützten Lernens: Theorie - Logik - Modelle. In M. Kindt (Hrsg.), *Projektelevaluation in der Lehre: Multimedia an Hochschulen zeigt Profil(e)* (S. 63-99). Münster: Waxmann.

- Bechtel, W. (1988). *Philosophy of science: An overview for cognitive science*. Hillsdale, NJ: Erlbaum.
- Berardi-Coletta, B., Buyer, L. S., Dominowski, R. L. & Rellinger, E. R. (1995). Metacognition and problem solving: A process-oriented approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 205-223.
- Berry, D. C. & Broadbent, D. E. (1987). Explanation and verbalization in a computer-assisted search task. *The Quarterly Journal of Experimental Psychology*, 39A, 585-609.
- Bielaczyc, K., Pirolli, P. L. & Brown, A. L. (1994). Collaborative explanations and metacognition: Successful learning activities in the acquisition of cognitive skills. In A. Ram & K. Eiselt (Eds.), *Proceedings of the 16th Annual Conference of the Cognitive Science Society* (pp. 39-44). Hillsdale, NJ: Erlbaum.
- Blachowicz, J. (1998). *Of two minds: The nature of inquiry*. Albany, NY: State University of New York Press.
- Blaye, A., Light, P., Joiner, R. & Sheldon, S. (1991). Collaboration as a facilitator of planning and problem solving on a computer-based task. *British Journal of Developmental Psychology*, 9, 471-483.
- Bless, H., Wänke, M., Bohner, G., Fellhauer, R. & Schwarz, N. (1994). Need for Cognition: Eine Skala zur Erfassung von Engagement und Freude bei Denkaufgaben. *Zeitschrift für Sozialpsychologie*, 25, 147-154.
- Blömeke, S. (2003). Lehren und Lernen mit neuen Medien - Forschungsstand und Forschungsperspektiven. *Unterrichtswissenschaft*, 31, 57-82.
- Bodemer, D., Plötzner, R. & Feuerlein, I. (2001). VISUALSTAT. Retrieved from <http://www.psychologie.uni-freiburg.de/visualstat/>
- Boos, M., Scharpf, U. & Fisch, R. (1991). Eine Methode zur Analyse von Interaktionsprozessen beim Problemlösen und Entscheiden in Sitzungen. *Zeitschrift für Arbeits- und Organisationspsychologie*, 35, 115-121.
- Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation* (2. überarb. Aufl.) Berlin: Springer.
- Boshuizen, H. P. A. & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science*, 16, 153-184.
- Breuer, F. (1977). *Wissenschaftstheorie für Psychologen: Eine Einführung*. Münster: Aschendorff.
- Brown, A. L. & Campione, J. C. (1994). Guided discovery in a community of learners. In K. M. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 229-270). Cambridge, MA: MIT Press.

- Brown, J. S. (1985). Process versus product: A perspective on tools for communal and informal electronic learning. *Journal of Educational Computing Research, 1*, 179-201.
- Brown, J. S., Collins, A. & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18* (1), 32-42.
- Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review, 31*, 21-32.
- Bruner, J. S., Goodnow, J. J. & Austin, G. A. (1956). *A study of thinking*. New York: Wiley.
- Cacioppo, J. T. & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*, 116-131.
- Carey, S. (1985). Are children fundamentally different kinds of thinkers and learners than adults? In S. Chipman, J. Segal & R. Glaser (Eds.), *Thinking and learning skills: Vol. 2* (pp. 485-517). Hillsdale, NJ: Erlbaum.
- Carlsen, D. D. & Andre, T. (1992). Use of a microcomputer simulation and conceptual change text to overcome student preconceptions about electric circuits. *Journal of Computer-Based Instruction, 19*, 105-109.
- Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215-281). New York: Academic Press.
- Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences, 6*, 271-315.
- Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P. & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145-182.
- Chi, M. T. H., Feltovich, P. J. & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.
- Chinn, C. A. & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research, 63*, 1-49.
- Cohen, E. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64*, 1-35.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed). Orlando, FL: Academic Press.
- Collins, A. & Brown, J. S. (1988). The computer as a tool for learning through reflection. In H. Mandl & A. Lesgold (Eds.), *Learning issues for intelligent tutoring systems* (pp. 1-18). Berlin: Springer.

- Collins, A., Brown, J. S. & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. Resnick (Ed.), *Knowing, learning, and instruction* (pp. 453-494). Hillsdale, NJ: Erlbaum.
- Cornelius, C. (2001). *Gegenseitiges Verständnis in Computerkonferenzen: Voraussetzung und Folgen konversationaler Kohärenz in Entscheidungsfindungsgruppen im Medienvergleich*. Münster: Waxmann.
- Damon, W. & Phelps, E. (1989). Critical distinctions among three approaches to peer learning. *International Journal of Educational Research*, 13, 9-19.
- Dansereau, D. F. (1988). Cooperative learning strategies. In C. E. Weinstein, E. T. Goetz & P. A. Alexander (Eds.), *Learning and study strategies: Issues in assessment, instruction and evaluation* (pp. 103-120). San Diego, CA: Academic Press.
- Dantas, A., Kemm, R. & Weaver, D. (2003). Hypothesis testing as a core component of a virtual experiment. In D. Lassner & C. McNaught (Eds.), *Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications 2003, Vol. 1 (ED-MEDIA)*, (pp. 2328-2335). Norfolk, VA: Association for the Advancement of Computing in Education.
- Darden, L. (1997). Recent work in computational scientific discovery. In M. Shafto & P. Langley (Eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (pp. 161-166). Mahwah, NJ: Erlbaum.
- Davis, J. H. (1973). Group decision and social interaction: A theory of social decision schemes. *Psychological Review*, 80, 97-125.
- de Jong, T. & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, 68, 179-201.
- de Jong, T., Martin, E., Zamarro, J.-M., Esquembre, F., Swaak, J. & van Joolingen, W. R. (1999). The interaction of computer simulation and learning support: An example from the physics domain. *Journal of Research in Science Teaching*, 36, 597-615.
- De Lisi, R. & Golbeck, S. L. (1999). Implications of Piagetian theory for peer learning. In A. M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 3-37). Mahwah, NJ: Erlbaum.
- Derry, S. J. & Lajoie, S. P. (1993). A middle camp for (un)intelligent instructional computing: An introduction. In S. J. Derry & S. P. Lajoie (Eds.), *Computers as cognitive tools* (pp. 1-11). Mahwah, NJ: Erlbaum.

- Derry, S. J., Levin, R. J., Osana, H. P., Jones, M. S. & Peterson, M. (2000). Fostering students' statistical and scientific thinking: Lessons learned from an innovative college course. *American Educational Research Journal*, 37, 747-773.
- Deusinger, I. M. (1986). *Die Frankfurter Selbstkonzeptskalen (FKSN)*. Göttingen: Hogrefe.
- Dillenbourg, P., Baker, M., Blaye, A. & O'Malley, C. (1996). The Evolution of Research on Collaborative Learning. In P. Reimann & H. Spada (Eds.), *Learning in humans and machines: Towards an interdisciplinary learning science* (pp. 189-211). Oxford: Elsevier.
- Dosher, B. A. (1998). Models of visual search: Finding a face in the crowd. In D. Scarborough & S. Sternberg (Eds.), *Methods, models, and conceptual issues: An invitation to cognitive science* (pp. 455-521). Cambridge, MA: MIT Press.
- Doise, W. & Mugny, G. (1984). *The social development of the intellect*. Oxford: Pergamon.
- Dunbar, K. (1993). Concept discovery in a scientific domain. *Cognitive Science*, 17, 397-434.
- Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 365-395). Cambridge, MA: MIT Press.
- Dunbar, K. (1997). How scientists think: On-line creativity and conceptual change in science. In T. B. Ward, S. M. Smith & J. Vaid (Eds.), *Conceptual structures and processes: Emergence, discovery, and change* (pp. 461-492). Washington, DC: American Psychological Association.
- Dunbar, K. (2000a). What scientific thinking reveals about the nature of cognition. In K. Crowley, C. D. Schunn & T. Okada (Eds.), *Designing for science: Implications from everyday, classroom, and professional settings* (pp. 115-140). Hillsdale, NJ: Erlbaum.
- Dunbar, K. (2000b). How scientists think in the real world: Implications for science education. *Journal of Applied Developmental Psychology*, 21, 49-58.
- Dunbar, K. (2001). The analogical paradox: Why analogy is so easy in naturalistic settings, yet so difficult in the psychological laboratory. In D. Gentner, K. J. Holyoak & B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 313-334). Cambridge, MA: MIT Press.
- Dunbar, K. (2002). Understanding the role of cognition in science: The *Science as Category* framework. In P. Carruthers, S. Stich & M. Siegal (Eds.), *The cognitive basis of science* (pp. 154-170). Cambridge, UK: Cambridge University Press.

- Dunbar, K. & Blanchette, I. (2001). The invivo/invitro approach to cognition: The case of analogy. *Trends in Cognitive Sciences*, 5, 334-339.
- Eagly, A. H. & Karau, S. J. (1991). Gender and the emergence of leaders: A meta-analysis. *Journal of Personality and Social Psychology*, 60, 685-710.
- Elstein, A. S., Shulman, L. S. & Sprafka, S. A. (1978). Medical problem solving: An analysis of clinical reasoning. Cambridge, MA: Harvard University Press.
- Ericsson, K. A. & Simon, H. A. (1993). *Protocol analysis* (Rev. ed.). Cambridge, MA: MIT Press.
- Ericsson, K. A. & Smith, J. (Eds.). (1991). *Toward a general theory of expertise: Prospects and limits*. Cambridge, UK: Cambridge University Press.
- Evans, J. St. B. T. (2002). The influence of prior belief on scientific thinking. In P. Carruthers, S. Stich & M. Siegal (Eds.), *The cognitive basis of science* (pp. 193-210). Cambridge, UK: Cambridge University Press.
- Fantuzzo, J. W., Riggio, R. E., Conelly, S. & Dimeff, L. A. (1989). Effects of reciprocal peer tutoring on academic achievement and psychological adjustment: A component analysis. *Journal of Educational Psychology*, 81, 173-177.
- Feist, G. J. & Gorman, M. E. (1998). The psychology of science: Review and integration of a nascent discipline. *Review of General Psychology*, 2, 3-47.
- Fischer, F. (2002). Gemeinsame Wissenskonstruktion - Theoretische und methodologische Aspekte. *Psychologische Rundschau*, 53, 119-134.
- Fleer, M. (1992). Identifying teacher-child interaction which scaffolds scientific thinking in young children. *Science Education*, 76, 373-397
- Frese, M., Albrecht, K., Kreuzer, R., von Papstein, P., Prümper, J. & Schulte-Göcking, H. (1995). Handlungsstile und Leistungsverhalten: Die Rolle von Plan- und Zielorientierung in Problem- und Lernsituationen. *Zeitschrift für Arbeits- und Organisationspsychologie*, 39, 67-77.
- Frese, M., Stewart, J. & Hannover, B. (1987). Goal orientation and planfulness. Action styles as personality concepts. *Journal of Personality and Social Psychology*, 52, 1182-1194.
- Gillies, R. (2004). The effects of cooperative learning on junior high school students during small group learning. *Learning and Instruction*, 14, 197-213.
- Glaser, R., Duschl, R. A., Schulze, S. & John, J. (1995). Student's understanding of objectives and procedures of experimentation in the science classroom. *The Journal of the Learning Sciences*, 4, 131-166.

- Goodyear, P., Njoo, M., Hijne, H. & van Berkum, J. J. A. (1991). Learning processes, learner attributes and simulations. *Education and Computing*, 6, 263-304.
- Graesser, A. C. & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, 31, 104-137.
- Grasshoff, G. & May, M. (1995). Methodische Analyse wissenschaftlichen Entdeckens. *Kognitionswissenschaft*, 5, 51-67.
- Gray, W. D. (2002). Simulated task environments: The role of high-fidelity simulations, scaled worlds, synthetic environments, and laboratory tasks in basic and applied cognitive research. *Cognitive Science Quarterly*, 2, 205-227.
- Greeno, J. G. (1989). A perspective on thinking. *American Psychologist*, 44, 134-141.
- Gruber, H., Mandl, H. & Renkl, A. (2000). Was lernen wir in Schule und Hochschule: Träges Wissen? In H. Mandl & J. Gerstenmaier (Hrsg.), *Die Kluft zwischen Wissen und Handeln: Empirische und theoretische Lösungsansätze* (S. 139-156). Göttingen: Hogrefe.
- Gruber, H., Law, L.-C., Mandl, H. & Renkl, A. (1996). Situated learning and transfer. In P. Reimann & H. Spada (Eds.), *Learning in humans and machines: Towards an interdisciplinary learning science* (pp. 168-188). Oxford: Elsevier.
- Guntermann, E. & Tovar, M. (1987). Collaborative problem solving with LOGO: Effects of group size and group composition. *Journal of Educational Computing Research*, 3, 313-334.
- Hackman, J. R. (1968). Effects of task characteristics on group products. *Journal of Experimental Social Psychology*, 4, 162-187.
- Hänsgen, K.-D. & Spicher, B. (1999). *EMS Eignungstest für das Medizinstudium in der Schweiz 1999: Bericht über die Durchführung und Ergebnisse* (Berichte des Zentrums für Testentwicklung, Nr. 5). Fribourg: Zentrum für Testentwicklung und Diagnostik.
- Hatano, G. & Inagaki, K. (1991). Sharing cognition through collective comprehension activity. In L. B. Resnick, J. M. Levine & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 331-348). Washington, DC: American Psychological Association.
- Hayes, N. (Ed.). (1997). *Doing qualitative analysis in psychology*. Hove, UK: Psychology Press.
- Heller, K. A., Kratzmeier, H. & Lengfelder, A. (1998). *Matrizen-Test-Manual: Ein Handbuch mit deutschen Normen* (Band 2). Göttingen: Beltz-Test GmbH.

