

Ruprecht-Karls-Universität Heidelberg
Psychologisches Institut



Strategien zum komplexen Problemlösen

Eine experimentelle Analyse der Rolle von Analogien
beim Wissenserwerb
in komplexen, dynamischen Systemen

Diplomarbeit
Heidelberg, August 2003

vorgelegt von
Carsten A. Schulz
Punkerstr. 8
69126 Heidelberg

Erstgutachter: Prof. Dr. J. Funke
(Allgemeine und Theoretische
Psychologie)

Zweitgutachter: Prof. Dr. Kh. Sonntag
(Arbeits-, Betriebs- und
Organisationspsychologie)

*„Wie lehrt man den richtigen Umgang mit Unbestimmtheit und Komplexität?
Es wird wohl auch hier auf die richtige Strategie ankommen.“*

Dörner, 1992, S. 300

Inhaltsverzeichnis

Zusammenfassung.....	5
Abstract	6
1. Einleitung.....	9
2. Problemlösen mit dynamischen Systemen.....	12
2.1 Definition und Einordnung des Problemlösens.....	12
2.2 Kriterien des komplexen Problemlösens.....	15
2.3 Wissenserwerb.....	18
2.4 Wissensanwendung.....	21
2.5 Wissensdiagnostik.....	22
2.5.1 Kausaldiagramme.....	23
2.5.2 Paarvergleich.....	24
2.6 Taxonomien.....	25
2.7 Der DYNAMIS-Ansatz.....	26
2.8 Emotion und Motivation.....	28
3. Stile, Strategien und Taktiken.....	31
3.1 Stile, Metakognition, basale Operationen.....	33
3.2 Strategien.....	34
3.2.1 Beschreibungen verschiedener Strategien aus der Denkpsychologie.....	34
3.2.2 Analogien als Strategie zum komplexen Problemlösen.....	38
3.3 Taktiken.....	43
3.3.1 Difference Reduction versus Hill Climbing.....	43
3.3.2 Nulleingriffe, Einzeleingriffe, starke Eingriffe.....	44
3.4 Arbeitsdefinitionen von Strategie und Taktik.....	47
4. Methoden.....	48
4.1 Einflussfaktoren der Problemlösegröße – ein Rahmenmodell.....	48
4.2 Voruntersuchung: Einzelfallstudien – Lautes Denken	50
4.3 Hypothesen.....	52
4.4 Pilotstudie.....	55
4.4.1 Ergebnisse der Pilotstudie.....	55
4.5 Experimentelles Design und Versuchsablauf.....	56
4.5.1 Beschreibung und Einordnung des DYNAMIS-Systems „WITS“.....	58
4.5.2 Überlegungen zur Explorationsphase.....	61

4.5.3 Das Treatment der Analogien und der Taktiken.....	62
4.6 Vorüberlegungen zur Stichprobengröße.....	65
4.7 Untersuchungsdurchführung über das Internet.....	66
4.8 Datenanalyseverfahren	67
5. Ergebnisse.....	69
5.1 Beschreibung der Stichprobe.....	69
5.2 Analyse der Nutzung der Analogien.....	74
5.3 Ergebnisse zur Hypothese I	77
5.4 Analyse der Nutzung der Taktiken.....	80
5.5 Ergebnisse zur Hypothese II.....	82
5.6 Allgemeine Ergebnisse zur Steuerungsleistung.....	84
5.7 Ergebnisse zur Hypothese III.....	87
5.8 Analyse der Zeiten.....	88
6. Diskussion.....	91
6.1 Die Hypothesen.....	91
6.1.1 Hypothese I – Die Analogien.....	91
6.1.2 Hypothese II – Die Taktiken.....	94
6.1.3 Hypothese III – Der Wissenserwerb	95
6.2 Die Internetuntersuchungen und ihre Abbrecher.....	96
6.3 Untersuchungsinstrument DYNAMIS.....	98
6.4 Geschlechtseffekte.....	100
6.5 Einfluss von Intelligenz und mathematischen Fertigkeiten.....	102
7. Ausblick und Fazit.....	104
7.1 Prozessauswertung.....	104
7.2 Analogie und komplexe dynamischen Szenarien.....	105
7.3 Fazit	106
Literaturverzeichnis.....	107
Anhang.....	116
Anhang A.....	117
Die DYNAMIS – Steuerdatei für das System „WITS“.....	117
Anhang B.....	118
Die Werbung in Emails, Newsgroups und Email-Verteilern.....	118
Anhang C.....	123
Instruktionstexte und Erklärungen zum DYNAMIS-System „WITS“.....	123

Zusammenfassung

In einem Online-Experiment wurde die Annahme geprüft, dass (1) der Einsatz von Analogien das Bearbeiten eines komplexen, dynamischen Systems vereinfacht. Unter anderem wird angenommen, dass der Einsatz von Analogien den Erwerb von Strukturwissen erleichtert. Der Erwerb von Strukturwissen wurde mit einer eigens für das Online-Experiment konzipierten Wissensdiagnostik erhoben.

Des Weiteren wurde (2) der Einfluss von spezifischen Eingriffsmustern auf das Strukturwissen analysiert. Dazu wird eine Unterscheidung zwischen diesen Eingriffsmustern, in dieser Arbeit als *Taktiken* bezeichnet, und übergeordneten Strategien eingeführt. Bei diesen Taktiken handelt es sich um so genannte „strong methods“ (Robertson, 2001) des komplexen Problemlösens. Sie werden in dieser Arbeit benannt und definiert.

Zusätzlich wurde geprüft, ob (3) Strukturwissen positiven Einfluss auf die Steuerleistungen hat, wie durch die Experimente von Preußler (1998) und Vollmeyer, Burns und Holyoak (1996) angenommen wird.

Zur Überprüfung der drei Annahmen diente ein abstraktes, dynamisches Szenario, welches in der Online-Version des Computerprogramms DYNAMIS umgesetzt wurde. In zwei unterschiedlichen Pilotstudien konnten unter anderem sieben typische Fehler von Versuchspersonen in solchen DYNAMIS-Szenarien ausfindig gemacht werden.

Die Ergebnisse zeigen, dass der Einsatz von Analogien den Wissenserwerb tendenziell erleichtert (1), die Nutzung von sinnvollen Taktiken zu mehr Strukturwissen führt (2) und Strukturwissen einen positiven Einfluss auf die Steuerleistung hat (3), was die benannten Experimente von Preußler (1998) bzw. Vollmeyer et al. (1996) bestätigt. Zusätzlich liegen durch die Unterscheidung einzelner Taktiken differenziertere Zusammenhänge zum Wissen und zur Steuerleistung vor.

Das über das Internet zu bedienende DYNAMIS-Szenario erwies sich als relativ schwer und teilweise überfordernd. Wahrscheinlich wäre bei einem einfacheren Szenario mit noch größeren Treatment-Effekten und Korrelationskoeffizienten zu rechnen gewesen. Die nahezu durchweg positiven Ergebnisse zeigen das Potential der Zusammenhänge der Taktiken mit dem Strukturwissen und machen ein differenzierteres Bild der Eingriffe in DYNAMIS-Szenarien möglich.

Abstract

The online experiment proved the hypothesis that (1) the employment of analogies facilitates the handling of a complex dynamic system. Among other things it is supposed that the employment of analogies simplifies the acquisition of structural knowledge. The acquisition of structural knowledge was inquired by the means of a knowledge diagnostic, which was conceived especially for the online experiment.

In addition to this, (2) the influence of specific input patterns on the structural knowledge was analysed. Moreover a differentiation between these input patterns (which are denominated in this work as *tactics*) and preceding strategies is introduced. These tactics are the so called “strong methods” (Robertson, 2001) of the complex problem solving domain. They are denominated and defined in this work.