- Herrenkohl, L. R. & Guerra, M. (1998). Participant structures, scientific discourse, and student engagement in fourth grade. *Cognition and Instruction, 16*, 431-473.
- Hill, G. W. (1982). Group versus individual performance: Are $N + 1$ heads better than one? *Psychological Bulletin, 91*, 517-539.
- Hmelo-Silver, C. E. & Pfeffer, M. G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive Science, 28*, 127-138.
- Hogan, D. M. & Tudge, J. R. H. (1999). Implications of Vygotsky's theory for peer learning. In A. M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 39-65). Mahwah, NJ: Erlbaum.
- Hovland, C. I. (1952). „Communication analysis“ of concept learning. *Psychological Review, 59*, 461-472.
- Howe, C., Tolmie, A., Duchak-Tanner, V. & Rattray, C. (2000). Hypothesis testing in science: Group consensus and the acquisition of conceptual and procedural knowledge. *Learning and Instruction, 10*, 361-391.
- Howe, C., Tolmie, A., & Rodgers, C. (1990). Physics in the primary school: Peer interaction and the understanding of floating and sinking. *European Journal of Psychology of Education, 4*, 459-475.
- Howe, C., Tolmie, A., & Rodgers, C. (1992). The acquisition of conceptual knowledge in science by primary school children: Group interaction and the understanding of motion down an incline. *British Journal of Developmental Psychology, 10*, 113-130.
- Hron, A., Hesse, F. W., Reinhard, P. & Picard, E. (1997). Strukturierte Kooperation beim computerunterstützten kollaborativen Lernen. *Unterrichtswissenschaft, 25*, 56-69.
- Huguet, P., Charbonnier, E. & Monteil, J.-M. (1999). Productivity loss in performance groups: People who see themselves as average do not engage in social loafing. *Group Dynamics: Theory, Research, and Practice, 3*, 118-131.
- Institut für Test- und Begabungsforschung (Hrsg.). (1995). *Der neue TMS: Originalversion des Tests für medizinische Studiengänge im besonderen Auswahlverfahren*. Göttingen: Hogrefe.
- Johnson, D. W. & Johnson, R. T. (1994). Learning Together. In S. Sharan (Ed.), *Handbook of cooperative learning methods* (pp. 51-65). Westport, CT: Greenwood Press.
- Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D. & Skon, L. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin, 89*, 47-62.

- Jonassen, D. H. (1991). What are cognitive tools? In P. Kommers, D. H. Jonassen & T. Mayes (Eds.). *Cognitive tools for learning* (pp. 1-6). Heidelberg: Springer.
- Kaartinen, S. & Kumpulainen, K. (2002). Collaborative inquiry and the construction of explanations in the learning of science. *Learning and Instruction, 12*, 189-212.
- Karmiloff-Smith, A. (1988). The child is a theoretician, not an inductivist. *Mind and Language, 3*, 183-195.
- Kerr, N. L., MacCoun, R. J. & Kramer, G. P. (1996). „When are N heads better (or worse) than one?“. Biased judgement in individuals versus groups. In E. H. Witte & J. H. Davis (Eds.), *Understanding group behavior: Vol. 1. Consensual action by small groups* (pp. 105-136). Mahwah, NJ: Erlbaum.
- Kerres, M. (2001). *Multimediale und telemediale Lernumgebungen: Konzeption und Entwicklung* (2. überarb. Aufl.). München: Oldenbourg.
- King, A. (1997). ASK to THINK-TEL WHY: A model of transactive peer tutoring for scaffolding higher level complex learning. *Educational Psychologist, 32*, 221-235.
- Klahr, D. (Ed.). (2000). *Exploring science: The cognition and development of discovery processes*. Cambridge, MA: MIT Press.
- Klahr, D. & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*, 1-48.
- Klahr, D., Dunbar, K. & Fay, A. L. (2000). Developmental aspects of scientific reasoning. In D. Klahr (Ed.), *Exploring science: The cognition and development of discovery processes* (pp. 83-131). Cambridge, MA: MIT Press.
- Klahr, D. & Simon, H. A. (1999). Studies of scientific discovery: Complementary approaches and convergent findings. *Psychological Bulletin, 125*, 524-543.
- Klayman, J. & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211-228.
- Klix, F. (1971). *Information und Verhalten*. Bern: Huber.
- Kneser, C., & Plötzner, R. (2001). Collaboration on the basis of complementary domain knowledge: Observed dialogue structures and their relation to learning success. *Learning and Instruction, 11*, 53-83.
- Knight, G. P. & Bohlmeier, E. M. (1990). Cooperative learning and achievement: Methods for assessing causal mechanisms. In S. Sharan (Ed.), *Cooperative learning: Theory and research* (pp. 1-22). New York: Praeger.
- Knorr-Cetina, K. (1984). *Die Fabrikation von Erkenntnis: Zur Anthropologie der Naturwissenschaft*. Frankfurt a. M.: Suhrkamp.

- Kolodner, J. L., Gray, J. T. & Burks Fasse, B. (2003). Promoting transfer through case-based reasoning: Rituals and practices in Learning by Design™ classrooms. *Cognitive Science Quarterly*, 3, 183-232.
- Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. Cambridge, MA: MIT Press.
- Krause, U.-M., Stark, R. & Mandl, H. (2004). Förderung des computer-basierten Wissenserwerbs durch kooperatives Lernen und eine Feedbackmaßnahme. *Zeitschrift für Pädagogische Psychologie*, 18, 125-136.
- Kraut, R. E., Egidio, C. & Galegher, J. (1990). Patterns of contact and communication in scientific research collaboration. In J. Galegher, R. E. Kraut & C. Egidio (Eds.), *Intellectual teamwork: Social and technological foundations of cooperative work* (pp. 149-171). Hillsdale, NJ: Erlbaum.
- Kratzmeier, H. & Horn, R. (1980). *Raven-Matrizen-Test: Advanced Progressive Matrices* (Deutsche Bearbeitung). Weinheim: Beltz Test Gesellschaft.
- Kriz, J., Lück, H. E., & Heidbrink, H. (1990). *Wissenschafts- und Erkenntnistheorie: Eine Einführung für Psychologen und Humanwissenschaftler* (2. durchgesehene Aufl.). Opladen: Leske + Budrich.
- Kubinger, K. D. & Farkas, M. G. (1991). Die Brauchbarkeit der Normen von Papier-Bleistift-Tests für die Computer-Vorgabe: Ein Experiment am Beispiel der SPM von Raven als kritischer Beitrag. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 12, 257-266.
- Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, 96, 674-689.
- Kuhn, D. (1991). *The Skills of Argument*. Cambridge, UK: Cambridge University Press.
- Kuhn, D., Amsel, E. & O'Loughlin, M. (1988). *The development of scientific thinking skills*. San Diego, CA: Academic Press.
- Kuhn, T. S. (1976). *Die Struktur wissenschaftlicher Revolutionen* (2. Aufl.). Frankfurt a. M.: Suhrkamp.
- Kulkarni, D. & Simon, H. A. (1988). The process of scientific discovery: The strategy of experimentation. *Cognitive Science*, 12, 139-175.
- Kumpulainen, K. & Mutanen, M. (1999). The situated dynamics of peer group interaction: An introduction to an analytic framework. *Learning and Instruction*, 9, 449-473.
- Lajoie, S. P. (1993). Computer environments as cognitive tools for enhancing learning. In S. P. Lajoie & S. J. Derry (Eds.), *Computers as cognitive tools* (pp. 261-288). Hillsdale, NJ: Erlbaum.

- Land, S. M. & Hannafin, M. J. (2000). Student-centered learning environments. In D. H. Jonassen & S. M. Land (Eds.), *Theoretical foundations of learning environments* (pp. 1-23). Mahwah, NJ: Erlbaum.
- Landauer, T. K. & Dumais, S. T. (1997). A solution to Plato's Problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211-240.
- Langley, P. (2000). The computational support of scientific discovery. *International Journal of Human-Computer Studies*, *53*, 393-410.
- Latour, B. (1986). Visualization and cognition: Thinking with eyes and hands. *Knowledge and Society: Studies in the Sociology of Culture Past and Present*, *6*, 1-40.
- Latour, B. & Woolgar, S. (1979). *Laboratory life: The social construction of scientific facts*. Princeton, NJ: Princeton University Press.
- Laughlin, P. R. (1980). Social combination processes of cooperative, problem-solving groups as verbal intellectual tasks. In M. Fishbein (Ed.), *Progress in social psychology* (Vol. 1). Hillsdale, NJ: Erlbaum.
- Laughlin, P. R., Bonner, B. L. & Altermatt, T. W. (1998). Collective versus individual induction with single versus multiple hypotheses. *Journal of Personality and Social Psychology*, *6*, 1481-1489.
- Laughlin, P. R. & Johnson, H. H. (1966). Groups and individual performance on a complementary task as a function of initial ability level. *Journal of Experimental Social Psychology*, *2*, 407-414.
- Laughlin, P. R., Kerr, N. L., Davis, J. H., Halff, H. M. & Marciniak, K. A. (1975). Group size, member ability, and social decision schemes on an intellectual task. *Journal of Personality and Social Psychology*, *31*, 522-535.
- Laughlin, P. R., Kerr, N. L., Munch, M. M. & Haggerty, C. A. (1976). Social decision schemes of the same four-person groups on two different intellectual tasks. *Journal of Personality and Social Psychology*, *33*, 80-88.
- Laughlin, P. R., Magley, V. J. & Shupe, E. I. (1997). Positive and negative hypothesis testing by cooperative groups. *Organizational Behavior and Human Decision Processes*, *69*, 265-275.
- Laughlin, P. R., VanderStoep, S. W. & Hollingshead, A. (1991). Collective versus individual induction: Recognition of truth, rejection of error, and collective information processing. *Journal of Personality and Social Psychology*, *61*, 50-67.
- Laurillard, D. (1987). Computers and the emancipation of students: Giving control to the learner. *Instructional Science*, *16*, 3-18.
- Lave, E. & Wenger, J. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.

- Leutner, D. (1993). Guided discovery learning with computer-based simulation games: Effects of adaptive and non-adaptive instructional support. *Learning and Instruction, 3*, 113-132.
- Levine, J. M. & Moreland, R. L. (1990). Progress in small group research. *Annual Review of Psychology, 41*, 585-634.
- Lewis, E. L., Stern, J. L. & Linn, M. C. (1993). The effect of computer simulations on introductory thermodynamics understanding. *Educational Technology, 33*(1), 45-58.
- MacWhinney, B., St. James, J., Schunn, C., Li, P. & Schneider, W. (2001). STEP - a system for teaching experimental psychology using E-Prime. *Behavioral Research Methods, Instruments, & Computers, 33*, 287-296.
- Manktelow, K. (1999). *Reasoning and thinking*. Hove, UK: Psychology Press.
- Masnick, A. M. & Klahr, D. (2003). Error matters: An initial exploration of elementary school children's understanding of experimental error. *Journal of Cognition and Development, 4*, 67-98.
- Mayer, R. E. (2001). *Multi-media learning*. Cambridge, UK: Cambridge University Press.
- Mayes, T. (1992). Cognitive tools: A suitable case for learning. In P. Kommers, D. H. Jonassen & T. Mayes (Eds.). *Cognitive tools for learning* (pp. 7-18). Heidelberg: Springer.
- McGrath, J. E. (1984). *Groups: Interaction and performance*. Englewood Cliffs, NJ: Prentice Hall.
- Meloth, M. S. & Deering, P. D. (1994). Task talk and task awareness under different cooperative learning conditions. *American Educational Research Journal, 31*, 138-165.
- Mercer, N., Wegerif, R. & Dawes, L. (1999). Children's talk and the development of reasoning in the classroom. *British Educational Research Journal, 25*, 95-111.
- Métraiiller, Y. (2002). *Kooperatives Problemlösen: Dyaden im Vergleich mit Einzelpersonen*. Unveröffentlichte Lizentiatsarbeit, Universität Basel, Basel.
- Meyer, T. N., Miller, T. M., Steuck, K. & Kretschmer, M. (1999). A multi-year large-scale field study of a learner controlled intelligent tutoring system. In S. P. Lajoie & M. Vivet (Eds.), *Proceedings of the Ninth International Conference on Artificial Intelligence in Education* (pp. 191-198). Amsterdam: IOS Press.
- Miller, T. (2004) Essay assessment with latent semantic analysis. *Journal of Educational Computing Research, 29*, 495-512.