Additionally it was proved, (3) whether structural knowledge has a positive influence on the controlling achievement, as supposed by Preußler (1998) and Vollmeyer, Burns, and Holyaok (1996) by means of their experiments.

To prove the three hypotheses, an abstract dynamic scenario was utilised. It was realised in the online version of the software DYNAMIS. In two different pilot studies seven typical mistakes of test subjects could be revealed in these DYNAMIS- scenarios.

The results illustrate, that the employment of analogies seems to facilitate the acquisition of knowledge (1), that the usage of reasonable tactics leads to more structural knowledge (2) and that structural knowledge has a positive influence on the controlling achievement (3), confirming the cited experiments of Preußler (1998) and Vollmeyer et al. (1996). Furthermore differentiated connections concerning knowledge and controlling achievement are available through the differentiation of singular tactics.

The Internet-based DYNAMIS scenario revealed to be relatively difficult and partly demanding too much. Probably a more simple scenario could have increased the effects of the treatment and the correlation coefficients. The nearly without exceptions positive results illustrate the potential of connections between tactics and structural knowledge and allow a differentiated view on interventions in DYNAMIS- scenarios.

Tabellenverzeichnis

<i>Tabelle 1: Oberflächen- und Strukturmerkmale von Problemen.....</i>	<i>41</i>
<i>Tabelle 2: Gegenüberstellung der Konzepte Strategie und Taktik.</i>	<i>47</i>
<i>Tabelle 3: Versuchsdesign des Experiments mit den einzelnen Phasen.....</i>	<i>57</i>
<i>Tabelle 4: Die Matrizen für das WITS-Szenario aus der Steuerdatei für DYNAMIS.....</i>	<i>58</i>
<i>Tabelle 5: Angaben der Versuchspersonen zu ihrem Berufsstand.....</i>	<i>72</i>
<i>Tabelle 6: Subjektive Einschätzung des Computerumgangs und Dauer der Nutzung.....</i>	<i>73</i>
<i>Tabelle 7: Mittelwerte und Standardabweichungen der Informationsaufrufe.....</i>	<i>75</i>
<i>Tabelle 8: Anzahl korrekter Angaben im Wissenstest vor dem Zieldurchgang.....</i>	<i>78</i>
<i>Tabelle 9: Anzahl falscher Angaben im Wissenstest vor dem Zieldurchgang.....</i>	<i>78</i>
<i>Tabelle 10: Mittelwerte (M) und Standardabweichung (s) aller Taktiken für die unspezifische (KG1) und die taktische (KG2) Kontrollgruppe sowie für die Experimentalgruppen mit einer (EG1) und drei (EG2) Analogien.....</i>	<i>81</i>
<i>Tabelle 11: Rangkorrelation (rs) nach Spearman für die Taktikvariablen mit den Strukturwissensvariablen.....</i>	<i>83</i>
<i>Tabelle 12: Regressionsmodell für die Güte der Systemsteuerung (GdS) mit und ohne die Note für Mathematik mit standardisierten und nicht standardisierten Koeffizienten und T-Werten. Achtung: Für GdS und die Noten gilt, je kleiner der Wert, desto besser die Leistung.....</i>	<i>86</i>
<i>Tabelle 13: Zusammenfassung der Regressionsmodelle für GdS.....</i>	<i>86</i>
<i>Tabelle 14: Produkt-Moment-Korrelation nach Pearson für den Zusammenhang zwischen den Wissensvariablen nach Durchgang 3 und den Zielvariablen mit den logarithmierten Abweichungen angegeben als Güte der Steuerungsleistung (GdS) nach Müller (1993)....</i>	<i>87</i>
<i>Tabelle 15: Mittelwerte und Standardabweichung der Bearbeitungsdauer (in Minuten) der vier Untersuchungsgruppen.....</i>	<i>88</i>
<i>Tabelle 16: Mittelwerte und Standardabweichung der Zeiten einzelner Phasen der Untersuchung.....</i>	<i>90</i>
<i>Tabelle 17: Mittelwert, Standardabweichung der Zielabweichungen (GdS) für die verschiedenen Treatmentgruppen.....</i>	<i>92</i>
<i>Tabelle 18: Mittelwerte und Standardabweichungen einiger soziodemographischer und abhängiger Variablen nach Geschlecht mit den Ergebnissen einer ANOVA und Angaben zur Effektgröße.....</i>	<i>101</i>
<i>Tabelle 19: Korrelationen der Abschlussnote mit Prädiktoren für die Steuerungsleistung, bivariate und für die Mathematiknote partialisierte Korrelation.....</i>	<i>102</i>
<i>Tabelle 20: Regressionsmodell zur Steuerungsleistung (GdS) mit standardisierten Koeffizienten.....</i>	<i>103</i>

Abbildungsverzeichnis

<i>Abbildung 1: Kausaldiagramm des DYNAMIS-Systems "Sinus".</i>	24
<i>Abbildung 2: Screenshot aus DYNAMIS: Das Testfenster für die Wissensdiagnostik zur Überprüfung des Strukturwissens.</i>	25
<i>Abbildung 3: Die Aktivationsfunktion nach dem Yerkes-Dodson-Gesetz.</i>	29
<i>Abbildung 4: Kontinuum der Ähnlichkeit. Je stärker Oberfläche und Struktur von Ziel und Quelle voneinander abweichen, desto unähnlicher sind die Probleme, bis sie nicht mehr vergleichbar sind. Wenn Struktur und Oberfläche gleich sind (links unten), dann sind die Probleme identisch.</i>	39
<i>Abbildung 5: Ein einfaches Rahmenmodell der Einflussfaktoren auf die Problemlösegüte (Steuerleistung) mit graphischer Darstellung der Hypothesen (Hyp.).</i>	49
<i>Abbildung 6: Kausaldiagramm des DYNAMIS-Systems "WITS".</i>	59
<i>Abbildung 7: Screenshot des "WITS"-Szenarios in Durchgang 1 für die Experimentalgruppen.</i>	63
<i>Abbildung 8: Screenshot des dritten Durchgangs der Experimentalgruppe (EG2).</i>	63
<i>Abbildung 9: Screenshot zum Durchgang 3 der Experimentalgruppe (EG2).</i>	64
<i>Abbildung 10: Altersverteilung in der Stichprobe.</i>	71
<i>Abbildung 11: Häufigkeiten der Befragung nach Erfahrung, Neigung und verbrachter Zeit mit Computern.</i>	73
<i>Abbildung 12: Bearbeitungsdauer des Experimentes.</i>	74
<i>Abbildung 13: Mittelwerte der Abfragen der Instruktionstexte.</i>	74
<i>Abbildung 14: Mittelwerte der Abfragen des Zieltextes.</i>	74
<i>Abbildung 15: Anzahl der Wechsel der Variablenbezeichner je Takt für Gruppen und Durchgänge getrennt. Gemittelt über die je $n = 14$ Versuchspersonen pro Gruppe.</i>	76
<i>Abbildung 16: Mittelwerte des Strukturwissens für Treffer (a) bzw. Fehler (b) in den Gruppen.</i>	79
<i>Abbildung 17: Gemittelte Zeiten der Gruppen für das gesamte Experiment in Minuten.</i>	89
<i>Abbildung 18: Gemittelte Zeiten der Gruppen für die vier Durchgänge in Minuten.</i>	89

1. Einleitung

Was die Forschung mit komplexen Problemen so spannend macht, ist auf der einen Seite die Häufigkeit, mit der Menschen mit solchen Problemen zu tun haben oder zumindest über sie nachdenken und diskutieren, auf der anderen Seite die Angst der Menschen vor Komplexität und davor, von vernetzten, dynamischen Systemen überfordert zu werden. Ein prominentes Beispiel, welches wohl nicht nur in den 1980er Jahren diskutiert wurde, ist die Umweltzerstörung, von Autoabgasen bis zum Waldsterben. Und nicht erst die 1990er Jahre brachten die Auseinandersetzungen um politische Systeme, Arbeitslosigkeit und Globalisierung in die öffentliche Diskussion. Ob das aktuelle Jahrzehnt geprägt sein wird von der Analyse militärischer Einsätze in Krisengebieten, Epidemien, Entwicklungshilfe, Freihandelszonen oder Überbevölkerung, muss sich erst noch zeigen. Neben diesen großen und eindeutig komplexen Problemen werden Menschen aber auch im Alltag oft mit der Frage konfrontiert, wie ein komplexes System funktioniert und wie es am besten so beeinflusst wird, dass das erwünschte Ergebnis zu Stande kommt. Vom Kauf eines Autos, der Entscheidung für eine Krankenkasse, bis hin zur Bedienung einer fremden Kaffeemaschine oder Heizungsanlage am Urlaubsort könnten im Alltag viele Beispiele für komplexe Probleme gefunden werden.

Trotz der Wichtigkeit von Kompetenzen im Umgang mit Komplexität ist das komplexe Problemlösen kein Schulfach und wird auch nicht an anderer Stelle gelehrt und vermittelt. So ist es auch eigentlich kein Wunder, dass oftmals Ängste und Zweifel an der eigenen Kompetenz bestehen (zu Emotionen und Kontrollüberzeugung beim komplexen Problemlösen siehe Spring 2002). In der heutigen Informationsgesellschaft wird ein jeder nahezu überflutet mit Daten, Statistiken und Fakten. Geht mit dieser Zunahme an Daten und an Wissen auch ein besserer Umgang mit Problemen einher? Einige prominente Wissenschaftler würden dies wohl eindeutig verneinen. Herbert Simon weist schon recht lange und erfolgreich auf die Grenzen menschlichen Denkens hin (Simon, 1972, 1992, 2001). Und auch in Deutschland gibt es mit Gerd Gigerenzer und seiner ABC Research Group deutliche Stimmen, welche die Unbegrenztheit menschlicher Vernunft anzweifeln (Gigerenzer & Selten, 2001; Gigerenzer, Todd & ABC Research Group, 1999).

Immer genaueres Detailwissen in Fachgebieten führt mehr und mehr zur Ausbildung von Spezialisten. Innerhalb der Fachgebiete hilft dieses Wissen sicherlich auch bei auftretenden Schwierigkeiten, was ist aber über die Problemlösefähigkeit in fachfremden Gebieten zu

sagen? Was hilft in neuen komplexen Situationen einer oder einem Problemlöser?

In dieser Arbeit sollte daher etwas untersucht werden, was fachgebietsübergreifend Wirkung zeigen kann und gleichzeitig recht einfach zu vermitteln ist. Laut Rollett (2002) wurden bereits eine große Anzahl verschiedenster Prädiktoren untersucht, ohne die große interindividuelle Varianz im Explorations-, Steuerungs- und Leistungsverhalten von Versuchspersonen aufzuklären. Zu diesen Prädiktoren gehören z.B. Intelligenzfaktoren, Persönlichkeitsfaktoren, Motive, kognitive Stile, religiöse und politische Orientierungen, Werthaltungen, Motivation, Metakognition, Attribution u.v.m.. Die Wahl eines Themas fiel in dieser Arbeit auf den Bereich der Strategien. Dieser Bereich wurde bisher in der Forschung zu komplexen Problemen wenig untersucht, aber es verbinden sich damit große Hoffnungen für die Verbesserung der Problemlöseleistung in komplexen Situationen.

Diese Diplomarbeit befasst sich innerhalb des Bereichs der Strategien mit dem Einsatz von Analogien als Strategien des komplexen Problemlösens und benutzt als Untersuchungswerkzeug ein Szenario auf Basis des DYNAMIS-Systems.

Das DYNAMIS-System ist ein Computerprogramm, das Funke bereits 1986 entwickelte. Es kann innerhalb einer Oberfläche verschiedenste dynamische Szenarien darstellen. Es ging aus einem Forschungsprojekt an der Universität Bonn hervor und wird inzwischen weit verbreitet in der Problemlöseforschung eingesetzt. So gibt es Forschungsgruppen in Bamberg, Berlin, Bonn und Trier, die das DYNAMIS-System verwenden. Zudem hat es die komplexe Problemlöseforschung der letzten zwanzig Jahre verändert und geprägt (Funke, 1986, 2003).

Das Programm selbst wird seit einigen Jahren an der Universität Heidelberg so modernisiert, dass es in neuen Programmiersprachen vorliegt und auch über das Internet benutzt werden kann. Dies ermöglicht, die Datenerhebung aus dem Psychologischen Institut auszulagern und deutschlandweit Versuchspersonen zu rekrutieren. Bei entsprechender Übersetzung der Dateien ist ein Einsatz über Sprachgrenzen hinaus leicht vorstellbar und ließe dann durchaus internationale Vergleiche zum komplexen Problemlösen zu.

Kapitel 2 geht neben der Darstellung des DYNAMIS-Ansatzes allgemein auf das komplexe Problemlösen ein und hebt besonders den Umgang mit Wissen bei komplexen Systemen hervor. Das Kapitel 3 befasst sich mit den verschiedenen Strategien und insbesondere mit Analogien beim Problemlösen. Zudem führt dieses Kapitel eine sprachliche Unterscheidung ein, die den Begriff der Strategie von dem der Taktik beim Problemlösen abhebt, da diese in

1 Um eine bessere Lesbarkeit der Arbeit zu gewährleisten, wird im folgenden, wenn nicht anders umschreibbar, nur noch eine der beiden Geschlechterformen verwendet. Gemeint sind aber stets männliche und weibliche Personen.

der Literatur sehr uneinheitlich verwendet werden. Es folgt der methodische Teil in Kapitel 4 mit den beiden Pilotstudien, den Hypothesen und der Versuchsbeschreibung, bevor in Kapitel 5 die Ergebnisse des Experimentes dargestellt werden. Die Ergebnisse werden in Kapitel 6 diskutiert, bevor in Kapitel 7 ein Ausblick gegeben und ein abschließendes Fazit gezogen wird.

2. Problemlösen mit dynamischen Systemen

Dieses Kapitel widmet sich dem Problemlösen mit dynamischen Systemen. Es soll vom Allgemeinen zum Speziellen die Bereiche darstellen, die für das Problemlösen in dieser Untersuchung relevant sein werden. Dabei wird in Abschnitt 2.1 mit verschiedenen Definitionen begonnen und neben den Kriterien die Probleme ausmachen, sollen in Abschnitt 2.2 auch die Kriterien genannt werden, die erfolgreiche Problemlöser ausmachen. Abschnitt 2.3 widmet sich Lernprozessen und ihrem Zusammenhang mit dem Wissenserwerb beim komplexen Problemlösen. Dabei wird auf Produktionssysteme hingewiesen (Anderson, 1983, 1989; Anderson & Thompson, 1989), die diesen Prozess mit hoher Genauigkeit darstellen können. In Abschnitt 2.4 wird es um die Wissensanwendung gehen, und es wird kurz auf die Kontroverse über die Nützlichkeit des Strukturwissens für die Steuerleistung eingegangen. Neben dem Wissenserwerb und der Wissensanwendung ist die Wissensdiagnostik nicht zu vergessen, da das Problem der individuellen Repräsentation des Wissens für die Problemlöseforschung von zentraler Bedeutung bleibt. Als Beispiele für Verfahren der Wissensdiagnostik werden Kausaldiagramme und Paarvergleiche in Abschnitt 2.5 besprochen.