- Mitroff, I. I. (1974). *The subjective side of science: A philosophical inquiry into the psychology of the Apollo moon scientists*. Amsterdam: Elsevier.
- Miyake, N. (1986). Constructive interaction and the iterative process of understanding. *Cognitive Science*, 10, 151-177.
- Moreland, R. L., Levine, J. M. & Wingert, M. L. (1996). Creating the ideal group: Composition effects at work. In E. H. Witte & J. H. Davis (Eds.), *Understanding group behavior: Vol. 2. Small group processes and interpersonal relations* (pp. 11-35). Mahwah, NJ: Erlbaum.
- Moshmann, D. & Geil, M. (1998). Collaborative reasoning: Evidence for collective rationality. *Thinking and Reasoning*, 4, 231-248.
- Mynatt, C. R., Doherty, M. E. & Tweney, T. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, 29, 85-95.
- Nägele, C., Gurtner, A., Tschan, F. & Semmer, N. (2001). *Dokumentation aller Fragebogen im Projekt „Individual and group reflexivity, and its influence on shared mental models and performance in computer mediated groups“* (Arbeitspapier 2001-01). Neuchâtel: Université de Neuchâtel, Groupe de Psychologie Appliquée.
- Neubauer, A. C., Urban, E. & Malle, B. F. (1995). Raven's Advanced Progressive Matrices: Computerunterstützte Präsentation versus Standardvorgabe. *Diagnostica*, 37, 204-212.
- Neuweg, H. G. (1999). *Könnerschaft und implizites Wissen: Zur lehrerlerntheoretischen Bedeutung der Erkenntnis- und Wissenstheorie Michael Polanyis*. Münster: Waxmann.
- Newell, A. & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Niegemann, H. M. (2001). *Neue Lernmedien konzipieren, entwickeln, einsetzen*. Bern: Huber.
- Njoo, M. & de Jong, T. (1993). Exploratory learning with a computer simulation for control theory: Learning processes and instructional support. *Journal of Research in Science Teaching*, 30, 821-844.
- Okada, T. & Simon, H. A. (1997). Collaborative discovery in a scientific domain. *Cognitive Science*, 21, 109-146.
- Opwis, K. (1992). *Kognitive Modellierung: Zur Verwendung wissensbasierter Systeme in der psychologischen Theoriebildung*. Bern: Huber.
- Palincsar, A. S. & Brown, A. L. (1984). Reciprocal teaching of comprehension fostering and comprehension monitoring activities. *Cognition and Instruction*, 1, 117-175.

- Papert, S. (1980). *Mindstorms: Children, computers and powerful ideas*. New York: Basic Books.
- Patel, V. L. & Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science*, 10, 91-116.
- Pea, R. D. (1985). Beyond amplification: Using the computer to reorganize mental functioning. *Educational Psychologist*, 20, 167-182.
- Pea, R. D. (1993). Practices of distributed intelligence and designs for education. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations* (pp. 47-87). Cambridge, MA: Cambridge University Press.
- Piaget, J. (1976). Écrits sociologiques. *Revue Européenne des Sciences Sociales*, 38-39, 44-197.
- Pilkington, R. M. & Parker-Jones, C. H. (1996). Interacting with computer-based simulation: The role of dialogue. *Computers and Education*, 27, 1-14.
- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451-502). San Diego, CA: Academic Press.
- Plötzner, R., Fehse, E., Kneser, C. & Spada, H. (1999). Learning to relate qualitative and quantitative problem representations in a model-based setting for collaborative problem solving. *The Journal of the Learning Sciences*, 8, 177-214.
- Poletiek, F. H. (2001). *Hypothesis-testing behaviour*. Hove, UK: Psychology Press.
- Polson, M. C. & Richardson, J. J. (Eds.). (1988). *Foundations of intelligent tutoring systems*. Hillsdale, NJ: Erlbaum.
- Posner, M. I. & Raichle, M. E. (1996). *Bilder des Geistes: Hirnforscher auf den Spuren des Denkens* (M. Mauch, Übers.). Heidelberg: Spektrum Akademischer Verlag.
- Reigeluth, C. M. & Schwartz, E. (1989). An instructional theory for the design of computer-based simulations. *Journal of Computer-Based Instruction*, 16, 1-10.
- Reijnen, E. (2001). *Visuelle Suche*. Unveröffentlichte Lizentiatsarbeit, Universität Basel, Basel.
- Reimann, P. (1991). Detecting functional relations in a computerized discovery environment. *Learning and Instruction*, 1, 45-64.
- Reips, U.-D. & Neuhaus, C. (2002). WEXTOR: A web-based tool for generating and visualizing experimental designs and procedures. *Behavior Research Methods, Instruments, & Computers*, 34, 234-240.

- Renkl, A. & Schweizer, K. (2000). Wer lernt wieviel? Zur Bedeutung von Intelligenzunterschieden. In K. Schweizer (Hrsg.), *Intelligenz und Kognition: Die kognitiv-biologische Perspektive der Intelligenz* (S. 85-104). Landau: Verlag Empirische Pädagogik.
- Reusser, K. (1993). Tutoring systems and pedagogical theory: Representational tools for understanding, planning, and reflection in problem solving. In S. P. Lajoie & S. J. Derry (Eds.), *Computers as cognitive tools* (pp. 143-177). Mahwah, NJ: Erlbaum.
- Rivers, R. H. & Vockell, E. (1987). Computer simulations to stimulate scientific problem solving. *Journal of Research in Science Teaching*, 24, 403-415.
- Rogers, Y. & Ellis, J. (1994). Distributed cognition: An alternative framework for analysing and explaining collaborative working. *Journal of Information Technology*, 9, 119-128.
- Rogoff, B. (1991). Social interaction as apprenticeship in thinking: Guidance and participation in spatial planning. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 349-364). Washington, DC: American Psychological Association.
- Roschelle, J. (1992). Learning by collaborating: Convergent conceptual change. *The Journal of the Learning Sciences*, 2, 235-276.
- Roschelle, J. & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *Computer supported collaborative learning* (pp. 69-97). Berlin: Springer.
- Rüegsegger, O. (2002). *Kognitive Werkzeuge beim internetbasierten wissenschaftlichen Problemlösen: Hypothesenunterstützung bei „virtue“*. Unveröffentlichte Lizentiatsarbeit, Universität Basel, Basel.
- Ruffman, T., Perner, J., Olson, D. R. & Doherty, M. (1993). Reflecting on scientific thinking: Children's understanding of the hypothesis-evidence relation. *Child Development*, 64, 1617-1636.
- Salomon, G. (1988). AI in reverse: Computer tools that turn cognitive. *Journal of Educational Computing Research*, 4, 123-139.
- Salomon, G. (1993). On the nature of pedagogic computer tools: The case of the Writing Partner. In S. P. Lajoie & S. Derry (Eds.), *Computers as cognitive tools* (pp. 179-195). Mahwah, NJ: Erlbaum.
- Salomon, G. & Perkins, D. (1998). Individual and social aspects of learning. *Review of Research in Education*, 23, 1-24.
- Sassenberg, K., Boos, M., Postmes, T. & Reips, U.-D. (2003). Studying the internet: A challenge for modern psychology. *Swiss Journal of Psychology*, 62, 75-77.

- Scardamalia, M., Bereiter, C., McLean, R. S., Swallow, J. & Woodruff, E. (1989). Computer-supported intentional learning environments. *Journal of Educational Computing Research*, 5, 51-68.
- Schauble, L., Glaser, R., Raghavan, K. & Reiner, M. (1991). Causal models and experimentation strategies in scientific reasoning. *The Journal of the Learning Sciences*, 1, 201-238.
- Schauble, L., Klopfer, L. E. & Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching*, 28, 859-882.
- Schraagen, J. M. (1993). How experts solve a novel problem in experimental design. *Cognitive Science*, 17, 285-309.
- Schrempp, I. & Sodian, B. (1999). Wissenschaftliches Denken im Grundschulalter. Die Fähigkeit zur Hypothesenprüfung und Evidenzevaluation im Kontext der Attribution von Leistungsergebnissen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 31, 67-77.
- Schulmeister, R. (1997). *Grundlagen hypermedialer Lernsysteme: Theorie, Didaktik, Design* (2. Aufl.). München: Oldenbourg.
- Schunn, C. D. & Anderson, J. R. (1999). The generality/specificity of expertise in scientific reasoning. *Cognitive Science*, 23, 337-370.
- Schunn, C. D. & Klahr, D. (2000). Multiple-space search in a more complex discovery microworld. In D. Klahr (Ed.), *Exploring science: The cognition and development of discovery processes* (pp. 161-199). Cambridge, MA: MIT Press.
- Searle, J. R. (1971). What is a speech act? In J. R. Searle (Ed.), *The philosophy of language*. Oxford: Oxford University Press.
- Shapiro, A. M. (2004). How including prior knowledge as a subject variable may change outcomes of learning research. *American Educational Research Journal*, 41, 159-189.
- Shrager, J. & Langley, P. (1990). Computational approaches to scientific discovery. In J. Shrager & P. Langley (Eds.), *Computational models of scientific discovery and theory formation* (pp. 1-25). San Mateo, CA: Morgan Kaufmann.
- Shute, V. J. (1993). A comparison of learning environments: All that glitters... In S. P. Lajoie & S. J. Derry (Eds.), *Computers as cognitive tools* (pp. 1-11). Mahwah, NJ: Erlbaum.
- Shute, V. S. & Glaser, R. (1990). A large-scale evaluation of an intelligent discovery world: Smithtown. *Interactive Learning Environments*, 1, 51-77.

- Shute, V. J., Glaser, R. & Raghavan, K. (1989). Inference and discovery in an exploratory laboratory. In P. L. Ackerman, R. J. Sternberg & R. Glaser (Eds.), *Learning and individual differences: Advances in theory and research* (pp. 279-326). New York: Freeman.
- Shute, V. J. & Towle, B. (2003). Adaptive E-learning. *Educational Psychologist*, 38, 105-114.
- Simon, H. A. (1966). Scientific discovery and the psychology of problem solving. In R. Colodny (Ed.), *Mind and cosmos* (pp. 22-40). Pittsburgh, PA: University of Pittsburgh Press.
- Simon, H. A., Langley, P. W. & Bradshaw, G. L. (1981). Scientific discovery as problem solving. *Synthese*, 47, 1-27.
- Sklar, L. (1995). Philosophy of science. In R. Audi (Ed.), *The Cambridge dictionary of philosophy* (pp. 611-615). Cambridge, UK: Cambridge University Press.
- Slavin, R. E. (1995). *Cooperative learning: Theory, research and practice* (2nd ed.). Boston, MA: Allyn & Bacon.
- Sodian, B., Zaitchik, D. & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Development*, 62, 753-766.
- Stark, R. & Mandl, H. (2000). Training in empirical research methods: Analysis of problems and intervention from a motivational perspective. In J. Heckhausen (Ed.), *Motivational psychology of human development* (pp. 165-183). Amsterdam: Elsevier.
- Stasser, G. (1991). Pooling of unshared information during group discussion. In S. Worchel, W. Wood, & J. A. Jeffery (Eds.), *Group process and productivity* (pp. 48-67). Newbury Park, CA: Sage Publications.
- Stasser, G., Taylor, L. A. & Hanna, C. (1989). Information sampling in structured and unstructured discussions of three- and six-person groups. *Journal of Personality and Social Psychology*, 57, 67-78.
- StatSoft, Inc. (2003). Electronic Statistics Textbook. Tulsa, OK: StatSoft. Retrieved from <http://www.statsoft.com/textbook/stathome.html>.
- Süß, H.-M. (1999). Intelligenz und komplexes Problemlösen: Perspektiven für eine Kooperation zwischen differentiell-psychometrischer und kognitionspsychologischer Forschung. *Psychologische Rundschau*, 59, 220-228.
- Suter, W. N. & Lindgren, H. C. (1989). *Experimentation in psychology: A guided tour*. Boston: Allyn & Bacon.
- Swaak, J. & de Jong, T. (2001). Discovery simulations and the assessment of intuitive knowledge. *Journal of Computer Assisted Learning*, 17, 284-294.

- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12, 257-285.
- Teasley, S. D. (1995). The role of talk in children's peer collaborations. *Developmental Psychology*, 2, 207-220.
- Thagard, P. (1997). Collaborative knowledge. *Noûs*, 31, 242-261.
- Thagard, P. (1998a). Ulcers and bacteria I: Discovery and acceptance. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biology and Biomedical Sciences*, 29, 107-136.
- Thagard, P. (1998b). Ulcers and bacteria II: Instruments, experiments, and social interactions. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biology and Biomedical Sciences*, 29, 317-342.
- Thurman, R. A. (1993). Instructional simulation from a cognitive psychology viewpoint. *Educational Technology Research and Development*, 41(4), 75-89.
- Treisman, A. M. & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12, 97-136.
- Triona, L. M. & Klahr, D. (2003). Point and click or grab and heft: Comparing the influence of physical and virtual instructional materials on elementary school student's ability to design experiments. *Cognition and Instruction*, 21, 149-173.
- Tschan, F. (2000). *Produktivität in Kleingruppen: Was machen produktive Gruppen anders und besser?* Bern: Huber.
- Tschan, F. (2002). Ideal cycles of communication (or cognitions) in triads, dyads, and individuals. *Small Group Research*, 33, 615-643.
- Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development*, 51, 1-10.
- Underwood, J., Underwood, G. & Wood, D. (2000). When does gender matter? Interactions during computer-based problem solving. *Learning and Instruction*, 10, 447-462.
- van Berkum, J. J. A., Hijne, H., de Jong, T., van Joolingen, W. R. & Njoo, M. (1991). Aspects of computer simulations in an instructional setting. *Education and Computing*, 6, 231-239.
- van Boxtel, C., van der Linden, J. & Kanselaar, G. (2000). Deep processing in a collaborative learning environment. In H. Cowie & G. van Aalsvoort (Eds.), *Social interaction in learning and instruction: The meaning of discourse for the construction of knowledge* (pp. 161-178). Amsterdam: Elsevier.
- van Joolingen, W. R. (1999). Cognitive tools for discovery learning. *International Journal of Artificial Intelligence in Education*, 10, 385-397.