Um später das spezielle Szenario, welches für die Untersuchung gewählt wurde, genauer einordnen zu können, ist ein sinnvoller Rahmen nötig. Hierfür wird die Taxonomie von Funke (1990) gewählt. Auf die Taxonomie wird kurz in Abschnitt 2.6 eingegangen. In Abschnitt 2.7 wird gezeigt, warum es zu der Entwicklung des DYNAMIS-Ansatzes kam und was diesen Ansatz ausmacht. Im letzten Abschnitt 2.8 sollen Emotion und Motivation im Hinblick auf das Bearbeiten komplexer Problemszenarien beschrieben werden. Sowohl die Datenerhebung über das Internet als auch die Länge der meisten dynamischen Probleme an sich weisen auf die wichtige Rolle hin, die gerade die Motivation bei solchen Experimenten spielt.

2.1 Definition und Einordnung des Problemlösens

Denken, Problemlösen oder problemlösendes Denken sind Begriffe, die nicht leicht voneinander getrennt behandelt werden können. Gemeinsam ist ihnen, dass man unter ihnen Teilgebiete der Allgemeinen Psychologie versteht. Ob Denken als Problemlösen definiert werden sollte (Dörner, 1979) oder ob Problemlösen nur ein Teilgebiet der Denkpsychologie ist und damit getrennt vom Denken definiert werden sollte (Hussy, 1984), ist wohl von der jeweiligen Perspektive abhängig.

Einige Sichtweisen sollen hier aufgezeigt werden, bevor auf eine Arbeitsdefinition von Funke (2003) eingegangen wird, die als Grundlage des in dieser Arbeit verwendeten Begriffs des

Problemlösens dienen soll.

Neben Dörner zählt auch Johnson zu den Autoren, die Denken und Problemlösen gemeinsam betrachten, er drückt seinen Ansatz folgendermaßen aus: „[...] *we can define thinking as problem solving and including mediators and mental experience to the extent that they assist the explanation of problem solving.*“ (Johnson, 1972). Andere Autoren wie Hussy ordnen das Problemlösen der Denkpsychologie unter. Dabei definiert er es als „[...] *das Bestreben, einen gegebenen Zustand (Ausgangszustand, Ist-Zustand) in einen anderen, gewünschten Zustand (Zielzustand, Soll-Zustand) über zuführen, wobei es gilt eine Barriere zu überwinden, die sich zwischen Ausgangs- und Zielzustand befindet.*“ (Hussy, 1984, S.114). Auch Dörner benutzt den Begriff der Barriere und klassifiziert Probleme nach der Art der Barriere, die davon abhängt, welche *Operatoren* vorhanden und welche benötigt sind (Dörner, 1976; siehe auch Abschnitt 'Taxonomien' auf Seite 25). Mit Operatoren sind die Mittel gemeint, mit denen man im jeweiligen Realitätsbereich Operationen ausführen kann (beim Schach z.B. die Rochade, der Springerzug etc.). Diese Operatoren können sich erstrecken von übergeordneten Strategien über bereits vorhandenes Lösungswissen aus ähnlichen Problemen bis hin zu einfachen Werkzeugen wie Hammer oder Zange (z.B. im Realitätsbereich der Autoreparatur).

Einer der Autoren, der Denken und Problemlösen getrennt voneinander betrachtet, ist Nickerson. Er weist auf die Wichtigkeit der Erforschung des Problemlösens und Denkens hin und auf die Notwendigkeit, die Erkenntnisse auch in Schulen und Universitäten zu lehren. Immer wieder können Mängel in der Problemlöseleistung der Bevölkerung aufgezeigt werden, woraufhin staatliche Programme initiiert werden, die Abhilfe schaffen sollen. So gab es bereits in den USA in den 1920er Jahren, ebenso wie in den 1940er Jahren größere Forschungsunternehmungen, die sich dem Problemlösen widmeten. Nun scheint wieder so eine Zeit gekommen zu sein, wo es nötig zu sein scheint, schreibt er in „Teaching Thinking and Problem Solving“ (Nickerson, 1998).

Wie Denken und Problemlösen definiert wird, hängt auch ab von den Untersuchungsmethoden, die gewählt werden (mehr über Modellfunktion von Werkzeugen siehe Ahrens, 1984; Gigerenzer, 1981, 1984, 1991). Der Großteil der Problemlöseforschung beschäftigt sich bis heute mit resultatorientierten Denksportaufgaben wie dem „Turm von Hanoi“ (TvH) oder den „Kannibalen und Missionaren“ (KuM). Untersuchungen beschränken sich auf Laborsettings, die gegenüber dem realen Leben recht trivial anmuten. Bezeichnet wird dieser Ansatz daher auch als triviales Problemlösen.

Gegenüber diesen trivialen Problemlösesettings der Denkpsychologie hebt sich vor allem in Deutschland seit Ende der 1970er Jahre die *komplexe* Problemlöseforschung ab. Angeregt durch den Einsatz von Computern zur Simulation von Mikrowelten als kleine komplexe Szenarien wurde eine Forschungswelle ausgelöst, die scheinbar ebenso komplex, vielzellig und dynamisch ist, wie ihr Gegenstand selbst. Hier ist vor allem Dietrich Dörner und sein aus über 2000 Variablen bestehendes System 'Lohhausen' zu nennen (Dörner, Kreuzig, Reither & Stäudel, 1983). Ein weiteres Beispiel ist der 'Tailorshop' von Putz-Osterloh und Lürer (1981) oder auch 'MORO' von Dörner, Stäudel und Strohschneider (1986; einen Überblick über weitere Szenarien liefert Funke, 1986, 1993, 2003).

Gemeinsam bleibt all diesen Bestrebungen, dass der Gegenstand der Denkpsychologie aus den psychologischen Laboratorien herausgeholt und auf mehr Realität bezogen werden soll. So schreiben Dörner et al. zum Denken in realen Situationen: „*Denken ist in der Realität immer eingebettet in eine bestimmte Situation, eingebettet in einen Handlungsablauf und muß sich diesem anpassen. Denken geschieht unter Zeitdruck, oft unter relativ extremen emotionalen Belastungen. Man denkt, weil man Angst hat vor den Folgen einer möglichen Entwicklung [...] man denkt nach, weil die gesamte Situation so komplex ist, daß man sie nicht durchschaut: Denken dient der Komplexitätsreduktion.*“ (Dörner, Schaub und Strohschneider, 1999, S.198)

Deutlich wird hier, dass es zur Untersuchung eines solchen Phänomenbereichs mehr braucht als einfache Streichholzprobleme, Rechenaufgaben oder „Turm von Hanoi“-Derivate. Mit dynamischen und stark vernetzten Szenarien, die als Mikrowelten in Computersimulationen ablaufen, wird versucht, das Problemlösen im Sinne des obigen Zitats realitätsnäher zu beschreiben und zu erforschen.

Nach Dörner und Schölkopf (1991) sind folgende fünf Komponenten für das kompetente Lösen komplexer Probleme notwendig (aus Knoblich, 2002):

1. eine adäquate Formulierung von Zielen,
2. ein adäquates Modell der Systemstruktur,
3. eine adäquate Prognose der zukünftigen Systementwicklung,
4. das adäquate Treffen von Entscheidungen und
5. eine adäquate Kontrolle von Hypothesen und Strategien zur Systemsteuerung.

Erfolgreiche Problemlöser zeichnen sich demzufolge dadurch aus, dass sie

- gezielt Systemwissen zu gewinnen versuchen, um eine gute Grundlage für Zukunftsprognosen zu haben,
- die Auswirkung früherer Eingriffe beobachten, bevor sie neue Eingriffe in das System vornehmen und
- bisher getroffene Entscheidungen und Strategien immer wieder kritisch reflektieren.