- van Joolingen, W. R. & de Jong, T. (1993). Exploring a domain with a computer simulation: Traversing variable and relation space with the help of a hypothesis scratchpad. In D. M. Towne, T. de Jong & H. Spada (Eds.), *Simulation-based experiential learning* (pp. 191-206). Berlin: Springer.
- van Joolingen, W. R. & de Jong, T. (1997). An extended dual search space model of scientific discovery learning. *Instructional Science*, 25, 307-346.
- van Joolingen, W. R. & de Jong, T. (1991). Supporting hypothesis generation by learners exploring an interactive computer simulation. *Instructional Science*, 20, 389-404.
- VanLehn, K. (1989). Problem solving and skill acquisition. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 527-579). Cambridge, MA: MIT Press.
- Veermans, K. & van Joolingen, W.R. (1998). Using induction to generate feedback in simulation based discovery learning environments. In B.P. Goettl, H.M. Halff, C.L. Redfield & V.J. Shute (Eds.), *Intelligent Tutoring Systems, 4th International Conference, ITS'98* (pp. 196-205). Berlin: Springer.
- Vollmeyer, R. & Burns, B. D. (1996). Hypotheseninstruktion und Zielspezifität: Bedingungen, die das Erlernen und Kontrollieren eines komplexen Systems beeinflussen. *Zeitschrift für Experimentelle Psychologie*, 43, 657-683.
- Vorberg, D. & Blankenberger, S. (1999). Die Auswahl statistischer Tests und Maße. *Psychologische Rundschau*, 50, 157-164.
- Voss, J. F., Tyler, S. W. & Yengo, L. A. (1983). Individual differences in the solving of social science problems. In R. F. Dillon & R. R. Schmeck (Eds.), *Individual differences in cognition* (pp. 205-232). New York: Academic Press.
- Vye, N. J., Goldman, S. R., Voss, J. F., Hmelo, C., Williams, S. & Cognitive and Technology Group at Vanderbilt (1997). Complex mathematical problem solving by individuals and dyads. *Cognition and Instruction*, 15, 435-484.
- Vygotsky, L. S. (1962). *Thought and language* (E. Hammann & G. Vakar, Ed. & Trans.). Cambridge, MA: MIT Press. (Original work published 1934)
- Vygotsky, L. S. (1978). *Mind in society: The development of higher order psychological processes* (M. Cole, V. John-Steiner, S. Scribner & E. Souberman, Eds.). Cambridge, MA: Harvard University Press.
- Wallach, D. (1998). *Komplexe Regelungsprozesse: Eine kognitionswissenschaftliche Analyse*. Wiesbaden: Deutscher Universitätsverlag.
- Wason, P. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology I* (pp. 135-151). Harmondsworth, UK: Penguin.

- Webb, N. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research*, 13, 21-39.
- Wegerif, R. (1996). Using computers to help coach exploratory talk across the curriculum. *Computers and Education*, 26, 51-60.
- Wells, G. (1993). Reevaluating the IRF sequence: A proposal for the articulation of theories of activity and discourse for the analysis of teaching and learning in the classroom. *Linguistics and Education*, 5, 1-37.
- Wertsch, J. V. (1985). *Vygotsky and the social formation of mind*. Cambridge, MA: Harvard University Press.
- Westermann, R. (2000). *Wissenschaftstheorie und Experimentalmethodik: Ein Lehrbuch zur psychologischen Methodenlehre*. Göttingen: Hogrefe.
- White, B. Y. & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, 16, 3-118.
- Windschitl, M. & Andre, T. (1998). Using computer simulations to enhance conceptual change: The roles of constructivist instruction and student epistemological beliefs. *Journal of Research in Science Teaching*, 35, 145-160.
- Winne, P. H. & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531-566). San Diego, CA: Academic Press.
- Wolfe, J. M., Cave, K. R. & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model of visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 419-433.
- Zajonc, R. B. (1965). Social facilitation. *Science*, 149, 269-274.
- Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science*, 21, 179-217.

Anhang

Anhang A: Eingangsfragebogen

Code _____

Liebe Probandin, lieber Proband,

vielen Dank für Ihre Bereitschaft zur Teilnahme an unserem Experiment. Im Laufe der Untersuchung werden Sie in einem simulierten psychologischen Labor Experimente zu einer bestimmten psychologischen Fragestellung durchführen. Natürlich legen wir bei unserer Untersuchung grossen Wert auf die anonyme Behandlung Ihrer Daten. Allerdings wollen wir das, was Sie im Hauptteil unserer Untersuchung tun, auf Video aufzeichnen. Deshalb bitten wir Sie gleich zu Anfang, uns dafür Ihr Einverständnis zu geben:

Ich bin damit einverstanden, dass während des Experiments ein Video aufgezeichnet wird und diese Daten zur weiteren Auswertung herangezogen werden.

Name: _____

Ort/Datum: _____ Unterschrift: _____

Code _____

Vortest zum Experiment im virtuellen Labor

Versuchspersonendaten

Bitte tragen Sie ein:

Alter: _____ Jahre

Geschlecht: weiblich männlich

Hauptfach: _____

Nebenfächer: _____

Semesterzahl im Hauptfach: _____

Ich habe schon einmal am Test für medizinische Studiengänge teilgenommen:

ja nein

Code _____

Fragebogen 1

Bitte beantworten Sie die folgenden Fragen möglichst rasch und ohne groß darüber nachzudenken.

1. Die Aufgabe, neue Lösungen für Probleme zu finden, macht mir wirklich Spaß.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

2. Ich würde lieber eine Aufgabe lösen, die Intelligenz erfordert, schwierig und bedeutend ist, als eine Aufgabe, die zwar irgendwie wichtig ist, aber nicht viel Nachdenken erfordert.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

3. Ich setze mir eher solche Ziele, die nur mit erheblicher geistiger Anstrengung erreicht werden können.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

4. Ich würde lieber etwas tun, das wenig Denken erfordert, als etwas, das mit Sicherheit meine Denkfähigkeit herausfordert.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

5. Ich trage nicht gerne die Verantwortung für eine Situation, die sehr viel Denken erfordert.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

6. Ich versuche, Situationen vorauszuahnen und zu vermeiden, in denen die Wahrscheinlichkeit groß ist, daß ich intensiv über etwas nachdenken muß.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

7. Ich habe es gern, wenn mein Leben voller kniffliger Aufgaben ist, die ich lösen muß.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

8. Es genügt mir, einfach die Antwort zu kennen, ohne die Gründe für die Antwort eines Problems zu verstehen.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

9. Ich kann Auffassungen von Bekannten oft nicht zustimmen, habe aber Hemmungen, meine Kritik offen vorzubringen.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

10. Es fällt mir schwer, meine Meinung vor einer grösseren Gruppe zu vertreten.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

11. Es fällt mir schwer, einer Gruppe gegenüber eine gegensätzliche Auffassung zu vertreten.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

12. In einer Lerngruppe zu lernen, bringt mir nichts.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

13. Ich finde es stimulierend, mit einer anderen Person gemeinsam eine Lösung für ein Problem auszuarbeiten.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

14. Beim Lernen verlasse ich mich ganz auf mich selbst.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

15. Ich mag es nicht, wenn ich mich bei einer Arbeit mit jemand anders abstimmen muss.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

16. Die Ideen einer anderen Person irritieren mich beim Lernen mehr, als dass sie mir nutzen.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

17. Ich habe Erfahrung mit der Arbeit in einer Lerngruppe.

- Nein.
 Habe ich schon probiert, aber wieder aufgegeben.
 Ich lerne eher sporadisch auch mal mit anderen Studierenden zusammen.
 Ich lerne regelmässig mit anderen zusammen.

18. Im Allgemeinen, wenn ich etwas tue . . .

	trifft genau auf mich zu	trifft etwas auf mich zu	ich bin in der Mitte	trifft etwas auf mich zu	trifft genau auf mich zu	
. . . dann mache ich es so, wie andere Leute es vorschlagen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	. . . dann mache ich das so, wie ich es meine, egal was andere Leute vorschlagen.

19. Im Allgemeinen, wenn ich etwas tue . . .

	trifft genau auf mich zu	trifft etwas auf mich zu	ich bin in der Mitte	trifft etwas auf mich zu	trifft genau auf mich zu	
. . . dann versuche ich andere Leute mit einzubeziehen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	. . . dann mache ich es alleine.

20. Im Allgemeinen, wenn ich etwas tue . . .

	trifft genau auf mich zu	trifft etwas auf mich zu	ich bin in der Mitte	trifft etwas auf mich zu	trifft genau auf mich zu	
. . . habe ich mich einmal entschieden, was ich tun werde, dann mache ich es auch sofort.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	. . . habe ich einmal entschieden, was ich tun werde, lasse ich mir mit der Umsetzung Zeit.

21. Im Allgemeinen, wenn ich etwas tue . . .

. . . nehme ich alle meine Ziele sehr ernst.	trifft genau auf mich zu <input type="checkbox"/>	trifft etwas auf mich zu <input type="checkbox"/>	ich bin in der Mitte <input type="checkbox"/>	trifft etwas auf mich zu <input type="checkbox"/>	trifft genau auf mich zu <input type="checkbox"/>	. . . meine Ziele mögen wohl wichtig sein, aber ich verliere sie manchmal aus den Augen.
--	---	---	--	---	---	--

22. Im Allgemeinen, wenn ich etwas tue . . .

. . . denke ich an das, was im Augenblick notwendig ist.	trifft genau auf mich zu <input type="checkbox"/>	trifft etwas auf mich zu <input type="checkbox"/>	ich bin in der Mitte <input type="checkbox"/>	trifft etwas auf mich zu <input type="checkbox"/>	trifft genau auf mich zu <input type="checkbox"/>	. . . denke ich über die langfristigen Folgen meines Handelns nach.
--	---	---	--	---	---	---

23. Im Allgemeinen, wenn ich etwas tue . . .

. . . finde ich es nicht notwendig, über frühere Fehler nachzudenken.	trifft genau auf mich zu <input type="checkbox"/>	trifft etwas auf mich zu <input type="checkbox"/>	ich bin in der Mitte <input type="checkbox"/>	trifft etwas auf mich zu <input type="checkbox"/>	trifft genau auf mich zu <input type="checkbox"/>	. . . finde ich es notwendig, über frühere Fehler nachzudenken.
---	---	---	--	---	---	---

24. Im Allgemeinen, wenn ich etwas tue . . .

. . . finde ich es notwendig, über verschiedene Vorgehensweisen nachzudenken, bevor ich mit dem Handeln anfangе.	trifft genau auf mich zu <input type="checkbox"/>	trifft etwas auf mich zu <input type="checkbox"/>	ich bin in der Mitte <input type="checkbox"/>	trifft etwas auf mich zu <input type="checkbox"/>	trifft genau auf mich zu <input type="checkbox"/>	. . . tue ich es auf die erstbeste Methode, die mir einfällt.
--	---	---	--	---	---	---

25. Im Allgemeinen, wenn ich etwas tue . . .

. . . habe ich keinen zweiten Plan für den Fall, dass etwas schiefeht.	trifft genau auf mich zu <input type="checkbox"/>	trifft etwas auf mich zu <input type="checkbox"/>	ich bin in der Mitte <input type="checkbox"/>	trifft etwas auf mich zu <input type="checkbox"/>	trifft genau auf mich zu <input type="checkbox"/>	. . . habe ich noch einen weiteren Plan im Hinterkopf, falls etwas schiefeht.
--	---	---	--	---	---	---

26. Im Allgemeinen, wenn ich etwas tue . . .

	trifft genau auf mich zu	trifft etwas auf mich zu	ich bin in der Mitte	trifft etwas auf mich zu	trifft genau auf mich zu	
. . . habe ich eine ungefähre Vorstellung davon, wie ich vorgehen will, aber nicht über- mäßig genaue Pläne.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	. . . mache ich sehr detaillierte Pläne.

27. Im Allgemeinen, wenn ich etwas tue . . .

	trifft genau auf mich zu	trifft etwas auf mich zu	ich bin in der Mitte	trifft etwas auf mich zu	trifft genau auf mich zu	
. . . mache ich nur Sachen, die zu dem beitragen, was ich erreichen will.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	. . . erlaube ich es mir auch, von dem, was ich erreichen will, abgelenkt zu werden.

28. Im Allgemeinen, wenn ich etwas tue . . .

	trifft genau auf mich zu	trifft etwas auf mich zu	ich bin in der Mitte	trifft etwas auf mich zu	trifft genau auf mich zu	
. . . denke ich erst daran, was ich zu tun habe, wenn ich es mache.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	. . . plane ich lange im Voraus, bevor ich etwas tun muss.

29. Im Allgemeinen, wenn ich etwas tue . . .

	trifft genau auf mich zu	trifft etwas auf mich zu	ich bin in der Mitte	trifft etwas auf mich zu	trifft genau auf mich zu	
. . . plane ich auch für Zwischenfälle, die an- dere für unwahrschein- lich halten.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	. . . zerbreche ich mir nicht den Kopf über Schwierigkeiten, die ganz selten auftreten.

30. Im Allgemeinen, wenn ich etwas tue . . .

	trifft genau auf mich zu	trifft etwas auf mich zu	ich bin in der Mitte	trifft etwas auf mich zu	trifft genau auf mich zu	
. . . kommt es manchmal vor, dass ich nicht wirk- lich alles mache, was ich mir vorgenommen habe.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	. . . tue ich absolut alles, was ich mir vorge- nommen habe.

31. Im Allgemeinen, wenn ich etwas tue . . .

	trifft genau auf mich zu	trifft etwas auf mich zu	ich bin in der Mitte	trifft etwas auf mich zu	trifft genau auf mich zu	
. . . denke ich nicht lange darüber nach, wie ich es machen will - ich mache es einfach.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	. . . denke ich eine ganze Weile darüber nach, wie ich es machen will.

32. Im Allgemeinen, wenn ich etwas tue . . .

	trifft genau auf mich zu	trifft etwas auf mich zu	ich bin in der Mitte	trifft etwas auf mich zu	trifft genau auf mich zu	
. . . verfolge ich sehr hartnäckig, was ich mir vorgenommen habe.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	. . . höre ich damit auf, wenn grössere Schwierigkeiten auftauchen.

33. Im Allgemeinen, wenn ich etwas tue . . .

	trifft genau auf mich zu	trifft etwas auf mich zu	ich bin in der Mitte	trifft etwas auf mich zu	trifft genau auf mich zu	
. . . bin ich mir nicht immer völlig klar darüber, was ich genau erreichen möchte.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	. . . weiss ich im Einzelnen, was ich erreichen möchte.

Anhang B: Instruktion visuelle Suche

einführung

Sie werden hier die visuelle Suche als Fragestellung für Ihre anschließenden Experimente im virtuellen Labor kennenlernen. Im Text werden einige englische Begriffe verwendet. Merken Sie sich diese Begriffe, denn sie finden im virtuellen Labor Verwendung. Das Verständnis des Textes wird in einem anschließenden Vortest geprüft. Auch Ihre Entlohnung als Versuchsperson wird teilweise leistungsabhängig berechnet. Lesen Sie daher besonders sorgfältig.

phänomen

Wie findet eine Person ihren Freund bei einem Fussballspiel in einer riesigen Menge von Zuschauern wieder? Wie kann man ein getarntes Tier überhaupt entdecken? In beiden Fällen soll ein bestimmter *Zielreiz* (target) in einer visuellen Anordnung gefunden werden, deren Elemente Ähnlichkeiten mit dem Zielreiz aufweisen. Diesen Prozess nennt man *visuelle Suche*.

Um den Prozess der visuellen Suche zu erforschen, zeigte man den Versuchspersonen *Displays*, die eine variable Menge an Objekten (z.B. Rechtecke) enthielten (siehe Abbildung 1). Alle Objekte eines Displays, die vom Zielreiz unterschiedlich sind, nennt man *Distraktoren*. Bei den Experimenten enthielt die eine Hälfte der Displays einen Zielreiz, die andere Hälfte nicht. Eine Versuchsperson bekam im Experiment dann viele Displays hintereinander in zufälliger Reihenfolge dargeboten. Dabei wurde den Versuchspersonen jeweils zunächst der Zielreiz gezeigt. Dieser wurde als sogenannter Fixationspunkt benutzt, das heisst, er wurde an der Position gezeigt, die auch den Mittelpunkt des anschliessend gezeigten Displays kennzeichnet, so dass die Versuchspersonen sich vor der visuellen Suche immer auf denselben Punkt konzentrierten. Daraufhin wurde das Display gezeigt. Die Aufgabe der Versuchspersonen war es, so schnell und so korrekt wie möglich zu entscheiden, ob der Zielreiz im Display vorhanden ist oder nicht. Es wurde untersucht, ob sich Veränderungen der Eigenschaften des Zielreizes und der Distraktoren auf die Reaktionszeiten auswirken.

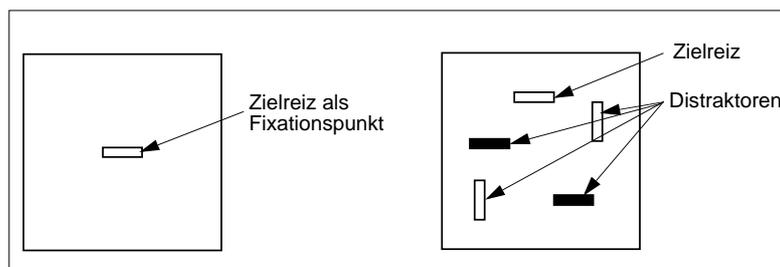


Abbildung 1: Displays für die visuelle Suche

Die wohl einflussreichste Theorie zum Thema der visuellen Suche ist die *Merkmalsintegrationstheorie*.

merkmalsintegrationstheorie

Die Merkmalsintegrationstheorie unterscheidet zwischen Elementarmerkmalen eines Objekts (z.B. Farbe, Grösse und Orientierung) und dem Objekt selbst (z.B. rotes, senkrechttes Rechteck). Nach dieser Theorie läuft die visuelle Suche in mehreren aufeinanderfolgenden Stufen ab.

1. Auf der präattentiven Stufe werden die Elementarmerkmale der Objekte auf dem Display identifiziert. Auf dieser Stufe laufen die Informationsverarbeitungsprozesse der Wahrnehmung parallel und unbewusst ab.
2. Die identifizierten Elementarmerkmale werden dann auf der Stufe der gerichteten Aufmerksamkeit zu einem Objekt zusammengesetzt. Auf dieser Stufe verläuft die Verarbeitung seriell und somit langsamer.
3. Wir nehmen ein dreidimensionales Objekt wahr (in experimentellen Settings werden häufig nur Displays mit zweidimensionalen Objekten eingesetzt).
4. Wir vergleichen dieses Objekt mit einer im Gedächtnis gespeicherten Repräsentation des Zielreizes.
5. Wenn wir dabei eine Übereinstimmung finden, identifizieren wir den Gegenstand als Zielreiz und die visuelle Suche ist damit beendet.

Die Merkmalsintegrationstheorie unterscheidet zwei Arten der Suche:

Bei einer *Merkmalsuche* (*feature search*, Abb. 2) wird nach einem Zielreiz gesucht, der sich nur bezüglich eines einzigen Elementarmerkmals wie Farbe, Orientierung u.a. von den Distraktoren unterscheidet. Im vorliegenden Beispiel unterscheiden sich der Zielreiz und die Distraktoren durch das Elementarmerkmal Farbe (pink vs. grün). Die visuelle Suche läuft in diesem Fall auf der ersten, parallel arbeitenden, präattentiven Stufe ab. Die Reaktionszeit erhöht sich mit der Zahl der Distraktoren geringfügig, wenn überhaupt.

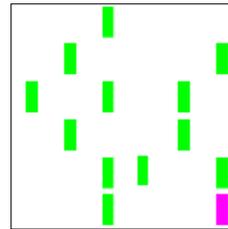


Abbildung 2: Merkmalsuche - Display mit pinkfarbenem Zielreiz.

Bei einer *Verbindungssuche* (*conjunction search*, Abb. 3) besteht der Zielreiz aus einer bestimmten Kombination oder Verbindung zweier oder mehrerer Elementarmerkmale. Diese sind gleichzeitig auch Bestandteil der anderen Elemente des Displays, wenn auch in einer anderen Kombination. Im vorliegenden Beispiel wird der Zielreiz durch eine Kombination der Elementarmerkmale Farbe und Orientierung definiert (pinkfarbene senkrechttes Rechteck). Es ergeben sich zwei Arten von Distraktoren: grüne, aufrechte und pinkfarbene, waagrechte Rechtecke. Man findet in solchen Fällen den Zielreiz nicht ohne

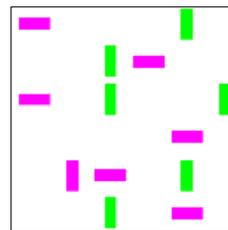


Abbildung 3: Verbindungssuche - Display mit pinkfarbenem, aufrechtem Zielreiz.

aktiv danach zu suchen oder die Aufmerksamkeit hin- und her wandern zu lassen. Hierfür ist die 2. Stufe der gerichteten Aufmerksamkeit, die seriell arbeitet, zuständig. Die Reaktionszeit nimmt deshalb mit der Anzahl Elemente mit falscher Merkmalskombination zu.

Die Reaktionszeit könnte neben den bereits dargestellten unterschiedlichen Suchprozessen auch von anderen Variablen beeinflusst werden. Einige davon werden im folgenden kurz dargestellt (Bei einigen Variablen ist auch die Farbwahrnehmung beteiligt. Für die Untersuchung werden daher jeweils nur Personen herangezogen, für die in einem Test nachgewiesen werden konnte, dass sie nicht farbenblind sind):

1. Ähnlichkeit der Farben der verschiedenen Objekte auf dem Display - ähnliche vs. unähnliche Farben (*similarity of colours*): Als unähnliche Farben gelten solche, die im Farbkreis weit auseinander liegen, z.B. Rot und Grün oder Blau und Gelb. Als ähnliche Farben gelten Farben wie Blau und Hellblau oder Rot und Pink.
2. Kombination der Farben (*colour combination*): Auch wenn die Farben der Objekte auf dem Display gleichweit voneinander entfernt sind, könnte es sein, dass die visuelle Suche bei Farbkombinationen aus einem bestimmten Farbspektrum schneller geht als bei anderen Farbkombinationen. Es könnte bei unähnlichen Farben beispielsweise sein, dass ein blauer Zielreiz auf einem Display mit blauen und gelben Objekten schneller gefunden wird als ein pinkfarbener Zielreiz auf einem Display mit pinkfarbenen und grünen Objekten. Bei ähnlichen Farben könnte man sich vorstellen, dass ein blauer Zielreiz bei blauen und hellblauen Objekten schneller gefunden wird als ein pinkfarbener Zielreiz bei pinkfarbenen und roten Objekten.
3. Anzahl der Objekte (Zielreiz und Distraktoren) auf dem Display (*set size*).
4. Exzentrizität des Zielreizes (*eccentricity*): Da sich die Versuchspersonen vor der Anzeige des Displays auf den Fixationspunkt konzentrieren, der genau die Mitte des Displays anzeigt, könnte es sein, dass die Reaktionszeit unterschiedlich ausfällt, je nach dem wo sich der Zielreiz auf dem Display befindet. In einem Experiment könnte man beispielsweise den Abstand des Zielreizes zur Displaymitte variieren in den Stufen geringer Abstand, mittlerer Abstand, grosser Abstand. Dies illustriert Abbildung 4. Ein Display wäre dann in die folgenden Zonen aufgeteilt:

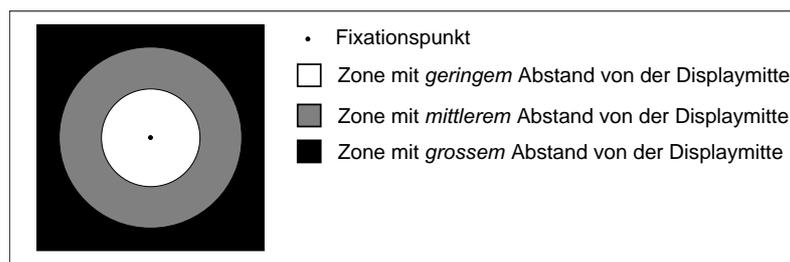


Abbildung 4: Einteilung des Displays in Zonen entsprechend dem Abstand vom Fixationspunkt.

Anhang C: Vortest

Code _____

Vortest

Bitte kreuzen Sie bei den folgenden Fragen jeweils die richtige Antwort an. Es gibt nur eine korrekte Lösung. Wenn Sie eine Frage aufgrund Ihres Wissens nicht beantworten können, lassen sie diese aus.

1. Was ist ein Zielreiz?
 - a) Der Zielreiz ist dasjenige Objekt, das angezeigt wird, wenn die visuelle Suche beendet und damit am Ziel ist.
 - b) Der Zielreiz ist dasjenige Objekt, nach dem die Versuchsperson bei der visuellen Suche suchen soll.
 - c) Mit „Zielreiz“ wird dasjenige Objekt auf dem Display bezeichnet, auf das sich die Versuchsperson vor Beginn der visuellen Suche konzentrieren soll und von dem aus sie die visuelle Suche starten soll.
 - d) Bei der visuellen Suche findet der Begriff „Zielreiz“ keine Verwendung.

2. Was ist ein Distraktor?
 - a) Als Distraktor wird jedes Objekt auf dem Display für die visuelle Suche bezeichnet, das sich vom Zielreiz in mindestens einem Merkmal unterscheidet.
 - b) Ein Distraktor ist eine Aufgabe, die die Versuchsperson als Auflockerung zwischen den einzelnen Aufgaben zur visuellen Suche bearbeiten soll.
 - c) Bei der visuellen Suche spielen Distraktoren keine Rolle.
 - d) Ein Distraktor ist ein Display ohne Zielreiz, das daher von der eigentlichen visuellen Suche ablenkt.