Nach der Durchsicht und dem Vergleich vieler aktueller und historischer Definitionen kommt Funke in seinem jüngsten Werk zum „Problemlösenden Denken“ zu einer Arbeitsdefinition, in der die Konzepte des Handlungsplans und der Repräsentation des Problembereichs betont werden. Damit wird auf den großen Teil bewussten Handelns hingewiesen und auf die zentrale Rolle des Gedächtnisses beim Problemlösen. Die Arbeitsdefinition lautet: *„Problemlösendes Denken erfolgt, um Lücken in einem Handlungsplan zu füllen, der nicht routinemäßig eingesetzt werden kann. Dazu wird eine Repräsentation erstellt, die den Weg vom Ausgangs- zum Zielzustand überbrückt.“* (Funke, 2003, S. 20).

Gerade die Repräsentation des Lösungsweges scheint noch weitgehend unerforscht. Inwieweit prozessorientierte Forschungsansätze oder die Untersuchung von Strategien zur Aufklärung beitragen können, muss sich erst noch zeigen. Dass dazu aber komplexe, realitätsnahe Simulationen wichtig und vielleicht sogar nötig sind, wird deutlich, betrachtet man die Unterschiede zu den trivialen Problemlösesettings. Nachfolgend soll daher auf die Kriterien des komplexen Problemlösens eingegangen und diese sollen näher spezifiziert werden.

2.2 Kriterien des komplexen Problemlösens

Ein komplexes Problem zeichnet sich durch vier wesentliche Kriterien aus (Dörner, 1992), die in diesem Abschnitt dargestellt werden. Die vier Kriterien sind:

- *Vernetztheit*,
- *Dynamik*,
- *Intransparenz* und
- *Polytelie*.

Oft wird noch der *Grad des Vorhandenseins freier Komponenten* aufgeführt. Diese freien Komponenten sind situationsabhängig und können daher nicht allgemein beschrieben werden. Sie nehmen jedoch einen großen Einfluss auf die Problemlöser. Wenn z.B. keine Ersatzteile in einer Reparaturwerkstatt vorhanden sind oder kein Flugzeug für geographische Beobachtungen, dann werden Problemanalysen jeweils anders angegangen und müssen auch anders ausfallen.

Nicht vergessen werden sollte auch der namensgebende Begriff der *Komplexität* für diese Art des Problemlösens. Dörner (1979) beschreibt die Komplexität mit der Anzahl der Elemente und der Vielfalt der Verknüpfungen. Durch sie werden Problemlösende zu komplexitätsreduzierenden Maßnahmen gezwungen, da deren kognitive und andere Ressourcen an ihre Grenzen stoßen. Bekannte Maßnahmen sind das Ausklammern von Merkmalen, die *Abstraktion* oder die *Reduktion* von vielen Einzelmerkmalen auf ein Grundmerkmal. Auch durch *Komplexbildung* wird Komplexität reduziert, indem ein sinnvolles Auflösungslevel gewählt wird. Beispielsweise wurden bei der Autoreparatur vor 20 Jahren noch Einzelteile betrachtet, auf Fehler untersucht und gegebenenfalls ausgewechselt, während heutzutage nur noch größere Einheiten wie Motor oder Stoßdämpfer betrachtet und ausgewechselt werden. Problematischerweise lässt sich die Komplexität nicht genau spezifizieren. Es lässt sich kaum ein Kriterium festlegen, welches bestimmt, wann Komplexität bei einem Problem vorliegt und wann nicht. So schreibt dann auch Dörner von der Subjektivität der Komplexität, da sie von den Problemlösenden abhängt und es kein objektiv bestimmbares Maß für sie gibt (Dörner, 1992, S. 62).

Die vier nachfolgend beschriebenen Merkmale können genauer bestimmt werden und weisen auf das Vorhandensein eines komplexen Problemes hin, wenn sie einzeln oder gemeinsam in einer Problemkonstellation auftreten.

Vernetztheit.

Allgemein spricht man im Rahmen des komplexen Problemlösens immer dann von Vernetztheit, wenn der Eingriff auf eine Variable verbunden ist mit Veränderungen einer oder mehrerer anderer Variablen.

Vernetztheit bedeutet nach Dörner (1995, S. 298), „[...] dass man – mehr oder weniger – nicht das **eine** tun kann, ohne auch etwas **anderes** zu tun.“ Typische Konstellationen von Vernetztheit treten z.B. in ökologischen Systemen auf. Entfernt man aus einem Biotop alle Frösche, so vermehren sich Insekten in stärkerem Ausmaß und Störche finden dort eventuell keine Nahrung mehr und verlassen das Biotop. Ein weiteres Beispiel für Vernetztheit sind medizinische Präparate, die meist verschiedene Nebenwirkungen haben. So können Kopfschmerztabletten bei empfindlichen Personen zu Magenbeschwerden führen oder gar zu Leberschäden. Bei komplexen, dynamischen Systemen hat sich im Sprachgebrauch der Begriff *Seiteneffekt* eingebürgert. Er hebt sich vom Begriff der *Nebenwirkung* ab, die als eine Form der Dynamik zu betrachten ist.

Dynamik.

Des Weiteren werden dynamische von stabilen Problemkonstellationen unterschieden. Als stabil ist ein System zu bezeichnen, wenn sich das System nur beim aktiven Eingriff der Problemlösenden verändert. Greifen sie hingegen nicht ein, darf sich das System auch nicht verändern. Ein dynamisches System ist dadurch gekennzeichnet, dass es sich auch ohne äußeres Zutun verändert. Ein Beispiel für Dynamik sind Wachstumsprozesse von Bevölkerungen, die mit steigender Bevölkerungszahl zunehmen. Von *Eigendynamik* wird dann gesprochen, wenn Variablen auf sich selbst wirken, wie z.B. der Zersetzungsprozess von radioaktiven Stoffen oder der biologische Verfall im Komposthaufen. Von *Nebenwirkung* wird gesprochen, wenn der Zustand einer Variablen den einer anderen direkt beeinflusst, so würde z.B. die Menge des Abfalls den Geruch beeinflussen oder die Menge des radioaktiven Stoffes die Strahlung.

Intransparenz.

Mit der Transparenz wird der Grad der Beobachtbarkeit beschrieben. Ein Schachspiel ist beispielsweise vollständig beobachtbar und alle Zusammenhänge sind den Spielenden bekannt. Auch die Positionen und damit der Wirkungsgrad der Figuren ist für den Beobachter gänzlich einzusehen. Das Problem eines Computerabsturzes hingegen ist für die meisten Nutzer eines solch komplexen Gerätes meist völlig intransparent. Wer kennt schon alle Prozesse, Tasks, Befehlsaufrufe und Festplattenzugriffe eines aktiven Computersystems? Welche dieser Aufgaben eine andere beeinflusst und durch einen Fehlgriff eventuell zum Absturz geführt hat, ist daher oft nicht nachvollziehbar und braucht zur Lösung für gewöhnlich die Tätigkeit des Hypothesenbildens. Man muss versuchen, das zu erschließen, was nicht sichtbar ist. Hypothesen über die innere Struktur bilden in solchen Fällen den Anfang der Suche nach der Lösung. Hier spielen Vorannahmen, Vorwissen und auch Weltbilder der Problemlösenden eine bedeutende Rolle (Dörner, 1995, S. 299).

Polytelie.