3. Was bedeutet der Begriff „Exzentrizität“ im Zusammenhang mit der visuellen Suche?
 - a) Mit Exzentrizität wird bezeichnet, wie ausgefallen und ungewöhnlich die Form eines Zielreizes ist. So weist ein Schmetterling als Zielreiz eine höhere Exzentrizität auf als ein Rechteck.
 - b) Der Grad der Exzentrizität drückt aus, in welchem Abstand zum Zentrum des Displays sich der Zielreiz befindet.
 - c) Die Exzentrizität besagt, ob sich die Distraktoren eher in der Mitte oder eher am Rand eines Displays befinden.
 - d) Der Begriff Exzentrizität hat für die Forschung zur visuellen Suche keine Bedeutung.

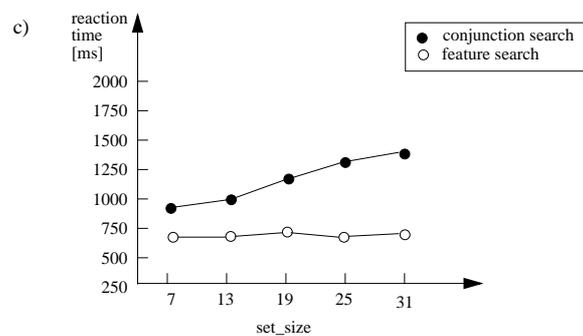
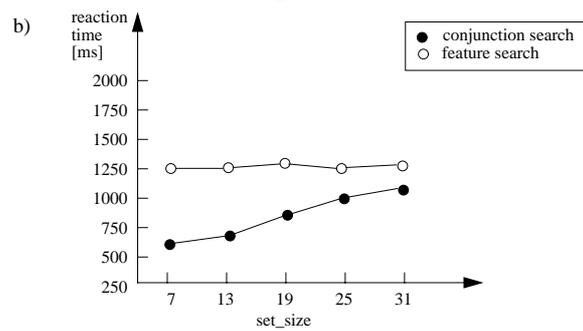
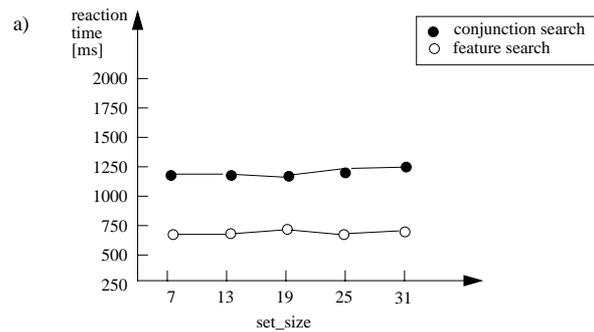
4. Welche Aussage über die Stufen der visuellen Suche ist richtig?
 - a) Auf der präattentiven Stufe findet serielle und unbewusste Informationsverarbeitung statt.
 - b) Auf der Stufe der gerichteten Aufmerksamkeit findet parallele und unbewusste Informationsverarbeitung statt.
 - c) Auf der Stufe der ungerichteten Aufmerksamkeit findet serielle und unbewusste Informationsverarbeitung statt.
 - d) Auf der präattentiven Stufe findet parallele und unbewusste Informationsverarbeitung statt.

5. Welche der folgenden Aussagen über die Merkmalsuche ist richtig?
 - a) Bei einer Merkmalsuche soll die Versuchsperson herausfinden, welches Merkmal (z.B. Farbe, Grösse, etc.) verschiedenen Objekten auf dem Display gemeinsam ist.
 - b) Bei einer Merkmalsuche wird nach einem Zielreiz gesucht, der sich nur durch ein einziges Merkmal wie Farbe, Grösse, etc. von allen anderen Objekten auf dem Display unterscheidet.
 - c) Bei einer Merkmalsuche wird nach einem Zielreiz gesucht, der sich erst durch eine bestimmte Kombination von zwei Merkmalen (z.B. Farbe und Grösse) von den anderen Objekten auf dem Display unterscheidet.
 - d) Bei einer Merkmalsuche soll die Versuchsperson nach einem Zielreiz suchen, der durch eine Kombination von Distraktoren gekennzeichnet ist.

6. Welche der folgenden Aussagen über die Verbindungssuche ist richtig?
 - a) Bei einer Verbindungssuche soll die Versuchsperson herausfinden, durch welches gemeinsame Merkmal (z.B. Farbe, Grösse, etc.) verschiedene Displays untereinander verbunden sind.
 - b) Bei einer Verbindungssuche wird nach einem Zielreiz gesucht, der sich nur durch ein einziges Merkmal wie Farbe, Grösse, etc. von allen anderen Objekten auf dem Display unterscheidet.
 - c) Bei einer Verbindungssuche wird nach einem Zielreiz gesucht, der sich erst durch eine bestimmte Kombination von zwei Merkmalen (z.B. Farbe und Grösse) von den anderen Objekten auf dem Display unterscheidet.
 - d) Bei einer Verbindungssuche soll die Versuchsperson eine Verbindung von Merkmalen (z. B. Farbe und Grösse) herausfinden, durch die sich verschiedene Typen von Displays unterscheiden.

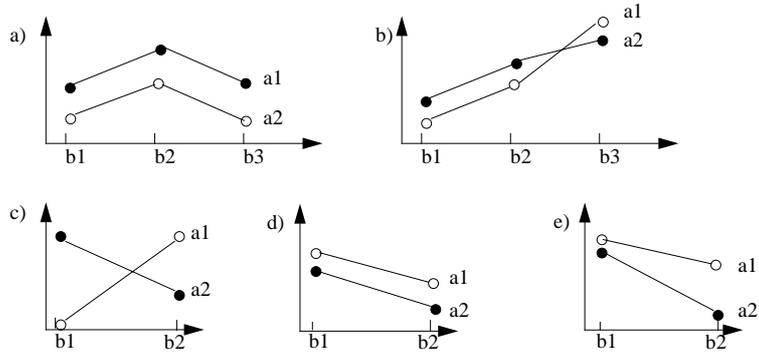
7. Welche der folgenden Aussagen über die Suchdauer von Verbindungs- und Merkmalsuche ist richtig?
- a) Bei gleicher Anzahl der Distraktoren gilt: Die Verbindungssuche dauert länger als die Merkmalsuche.
 - b) Bei gleicher Anzahl der Distraktoren gilt: Die Merkmalsuche dauert länger als die Verbindungssuche.
 - c) Bei gleicher Anzahl der Distraktoren gilt: Die Verbindungssuche dauert gleich lange wie die Merkmalsuche.
8. Welche der folgenden Aussagen ist richtig?
- a) Bei der Merkmalsuche gilt: Je mehr Distraktoren in einem Display vorkommen, desto kürzer dauert die Suche.
 - b) Bei der Merkmalsuche hängt es von der Zusammensetzung der Merkmale ab, ob die Anzahl der Distraktoren einen Einfluß auf die Suchdauer hat oder nicht.
 - c) Bei der Merkmalsuche gilt: Es kommt nicht auf die Anzahl der Distraktoren in einem Display an; die Suche dauert immer gleich lange.
 - d) Bei der Merkmalsuche gilt: Je mehr Distraktoren in einem Display vorkommen, desto länger dauert die Suche.
9. Welche der folgenden Aussagen ist richtig?
- a) Bei der Verbindungssuche gilt: Es kommt nicht auf die Anzahl der Distraktoren in einem Display an; die Suche dauert immer gleich lange.
 - b) Bei der Verbindungssuche gilt: Je mehr Distraktoren in einem Display vorkommen, desto kürzer dauert die Suche.
 - c) Bei der Verbindungssuche hängt es von der Zusammensetzung der Merkmale ab, ob die Anzahl der Distraktoren einen Einfluß auf die Suchdauer hat oder nicht.
 - d) Bei der Verbindungssuche gilt: Je mehr Distraktoren in einem Display vorkommen, desto länger dauert die Suche.

10. Welcher der folgenden Interaktionsplots gibt die Verhältnisse bei der Merkmals- und Verbindungssuche im Bezug auf die Anzahl der Elemente auf dem Display korrekt wieder?

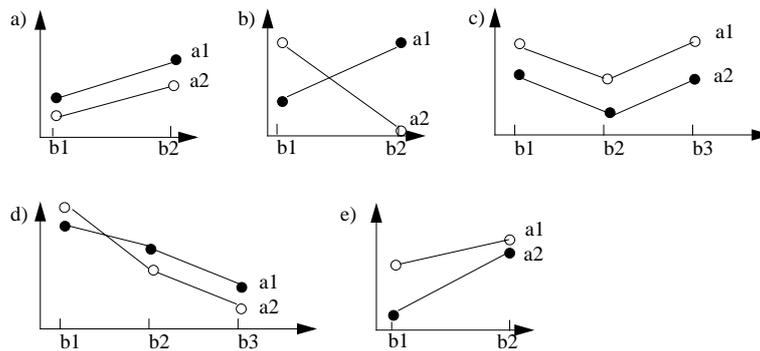


11. Geben Sie ein Beispiel für ein 2 x 2 Design.

12. Welche der abgebildeten Plots zeigen einen Haupteffekt? Hier sind mehrere richtige Antworten möglich. Kreuzen Sie dann alle korrekten Antworten an. Gehen sie bei Ihrer Entscheidung davon aus, daß ein Effekt, wenn er sich in der Graphik abzeichnet, auch signifikant ist.



13. Welche der abgebildeten Plots zeigen eine Interaktion? Hier sind mehrere richtige Antworten möglich. Kreuzen Sie dann alle korrekten Antworten an. Gehen sie bei Ihrer Entscheidung davon aus, daß ein Effekt, wenn er sich in der Graphik abzeichnet, auch signifikant ist.



14. Warum reicht es für die Interpretation experimenteller Ergebnisse nicht aus, sich Mittelwerte und Streuungen anzuschauen?
- Aus den Mittelwerten und Streuungen kann man keine kausalen Erklärungen ableiten.
 - Um die Mittelwerte und Streuungen sinnvoll interpretieren zu können, muß auch das α -Niveau angegeben sein.
 - Diese Werte allein lassen keine Aussage über die statistische Bedeutsamkeit der Ergebnisse zu.
 - Man muß auch das Konfidenzintervall bei der Interpretation beachten.
15. Was wird durch das statistische Signifikanzniveau beschrieben?
- Das Signifikanzniveau legt die Wahrscheinlichkeit fest, ab der die Alternativhypothese angenommen werden kann.
 - Das Signifikanzniveau legt die Wahrscheinlichkeit fest, dass für die erhaltenen Ergebnisse die Alternativhypothese angenommen wird, obwohl die Nullhypothese gültig ist.
 - Das Signifikanzniveau legt fest, ab welcher Grenze die Alternativhypothese als Erklärung für die gefundene Evidenz ausgeschlossen werden kann.

16. Christine und Daniela sind zwei fortgeschrittene Psychologiestudentinnen und wollen ihre Lizenziatsarbeit als Gemeinschaftsprojekt anfertigen. Sie wollen eine Untersuchung zum Thema Geschmackswahrnehmung durchführen. Aus der Literatur haben sie eine Hypothese abgeleitet, nämlich daß der wahrgenommene Grad der Süße von Lebensmitteln von der Tageszeit abhängig ist. Dieser soll auf einem Fragebogen erhoben werden. Den beiden ist nicht ganz klar, ob es auch noch andere Einflußfaktoren gibt. Für das Design haben sie allerdings unterschiedliche Vorschläge.

Christines Vorschlag sieht vor, zunächst ein Experiment durchzuführen, bei dem dieselben Versuchspersonen am selben Tag morgens und abends jeweils eine süße und eine neutrale Speise kosten und auf dem Fragebogen einschätzen sollen. Von diesem Experiment ausgehend möchte sie eventuell später weitere Experimente planen.

Daniela schlägt ein Experiment vor, bei dem die Versuchspersonen im Laufe des Tages dreimal einen Geschmackstest absolvieren sollen. Sie möchte den Geschmackstest auf andere Geschmacksrichtungen erweitern und will die Versuchspersonen neben mehreren süßen und neutralen Speisen auch salzige und bittere Geschmacksrichtungen kosten lassen. Die Reihenfolge der Geschmacksrichtungen soll randomisiert sein. Außerdem möchte sie die Versuchspersonen in zwei Gruppen einteilen. Die eine Gruppe soll vor den Mahlzeiten den Geschmackstest absolvieren, die andere Gruppe nach den Mahlzeiten. Damit soll der Blutzuckerspiegel kontrolliert werden.

Welchen Vorschlag favorisieren Sie?

Christines Vorschlag

Danielas Vorschlag

Begründen Sie Ihre Entscheidung. Nennen Sie dabei Vor- und Nachteile der beiden vorgeschlagenen Experimentaldesigns.

17. Eric hat die Daten für seine Lizentiatsarbeit endlich beisammen. Er hat ein Experiment zur Wirkung eines aufmerksamkeitssteigernden Medikaments auf die Reaktionszeit unter Schlafentzug durchgeführt. Aus früheren Studien ist bereits bekannt, dass ohne Medikamente sich die Reaktionszeit bei abnehmender Schlafdauer kontinuierlich verlängert. Das Medikament soll nun die Auswirkungen von Schlafentzug ausgleichen. Erics Hypothese war, daß das Medikament zwar die Reaktionszeit in einem Aufmerksamkeitstest verkürzt, die Kurve der mit abnehmender Schlafdauer kontinuierlich verlangsamten Reaktionszeit aber parallel zur Kurve der Reaktionszeit bei Schlafentzug ohne Medikamentengabe ansteigt. Da er als Hilfsassistent in einem Schlaflabor an einer großen Studie mitarbeitete, konnte er seine Versuchspersonen jeweils in Meßwiederholung unter allen Bedingungen untersuchen. Eric hat die folgenden Daten erhalten:

Deskriptive Daten: Reaktionszeiten beim Aufmerksamkeitstest

	mit Medikament			ohne Medikament		
	M	SD	n	M	SD	n
Schlafdauer						
5 Stunden	680.7	111.7	30	886.3	157.3	30
4 Stunden	661.7	99.3	30	967.0	177.6	30
3 Stunden	707.7	146.4	30	1187.0	214.3	30
2 Stunden	664.7	82.0	30	1224.7	254.0	30
1 Stunde	683.7	119.8	30	1394.7	249.0	30

Varianzanalyse (Meßwiederholung)

source	df	F
within subjects (S)	29	(143672.0)
between subjects		
Schlafdauer (A)	4	41.82***
mit/ohne Medikament (B)	1	534.86***
A x B	4	35.82***
A x S	116	(15904.1)
B x S	29	(28681.9)
A x B x S	116	(16926.7)

Anmerkungen: Werte innerhalb der Klammern repräsentieren die mittleren Quadratwerte.
* p<.05, ** p<.01, *** p<.001

Wie sind die Daten im Hinblick auf Eric's Hypothese zu interpretieren? Ergeben sich daraus gegebenenfalls neue Perspektiven?

Anhang D: Nachtest

Code _____

Nachtest

Bitte kreuzen Sie bei den folgenden Fragen jeweils die richtige Antwort an. Es gibt nur eine korrekte Lösung. Kreuzen Sie bei jeder Aufgabe eine Antwort an!

1. Was ist ein Zielreiz?
 - a) Ein Zielreiz ist die räumliche Anordnung verschiedener Objekte auf einem Display, die sich eine Versuchsperson bei den Aufgaben zur visuellen Suche einprägen soll.
 - b) Der Zielreiz ist dasjenige Objekt, das die Versuchsperson bei der visuellen Suche finden soll.
 - c) Bei der visuellen Suche findet der Begriff „Zielreiz“ keine Verwendung.
 - d) Der Zielreiz leuchtet auf, wenn man das richtige Objekt bei der visuellen Suche gefunden hat.

2. Welche der folgenden Aussagen über die Merkmalssuche ist richtig?
 - a) Bei einer Merkmalssuche soll die Versuchsperson herausfinden, welches Merkmal (z.B. Farbe, Grösse, etc.) verschiedenen Displays gemeinsam ist.
 - b) Bei einer Merkmalssuche wird nach einem Zielreiz gesucht, der sich nur durch ein einziges Merkmal wie Farbe, Grösse, etc. von allen anderen Objekten auf dem Display unterscheidet.
 - c) Bei einer Merkmalssuche wird nach einem Zielreiz gesucht, der sich erst durch eine bestimmte Kombination von zwei Merkmalen (z.B. Farbe und Grösse) von den anderen Objekten auf dem Display unterscheidet.
 - d) Bei einer Merkmalssuche soll die Versuchsperson herausfinden, in welchem Merkmal (z.B. Farbe, Grösse, etc.) sich verschiedene Displays unterscheiden.

3. Welche der folgenden Aussagen über die Verbindungssuche ist richtig?
 - a) Bei einer Verbindungssuche soll die Versuchsperson herausfinden, durch welches gemeinsame Merkmal (z.B. Farbe, Grösse, etc.) verschiedene Displays untereinander verbunden sind.
 - b) Bei einer Verbindungssuche wird nach einem Zielreiz gesucht, der sich nur durch ein einziges Merkmal wie Farbe, Grösse, etc. von allen anderen Objekten auf dem Display unterscheidet.
 - c) Bei einer Verbindungssuche wird nach einem Zielreiz gesucht, der sich erst durch die Kombination von zwei Merkmalen (z.B. Farbe und Grösse) von den anderen Objekten auf dem Display unterscheidet.
 - d) Bei einer Verbindungssuche soll die Versuchsperson eine Verbindung von Merkmalen (z. B. Farbe und Grösse) herausfinden, durch die sich verschiedene Typen von Displays unterscheiden.

4. Was ist ein Distraktor?
 - a) Ein Distraktor ist eine Aufgabe, die die Versuchsperson parallel zur visuellen Suche bewältigen soll und die nichts mit der visuellen Suche zu tun hat (z.B. rückwärts Zählen), sie jedoch von der Aufgabe der visuellen Suche ablenkt.
 - b) Bei der visuellen Suche spielen Distraktoren keine Rolle.
 - c) Ein Distraktor ist ein Objekt, das sich auf dem Display für die visuelle Suche bewegt und die Augen von der eigentlichen visuellen Suche ablenkt.
 - d) Als Distraktor wird jedes Objekt auf dem Display für die visuelle Suche bezeichnet, das sich in mindestens einem Merkmal vom Zielreiz unterscheidet.

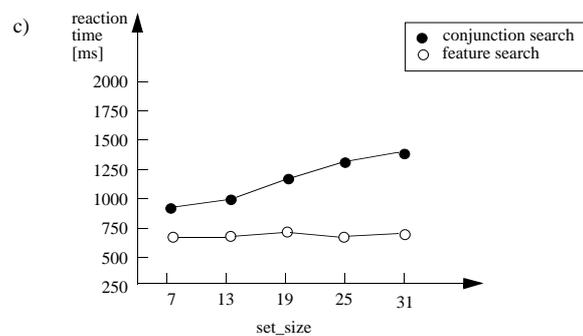
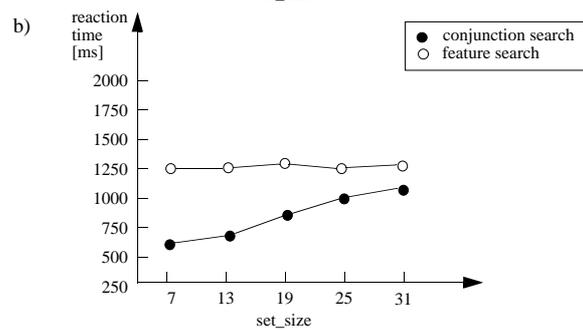
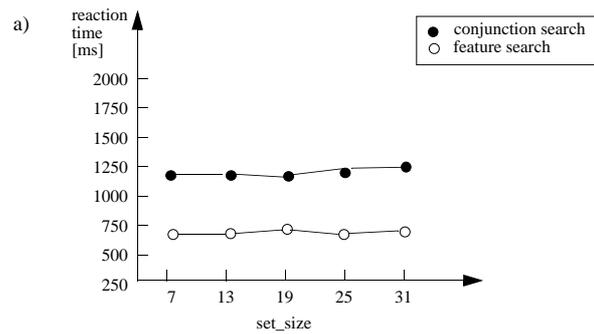
5. Welche Aussage über die Stufen der visuellen Suche ist richtig?
 - a) Auf der Stufe der gerichteten Aufmerksamkeit findet serielle und unbewusste Informationsverarbeitung statt.
 - b) Auf der Stufe der gerichteten Aufmerksamkeit findet serielle Informationsverarbeitung statt.
 - c) Auf der präattentiven Stufe findet serielle und unbewusste Informationsverarbeitung statt.
 - d) Auf der präattentiven Stufe findet parallele und bewusste Informationsverarbeitung statt.

6. Was bedeutet der Begriff „Exzentrizität“ im Zusammenhang mit der visuellen Suche?
 - a) Die Exzentrizität legt fest, ob ein Display eher am Anfang oder am Ende einer Reihe von Displays gezeigt wird oder eher in der Mitte.
 - b) Der Begriff Exzentrizität hat für die Forschung zur visuellen Suche keine Bedeutung.
 - c) Mit Exzentrizität ist gemeint, in welchem Abstand zum Zentrum des Displays sich der Zielreiz befindet.
 - d) Die Exzentrizität besagt, ob sich die Distraktoren eher in der Mitte oder eher am Rand eines Displays befinden.

7. Welche der folgenden Aussagen über die Suchdauer von Verbindungs- und Merkmalssuche ist richtig?
 - a) Wenn die Anzahl der Distraktoren gleich gross ist, dauert die Verbindungssuche gleich lange wie die Merkmalssuche.
 - b) Wenn die Anzahl der Distraktoren gleich gross ist, dauert die Verbindungssuche länger als die Merkmalssuche.
 - c) Wenn die Anzahl der Distraktoren gleich gross ist, dauert die Merkmalssuche länger als die Verbindungssuche.

8. Welche der folgenden Aussagen über die Suchdauer bei der Merkmalssuche ist richtig?
- a) Bei der Merkmalssuche gilt: Je mehr Distraktoren in einem Display vorkommen, desto länger dauert die Suche.
 - b) Bei der Merkmalssuche gilt: Je weniger Distraktoren in einem Display vorkommen, desto länger dauert die Suche.
 - c) Bei der Merkmalssuche hängt es von der Zusammensetzung der Merkmale ab, ob die Anzahl der Distraktoren einen Einfluß auf die Suchdauer hat oder nicht.
 - d) Bei der Merkmalssuche gilt: Es kommt nicht auf die Anzahl der Distraktoren in einem Display an; die Suche dauert immer gleich lange.
9. Welche der folgenden Aussagen über die Suchdauer bei der Verbindungssuche ist richtig?
- a) Bei der Verbindungssuche gilt: Je mehr Distraktoren in einem Display vorkommen, desto länger dauert die Suche.
 - b) Bei der Verbindungssuche gilt: Je weniger Distraktoren in einem Display vorkommen, desto länger dauert die Suche.
 - c) Bei der Verbindungssuche hängt es von der Zusammensetzung der Merkmale ab, ob die Anzahl der Distraktoren einen Einfluß auf die Suchdauer hat oder nicht.
 - d) Bei der Verbindungssuche gilt: Es kommt nicht auf die Anzahl der Distraktoren in einem Display an; die Suche dauert immer gleich lange.

10. Welcher der folgenden Interaktionsplots gibt die Verhältnisse bei der Merkmals- und Verbindungssuche im Bezug auf die Anzahl der Elemente auf dem Display korrekt wieder?



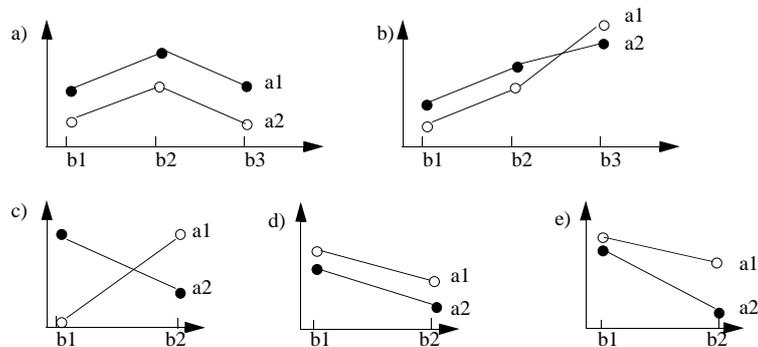
11. Welche der folgenden Aussagen ist richtig?
- a) Bei der Verbindungssuche sind die Reaktionszeiten umso länger, je ähnlicher sich der Zielreiz und die Distraktoren sind.
 - b) Bei der Verbindungssuche sind die Reaktionszeiten umso kürzer, je ähnlicher sich der Zielreiz und die Distraktoren sind.
 - c) Bei der Verbindungssuche hat die Ähnlichkeit zwischen dem Zielreiz und den Distraktoren keinen Einfluss auf die Reaktionszeiten.
12. Welche der folgenden Aussagen ist richtig?
- a) Bei der Merkmalssuche gilt: Die Suche nach einem blauen Zielreiz inmitten gelber Distraktoren dauert länger als die Suche nach einem blauen Zielreiz inmitten hellblauer Distraktoren.
 - b) Bei der Merkmalssuche gilt: Die Suche nach einem blauen Zielreiz inmitten gelber Distraktoren dauert weniger lange als die Suche nach einem blauen Zielreiz inmitten hellblauer Distraktoren.
 - c) Bei der Merkmalssuche gilt: Die Suche nach einem blauen Zielreiz inmitten gelber Distraktoren dauert gleich lange wie die Suche nach einem blauen Zielreiz inmitten hellblauer Distraktoren.
13. Welche der folgenden Aussagen ist richtig?
- a) Bei der Verbindungssuche gilt: Die Suche nach einem pinkfarbenen Zielreiz inmitten grüner und pinkfarbener Distraktoren dauert länger als die Suche nach einem pinkfarbenen Zielreiz inmitten roter und pinkfarbener Distraktoren.
 - b) Bei der Verbindungssuche gilt: Die Suche nach einem pinkfarbenen Zielreiz inmitten grüner und pinkfarbener Distraktoren dauert weniger lange als die Suche nach einem pinkfarbenen Zielreiz inmitten roter und pinkfarbener Distraktoren.
 - c) Bei der Verbindungssuche gilt: Die Suche nach einem pinkfarbenen Zielreiz inmitten grüner und pinkfarbener Distraktoren dauert gleich lange wie die Suche nach einem pinkfarbenen Zielreiz inmitten roter und pinkfarbener Distraktoren.
14. Welche der folgenden Aussagen ist richtig?
- a) Bei der Merkmalssuche gilt: Die Suche nach einem pinkfarbenen Zielreiz inmitten grüner Distraktoren dauert länger als die Suche nach einem blauen Zielreiz inmitten gelber Distraktoren.
 - b) Bei der Merkmalssuche gilt: Die Suche nach einem pinkfarbenen Zielreiz inmitten grüner Distraktoren dauert weniger lange als die Suche nach einem blauen Zielreiz inmitten gelber Distraktoren.
 - c) Bei der Merkmalssuche gilt: Die Suche nach einem pinkfarbenen Zielreiz inmitten grüner Distraktoren dauert gleich lange wie die Suche nach einem blauen Zielreiz inmitten gelber Distraktoren.