Unter Polytelie wird das gleichzeitige Vorhandensein mehrerer Ziele verstanden. Ausführlich beschreibt Dörner (1992) unterschiedliche Ziele und ihre Implikationen füreinander. Man unterscheidet zunächst zwischen *positiven* Zielen, die einen Soll-Zustand erreichen oder beibehalten wollen, und *negativen* Zielen, also solchen, bei denen ein Ist-Zustand einfach nur verändert werden soll. Die meisten negativen Ziele sind *global* beschrieben. Bei globalen Zielen gibt es keine weiteren Angaben zum Ziel. Im Gegensatz dazu sind bei *spezifischen* Zielen viele Parameter des Soll-Zustandes beschrieben, so dass sich dieser leicht vorstellen

lässt. Zusätzlich unterscheidet Dörner noch in *unklare* und *klare* Ziele. Bei unklaren Zielen ist oft nicht bekannt, wie der Soll-Zustand genau aussehen soll. Dabei handelt es sich meist nur um Oberbegriffe wie „Bekämpfung der Arbeitslosigkeit“, die wiederum aus Mehrfachzielen bestehen. Dörner schlägt als Strategie der Bearbeitung solcher unklaren Ziele die *Dekomposition* vor, also die Zerlegung in Teilkomponenten des Gesamtziels (Dörner, 1992, S. 81f.). Klare Ziele sind solche, von denen bereits ein recht genaues Bild vorliegt, mit genauer Definition der nötigen Parameter oder ohne.

Enthält ein Problem die Komponente der Polytelie und die der Vernetztheit, so ergeben sich weitere Konflikte für die Zielbehandlung, denn Vernetztheit der Variablen heißt auch Vernetztheit der Ziele. So kann es sein, dass durch Erreichen des einen Ziels ein anderes mit erreicht wird. Dies wird als *positive* Verknüpfung der Ziele bezeichnet. Die Ziele können sich aber auch in ihrer Erreichung ausschließen, dann sind sie *negativ* verknüpft. In beiden Fällen sind Ziele nicht *direkt*, sondern über Drittvariablen oder Zeitverzögerungen *indirekt* miteinander verknüpft. Dies ist meist erheblich schwerer zu erkennen und führt zu vielen für Menschen typische Fehler. Dörner (1992) beschreibt das Unglück von Tschernobyl und weist darin unter anderem auf die Schwierigkeiten im Erkennen der zeitverzögerten Reaktionen hin.

Alle vier Kriterien müssen von Problemlösenden beachtet werden, wenn sie zu guten Steuerleistungen in den Szenarien, also guten Lösungswegen für die komplexen Probleme kommen wollen. Die Beachtung der Kriterien gelingt unterschiedlich gut, was auch abhängig davon ist, wieviel über die Kriterien bekannt ist. Ziele sind oftmals bekannt oder müssen zumindest selbst generiert werden, sie liegen daher direkt auf der Hand. Auch die Intransparenz eines Systems kann meist recht schnell erkannt werden. Wie sieht es aber mit den Vernetzungen im System aus? In welche Richtung liegen die Zusammenhänge und wie stark wirken sich Dynamiken im System aus? Diese Fragen sind verbunden mit dem Wissen, was über ein System erworben und angewendet werden kann, welches Inhalt der nächsten beiden Abschnitte sein wird.

2.3 Wissenserwerb

„Menschliches Lernen ist eng mit Problemlösen verknüpft. [...] So wie die Flexibilität menschlichen Problemlösens letztlich nur auf dem Hintergrund der Lernmechanismen verstanden werden kann, die zum Erwerb des im Rahmen dieser Leistungen genutzten Wissens führen [...]“ (Plötzner & Spada, 1995, S. 383)

Einige dieser wichtigen Lernmechanismen, auf die Plötzner und Spada hinweisen, sollen in

diesem Abschnitt etwas genauer betrachtet werden. Komplexe Probleme gehören zu den Verfahren, bei denen vorhandenes Wissen aktiviert werden soll und muss, beispielsweise bei analogienbasiertem Lernen. Es sind aber auch Problemstellungen denkbar, bei denen alles Wissen aus der Fragestellung abgeleitet werden muss, zunächst also kein Vorwissen nötig ist. Gut beschrieben werden diese vorwissensfreien Lernmechanismen inzwischen für regelbasierte Produktionssysteme (Anderson & Lebiere, 1998). Hauptaugenmerk soll in diesem Abschnitt jedoch auf dem analogienbasierten Wissenserwerb liegen.

Beim maschinellen Lernen, also bei dem Forschungsansatz, der Lernmechanismen so genau beschreibt, dass ihre Umsetzung in Rechnerprogramme möglich wird, unterscheidet man *vorwissensarme* und *vorwissensintensive* Verfahren. Probleme wie der „Turm von Hanoi“ oder auch die Verfahren zum Konzepterwerb (Bruner, Goodnow & Austin, 1967; Greeno & Simon, 1984; Hunt, 1962) sind vorwissensarme induktive Lernverfahren. Bei ihnen wird das Wissen aus den Instruktionen oder den Beschreibungen der Beispiele induktiv erschlossen. Diese Beschreibungen enthalten bei den vorwissensarmen Problemen soviel Informationen, dass kein weiteres bereichsspezifisches Wissen vorhanden sein muss. Plötzner und Spada nennen diese Verfahren *ähnlichkeitsbasiert*. Sie erklären: „*Um den Raum möglicher Hypothesen geeignet einschränken zu können, erfordert fehlendes bereichsspezifisches Wissen bei induktiven Lernverfahren [...] oftmals die Vorgabe großer Beispielmengen.*“ (Plötzner & Spada, 1995, S. 385).

Demgegenüber sind *analogienbasierte* Verfahren ein Beispiel für eine Klasse wissensintensiver deduktiver Lernmechanismen (mehr zum Analogisieren in Abschnitt 'Analogien als Strategie zum komplexen Problemlösen' auf Seite 38). Bereichsspezifisches Wissen kann in konstruktiven Deduktionsverfahren auf unterschiedliche Art und Weise Anwendung finden. So kann es etwa dazu genutzt werden, um plausible Hypothesen auf der Grundlage einiger weniger Beispiele zu konstruieren und dann bestimmte Hypothesen zu falsifizieren und von einer weiteren Betrachtung auszuschließen. Bereichsspezifisches Wissen kann auch zur Formulierung von Erklärungs- und Kausalhypothesen herangezogen werden (Plötzner & Spada, 1995, S. 385). Erworbenes Wissen über Systemzusammenhänge kann dementsprechend genutzt werden, hypothesengeleitet vorzugehen und Strukturen zu prüfen. Erforscht werden solche Erkenntnisse mit kognitiven Modellierungssystemen, wie z.B. den regelbasierten Produktionssystemen.