15. Welche der folgenden Aussagen ist richtig?

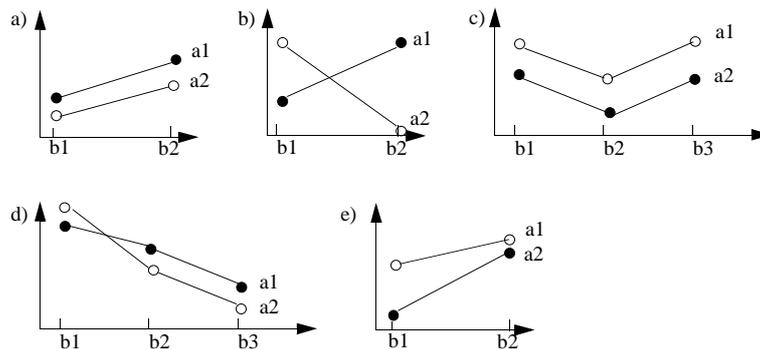
- a) Bei der Verbindungssuche gilt: Die Suche nach einem pinkfarbenen Zielreiz inmitten roter und grüner Distraktoren dauert länger als die Suche nach einem blauen Zielreiz inmitten gelber und blauer Distraktoren.
- b) Bei der Verbindungssuche gilt: Die Suche nach einem pinkfarbenen Zielreiz inmitten roter und grüner Distraktoren dauert weniger lange als die Suche nach einem blauen Zielreiz inmitten gelber und blauer Distraktoren.
- c) Bei der Verbindungssuche gilt: Die Suche nach einem pinkfarbenen Zielreiz inmitten roter und grüner Distraktoren dauert gleich lange wie die Suche nach einem blauen Zielreiz inmitten gelber und blauer Distraktoren.

16. Geben Sie ein Beispiel für ein 2 x 2 Design.

17. Welche der abgebildeten Plots zeigen eine Interaktion? Hier sind mehrere richtige Antworten möglich. Kreuzen Sie dann alle korrekten Antworten an. Gehen sie bei Ihrer Entscheidung davon aus, daß ein Effekt, wenn er sich in der Graphik abzeichnet, auch signifikant ist.



18. Welche der abgebildeten Plots zeigen einen Haupteffekt? Hier sind mehrere richtige Antworten möglich. Kreuzen Sie dann alle korrekten Antworten an. Gehen sie bei Ihrer Entscheidung davon aus, daß ein Effekt, wenn er sich in der Graphik abzeichnet, auch signifikant ist.



19. Warum reicht es für die Interpretation empirischer Ergebnisse nicht aus, sich Mittelwerte und Streuungen anzuschauen?
- Aus den Mittelwerten und Streuungen kann man keine Prognosen für das zukünftige Verhalten ableiten.
 - Die statistische Bedeutsamkeit der Ergebnisse kann anhand dieser Werte nicht beurteilt werden.
 - Um die Werte sinnvoll interpretieren zu können, muß auch das α -Niveau angegeben sein.
 - Man muß auch die Verteilung der Mittelwerte bei der Interpretation beachten.
20. Was wird durch das statistische Signifikanzniveau beschrieben?
- Das Signifikanzniveau legt fest, ab wann die Wahrscheinlichkeit, dass fälschlicherweise die Nullhypothese angenommen wird, gering genug ist, um die Gültigkeit der Alternativhypothese anzunehmen.
 - Das Signifikanzniveau legt die Wahrscheinlichkeit fest, ab der die Nullhypothese als Erklärung für die Ergebnisse angenommen werden kann.
 - Das Signifikanzniveau legt die Wahrscheinlichkeit fest, dass für die erhaltenen Ergebnisse die Alternativhypothese angenommen wird, obwohl die Nullhypothese gültig ist.

21. Aline und Bettina sind zwei Psychologiestudentinnen und wollen ihre Lizentiatsarbeit als Gemeinschaftsprojekt anfertigen. Als Thema haben sie sich Lernen mit Musik vorgenommen. Aus einer kognitiven Theorie des Lernens haben sie die Hypothese abgeleitet, daß der Lernerfolg unterschiedlich ausfällt, je nach dem, ob mit oder ohne Musik gelernt wird. Der Lernerfolg soll nach dem Lernen durch einen Wissenstest zum entsprechenden Lernstoff erhoben werden. Weitere Einflußfaktoren wurden in ihren Besprechungen diskutiert. Für das Design haben die beiden allerdings unterschiedliche Ideen.

Aline schlägt ein Experiment mit zwei Gruppen vor, die mit unterschiedlichem Lernstoff konfrontiert werden. Während die eine Gruppe deklaratives Wissen erwerben soll, in diesem Fall Definitionen psychischer Störungen, soll die andere Gruppe prozedurales Wissen erwerben: Diese Gruppe soll nach dem Experiment in der Lage sein, verschiedene voneinander unabhängige statistische Größen zu berechnen. Die Versuchspersonen sollen verschiedene psychische Störungen bzw. statistische Berechnungen nacheinander mit Musik verschiedener Richtungen lernen, in diesem Fall mit Entspannungsmusik, Hip Hop und in einer Kontrollbedingung ohne Musik. Nach jeder Lernphase erfolgt ein Wissenstest mit drei Parallelitems, die mit den beiden Musikstilen bzw. ohne Musik durchgeführt werden, wobei die Musik beim Nachtest in randomisierter Reihenfolge dargeboten wird. Nach jedem Test folgt eine Pause mit Ablenkungsaufgaben, damit beim Lernen keine Interferenzen entstehen.

Bettinas Vorschlag sieht ein Experiment vor, bei dem zunächst nur das Lernen deklarativen Wissens überprüft werden soll. Die Versuchspersonen sollen jeweils mit Musik (hier Entspannungsmusik) und ohne Musik lernen und auch zwei Paralleltests mit und ohne Musik absolvieren. Auch hier soll die Reihenfolge der Musik bei den Tests zufällig sein. Ebenso soll eine Pause mit Ablenkungsaufgaben zwischen den beiden Lernphasen eingelegt werden. Von diesem Experiment ausgehend möchte sie dann gegebenenfalls weitere Experimente planen.

Welchen Vorschlag favorisieren Sie?

Alines Vorschlag

Bettinas Vorschlag

Begründen Sie Ihre Entscheidung. Nennen Sie dabei Vor- und Nachteile der beiden vorgeschlagenen Experimentaldesigns.

22. Fabian hat für seine Lizentiatsarbeit in einem Projekt eine Untersuchung über die Behandlung von Kindern mit Aufmerksamkeitsdefizit-Syndrom durchgeführt. Er untersuchte, ob sich die medikamentöse Therapie, wenn sie mit einer Psychotherapie verknüpft wird, auch auf die Konzentrationsfähigkeit der Kinder positiv auswirkt. Die Beratungsstelle, die für die Rekrutierung der Untersuchungsteilnehmer zuständig war, konnte zu diesem Zeitpunkt nur eine begrenzte Anzahl von Kindern für die Studie gewinnen. Daher entschied man sich für ein Meßwiederholungsdesign, bei dem die Kinder in einer ersten Phase nur mit Medikament behandelt wurden und in einer zweiten Phase zusätzlich zur Medikamentengabe an einer Psychotherapie teilnahmen. In beiden Versuchsphasen wurde der Behandlungsverlauf alle zwei Wochen überprüft und dabei auch die Konzentrationsfähigkeit gemessen. Aus einer Pilotstudie wusste Fabian, dass sich die Psychotherapie bei diesen Kindern vor allem am Anfang stark auswirkt, diese Wirkung aber im Therapieverlauf eher abnimmt, weshalb den kleinen Patienten gerade von ärztlicher Seite eher nicht zu einer Psychotherapie geraten wird. Fabian hatte auch für seine Hauptuntersuchung die Hypothese, dass die Wirkung der Psychotherapie im Behandlungsverlauf eher abnimmt. Er erhielt die folgenden Ergebnisse:

Deskriptive Daten: Erreichte Punktzahl im Konzentrationstest

	Medikament			Medikament und Psychotherapie		
	M	SD	n	M	SD	n
Therapiedauer						
2 Wochen	67.0	10.5	20	88.9	13.6	20
4 Wochen	63.2	11.0	20	96.3	17.6	20
6 Wochen	70.1	12.2	20	115.3	25.6	20
8 Wochen	65.6	8.7	20	124.8	25.2	20
10 Wochen	64.5	12.4	20	135.7	23.9	20

Varianzanalyse (Meßwiederholung)

source	df	F
within subjects (S)	19	(1450.5)
between subjects		
Therapiedauer (A)	4	21.7***
Medikament mit/ohne Psychotherapie (B)	1	313.8***
A x B	4	25.1***
A x S	76	(173.4)
B x S	19	(341.6)
A x B x S	76	(158.3)

Anmerkungen: Werte innerhalb der Klammern repräsentieren die mittleren Quadratwerte.
* p<.05, ** p<.01, *** p<.001

Wie sind die Daten im Hinblick auf Fabians Hypothese zu interpretieren? Ergeben sich daraus gegebenenfalls neue Perspektiven?

Anhang E: Fragebogen für Dyaden nach dem Experiment

d

Code _____

Fragebogen 2

Bitte beantworten Sie die folgenden Fragen möglichst rasch und ohne groß darüber nachzudenken. Die Antworten werden anonym behandelt.

1. Wie gut kennen Sie Ihre Kooperationspartnerin/Ihren Kooperationspartner?

- Habe ich gerade eben kennen gelernt.
- Vom Sehen schon länger.
- Wir kennen uns schon länger.
- Wir sind befreundet.

2. Die Zusammenarbeit war konstruktiv.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

3. Wir haben die Probleme hinreichend beschrieben.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

4. Wir haben häufig die Ziele unserer Aufgabenstellung diskutiert.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

5. Unser Umgangston war mir überhaupt nicht sympathisch.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

6. Wir haben lange gebraucht, um Meinungsverschiedenheiten zu klären.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

7. Ich würde nächstes Mal die Sequenz an durchzuführenden Experimenten anders aufbauen. (Gehen Sie nicht vom Wissen über die visuelle Suche aus, das Sie nun erworben haben, sondern von methodischen Aspekten).

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

Beschreiben Sie hier, was Sie ändern würden.

8. Unser Weg, die Aufgabe anzugehen, war in meinen Augen effektiv.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

9. Mein Lernpartner hat sich bemüht, mir Dinge, die ich nicht gleich verstanden habe, zu erklären.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

10. Wir haben unsere Ideen miteinander abgestimmt, bevor wir etwas eingegeben haben.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

11. Meine Lernpartnerin/mein Lernpartner hat unsere Entscheidungen stärker beeinflusst als ich.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

12. Ich bin mit unseren Ergebnissen zufrieden.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

13. Meine eigene Leistung in diesem Experiment war sehr gut.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

14. Ich bin mit unserer Zusammenarbeit zufrieden.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

15. Ich würde gerne wieder mit meiner Lernpartnerin/meinem Lernpartner zusammenarbeiten.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

16. Ich habe meine Lernpartnerin/meinen Lernpartner als kompetent erlebt.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

17. Wir haben uns die Arbeit mit der Eingabe am Computer gleichmässig aufgeteilt.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

18. Es war schwierig, sich in *virtue* zurechtzufinden.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

19. Mir hat diese Form des Lernens am Computer Spaß gemacht.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

20. Man kann gut zu zweit am Computer eine Aufgabe in *virtue* lösen.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

21. Ich fand das Tippen lästig.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

22. In der Bedienung von Computern fühle ich mich völlig unsicher.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

23. Mit der Bedienung von *virtue* bin ich äusserst vertraut.

völlig unzutreffend 1 2 3 4 5 6 7 trifft zu

24. Arbeiten Sie viel am Computer?

- So gut wie nie.
- Selten.
- Mehrere Stunden in der Woche.
- Mehrere Stunden am Tag.

25. Können Sie schnell tippen?

- Nein.
- Ja.
- Professionell.

26. Hier ist Raum für Ihre Kommentare:

Vielen Dank!