Produktionssysteme dienen mit ihren genau zu spezifizierenden Regelwerken als Theorien für das analogienbasierte Lernen und Problemlösen. Ein Beispiel eines solchen Modellierungssystems ist PUPS (Penultimate Production System; Anderson, 1989; Anderson

& Thompson, 1989). PUPS stellt eine Weiterentwicklung der Teile des Produktionssystems ACT* dar (Adaptive Control of Thought, Anderson, 1983), die sich mit Analogien beschäftigen. In PUPS werden Beispiele als deklarative Strukturen in Form von Schemata encodiert. Jedes Schema umfasst eine Menge von Merkmalen, die Kategorien zugeordnet werden müssen. Ziel ist es, möglichst abstrakte Prinzipien, die allgemein einsetzbar sind, aus den konkreten Beispielen zu abstrahieren. Dazu werden unter anderem zwei Verfahren postuliert: die *Konstanteneliminierung* und die *funktionale Elaboration*. Durch die Konstanteneliminierung werden Merkmale, die eine bestimmte „Form“ haben, wie z.B. alle ganzen Zahlen, zu Kategorien generalisiert. Beispielsweise wird bei mehreren Additionsaufgaben, wie „addiere 2 und 3“ ($+ 2 3$) und „addiere 6 und 8“ ($+ 6 8$), das allgemeine Prinzip aufgestellt: „addiere x und y“ ($+ x y$). Wird nun die Aufgabe abgewandelt in „multipliziere 6 und 8“ ($* 6 8$) nutzen die Additionsbeispiele nichts mehr. Die Lösung dieser Aufgabe muss aus dem Vorwissen des Lernalers mit dem Verfahren der *funktionalen Elaboration* abgeleitet werden. In diesem Fall führt es dazu, dass der Operator „+“ durch den Operator „*“ ersetzt wird. Bei erfolgreicher Anwendung wird das allgemeine Prinzip zu folgender Regel weiter abstrahiert:

<p style="text-align: center;">Wenn</p> <p style="text-align: center;">zwei ganze Zahlen x und y mit einer Funktion f zu verknüpfen sind und der Operator Op die Funktion f implementiert,</p> <p style="text-align: center;">dann</p> <p style="text-align: center;">benutze die Form (Op x y).</p>
--

Beim Wissenserwerb über Analogien müssen beide Verfahren ineinander greifen, das induktive wie auch das deduktive Lernverfahren. Im obigen Beispiel kann nur aufgrund von mathematischem Vorwissen auf die Verbindung zwischen den Operatoren der Addition und der Multiplikation deduktiv geschlossen werden und die neue „Wenn-Dann“-Regel als allgemeines Prinzip aufgestellt werden.

Der Ablauf der Verfahren bei den Produktionssystemen gibt ein recht genaues Bild des Wissenserwerbs bei Versuchspersonen wieder, wie am Beispiel von fehlerhaften Problemlösungen von Schülern in verschiedenen Bereichen demonstriert werden konnte (dieses und weitere Beispiele zum kognitiven Modellieren siehe Anderson, 1989).

Beim komplexen Problemlösen werden aus dem bereichsspezifischen Vorwissen Hypothesen deduktiv abgeleitet und aus den Beschreibungen und dem Prozess der Veränderung des Systems (zur Explorationsphase siehe Seite 61) der Lernzuwachs induktiv erschlossen.

Wurde hier beschrieben, welche Unterschiede es beim Lernen und damit beim Wissenserwerb mit unterschiedlichen Problemstellungen gibt, soll nachfolgend etwas genauer auf das Wissen eingegangen werden, welches speziell beim Lösen komplexer dynamischer Szenarien genutzt werden muss.

2.4 Wissensanwendung

Zwei Arten des Wissens spielen eine zentrale Rolle beim komplexen Problemlösen, zum einen das Eingriffswissen und zum anderen das Strukturwissen (Preußler, 1998 S.219).

Unter Eingriffswissen versteht man die Kenntnis von spezifischen Steuerungsprozeduren. Es liegt in „Wenn-Dann“-Regeln vor und wird vor allem dann erworben, wenn in der so genannten Explorationsphase, die der Erkundung des Systems dient, bereits die Ziele vorgegeben werden (Vollmeyer, Burns & Holyoak, 1996).

Dieses Wissen ist hilfreich, um bekannte Ziele anzusteuern, führt jedoch bei leichter Veränderung von Systemvariablen oder bei veränderten Zielen nicht mehr zum Erfolg. Es scheint sogar so zu sein, dass Eingriffswissen bei leicht veränderten Systemen zu verminderten Ergebnissen führt, also scheint es in solchen Transferaufgaben sogar zu stören (Vollmeyer et al., 1996).

Das *Strukturwissen* bezieht sich auf die Variablen des Systems und deren kausales Wirkungsgefüge (Putz-Osterloh, 1993b). Es beinhaltet also Wissen darüber, welche Variable auf welche wirkt, in welcher Richtung und wie stark. Diese drei Arten des Strukturwissens (Funke, 1992) bauen aufeinander auf. So muss, wer die *Wirkstärke* kennen will, die *Wirkrichtung* kennen und mit dieser muss dann auch der *Wirkzusammenhang* zweier Variablen feststehen.

Es konnte gezeigt werden, dass explizites, also verbalisierbares Strukturwissen die Steuerung von Systemen erleichtert, wenn das Strukturwissen anwendungsorientiert erworben wurde (Preußler, 1998). Berry und Broadbent zeigten in mehreren Experimenten den Einfluss von Wissensvermittlung über ein System, bevor es Versuchspersonen steuern mussten (Berry & Broadbent, 1984). Dabei stellte sich heraus, dass die Wissensvermittlung zwar zu mehr Strukturwissen führte, jedoch nicht zu besseren Steuerleistungen. In einzelnen Experimenten ergab sich sogar ein negativer Zusammenhang zwischen Wissensvermittlung und Steuerleistung. Erst als sie die Versuchspersonen bat, während der Steuerung laut zu denken, bewirkte dies den erwünschten Effekt, dass die Wissensvermittlung zu mehr Strukturwissen und besseren Steuerleistungen führte. Preußler folgert daraus, dass „[...] vorab vermitteltes Wissen über ein System nur dann für die Steuerung hilfreich [ist], wenn

dafür gesorgt wird, daß dieses Wissen auch implementiert und bei der Steuerung genutzt wird.“ (Preußler, 1998, S. 220).

Nach Funke (2001) zeigen eine ganze Reihe weiterer Studien eine positive Korrelation zwischen Erwerb von Strukturwissen und Steuerungsleistung (Beckmann, 1994; Beckmann und Guthke, 1995; Funke, 1993; Kersting, 1999; Müller, 1993; Preußler, 1996, 1998; Sanderson, 1989; Süß, 1999, zitiert nach Funke, 2001, p. 81).

In der vorliegenden Arbeit ist, wie von Preußler (1998) gefordert, von anwendungsorientiertem Wissen auszugehen, da sich die Versuchspersonen ihr Wissen während der Explorationsphase durch aktives Eingreifen selbst erwerben mussten. Wie viel Wissen dabei jeweils genau vorliegt, ist Aufgabe einer geeigneten Wissensdiagnostik.

2.5 Wissensdiagnostik

Neben der Frage des Wissenserwerbs ist für die komplexe Problemlöseforschung die Frage der richtigen Wissensdiagnostik von zentraler Bedeutung. Dabei soll möglichst präzise der aktuelle Stand des Eingriffswissens und des Strukturwissens der Versuchsteilnehmenden abgefragt werden. Dabei liegt die Schwierigkeit darin, eine angemessene Diagnostik zu finden, die das Wissen der Versuchsperson abbildet, ohne durch die Art der Abfrage Wissen zu generieren, das ohne das Diagnostikinstrument nicht vorläge.

Da das *Eingriffswissen* direkt mit der Steuerleistung zusammenhängt, ist eine Diagnostik hier etwas leichter. Eingriffswissen kann ja nur erworben werden, wenn Ziele vorgegeben sind, so kann man einen numerischen Wert errechnen, der den Abstand vom Systemzustand zum Zielwert angibt. Auch eine Prozessdiagnostik ist denkbar, indem man den Vorschlag von Beckmann (1994) aufgreift und den aktuellen Systemzustand in die Berechnung aufnimmt. Er schlägt vor, den optimalen Eingriffsvektor zu berechnen und die Abweichung von diesem Eingriffsvektor zu bestimmen. Unter Eingriffsvektor sind die Eingaben der zu verändernden Variablen zu verstehen. Die Anzahl der Variablen hängt vom jeweiligen System ab, was zu mehrdimensionalen Vektoren führt. Zwar schlägt er dies als Maß für die Bestimmung der Steuerleistung vor, doch gilt auch, dass je geringer diese Abweichung ist, desto mehr Eingriffswissen muss bei der Versuchsperson vorliegen. Deutlich wird hier der Vorteil von formalen Systemen, bei denen Ziele genau bestimmbar sind und somit auch der Weg der Zielerreichung zu jedem Zeitpunkt algebraisch festgelegt werden kann.

Viel schwieriger ist die Bewertung und Darstellung des *Strukturwissens* einer Versuchsperson. Erschwert wird die Wissensdiagnostik vor allem durch die

Repräsentationsform, in der das Wissen bei der Versuchsperson vorliegt und dann abgefragt werden soll. So kann es durchaus passieren, dass eine Versuchsperson viel Wissen über ein System hat, dieses aber in einer sehr persönlichen und individuellen Repräsentationsform für sie vorliegt, die mit derjenigen der Wissensdiagnostik nicht übereinstimmt. In einem solchen Fall muss die Versuchsperson noch die geistige Arbeit leisten, ihr individuell repräsentiertes Wissen in die Abfrageform der Wissensdiagnostik zu transformieren. Dieser Arbeitsschritt ist wiederum fehleranfällig und macht die Entscheidung für eine Wissensdiagnostik um so wichtiger.

An dieser Stelle soll auch auf neuere Arbeiten von Wolfram Rollett hingewiesen werden, die den Anspruch haben festzustellen, wie viel Strukturwissen eine Versuchsperson sich überhaupt angeeignet haben kann. Dazu werden alle Eingaben analysiert und auf ihren Informationsgehalt hin überprüft (Rollett, 2002, in Vorb.).

Von den verschiedenen Verfahren, die zur Wissensdiagnostik eingesetzt werden, sollen hier zwei vorgestellt werden. Zunächst die Kausaldiagramme, die eine anschauliche Form darstellen, das Strukturwissen graphisch abzubilden. Kausaldiagramme wurden vor allem durch Funke oder Müller (Funke, 1985; Funke & Müller, 1988; Müller, Funke & Rasche, 1988) zur Wissensbestimmung benutzt. Zum zweiten sollen die Paarvergleiche dargestellt werden, die in etwas abgewandelter Form in dem Experiment zum Einsatz kamen, auf dem diese Diplomarbeit basiert.

2.5.1 Kausaldiagramme

Mit Kausaldiagrammen lassen sich alle Strukturen, die in einem komplexen dynamischen System bestehen, welches auf Basis von linearen Strukturgleichungssystemen umgesetzt wird, graphisch darstellen. Wie in Abbildung 4 zu sehen ist, werden Variablen durch ihre Namen in Kästchen dargestellt und Verbindungen zwischen ihnen als Linien. Weiß eine Versuchsperson auch die Richtung der Verbindung, so wird dies durch einen Pfeil markiert. Eine Zahl an einem solchen Pfeil zeigt die Stärke der Verbindung an.

So lassen sich auch

- *Seiteneffekte*, eine Eingabevariable (links) wirkt auf zwei Ausgabevariablen (rechts), wobei der Seiteneffekt wesentlich kleiner ist als der Haupteffekt,
 - *Nebenwirkungen*, ein Verbindungspfeil zwischen zwei Ausgabevariablen, und
 - *Eigendynamiken*, eine Ausgabevariable wirkt auf sich selbst, dargestellt durch einen gebogenen Pfeil,
- darstellen.

Kausaldiagramm zum Sinus-System

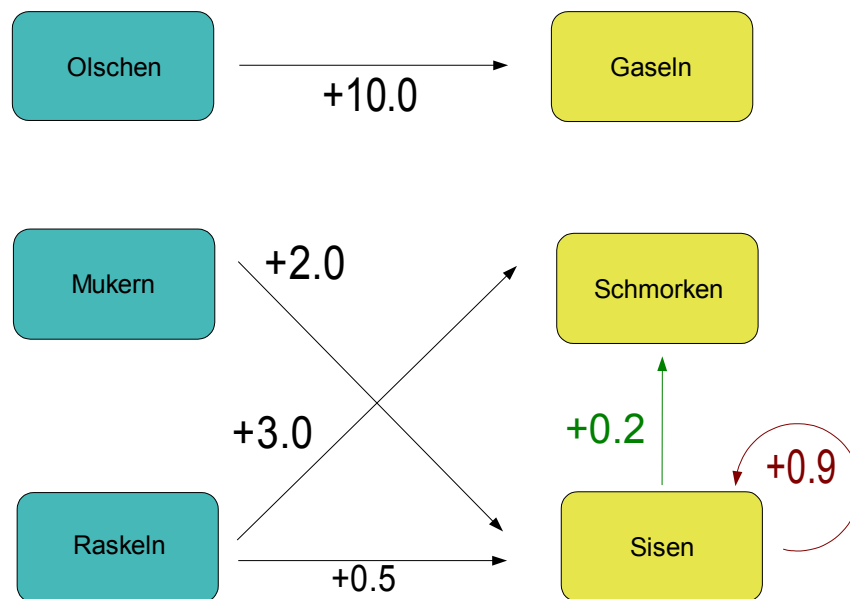


Abbildung 1: Kausaldiagramm des DYNAMIS-Systems "Sinus".

Je mehr Verbindungen innerhalb eines Systems bestehen, desto größer ist die Vernetztheit (Dörner, 1992, 1995) und um so schwieriger ist es, das System zu verstehen und zu steuern.

2.5.2 Paarvergleich

Beim Paarvergleich werden am Bildschirm jeweils zwei Variablenamen präsentiert, und die Versuchspersonen sollen entscheiden, ob zwischen den Variablen im System eine Beziehung besteht oder nicht. Dadurch lässt sich nur Relationswissen, also Wissen über einen bestehenden direkten Einfluss, überprüfen, nicht jedoch Wissen über Richtung und Stärke dieser Beziehung. Angewendet wurde dieses Verfahren beispielsweise von Preußler (1998). Eine differenziertere Wissensdiagnostik über Kausaldiagramme war in Preußlers Experimenten nicht sinnvoll, da die Experimentalgruppe mit Kausaldiagrammen vertraut war, die Kontrollgruppe jedoch nicht (siehe Preußler, 1998, S. 223).

In der experimentellen Untersuchung dieser Diplomarbeit soll eine Mischung aus beiden Verfahren gewählt werden. Dabei muss auf eine graphische Darstellung wie bei den Kausaldiagrammen verzichtet werden, da diese mit dem Treatment der Analogien konfundiert wäre. Die graphische Darstellung könnte als visuelle Analogie zu Verzerrungen führen. Dennoch sollen wie bei den Kausaldiagrammen alle Verbindungen dargestellt werden, die im System möglich sind. Dies geschieht über eine systematische Darstellung aller Paarbeziehungen. Die Versuchsperson wählt zum jeweiligen Paar über Drop-down-Menüs

aus, ob eine Beziehung besteht, welcher Art sie ist und kann die Stärke der Verbindung angeben (siehe Abbildung 2).

Variable	Zusammenhang_zu	Richtung	Stärke
A	A	keinen	.
	B	keinen	.
B	A	+	0.25
	B	+	1.1
I	A	keinen	.
	B	+	1.0
II	A	-	1.0
	B	-	2.0

Abbildung 2: Screenshot aus DYNAMIS: Das Testfenster für die Wissensdiagnostik zur Überprüfung des Strukturwissens.

So wichtig wie es ist, Denken und Problemlösen zu definieren und auch Wissen über Systeme strukturieren zu können, so wichtig ist es auch für den Forschungsbereich mit dynamischen Systemen, diese genau zu klassifizieren und vergleichbar zu machen. Nachfolgend daher ein kurzer Abschnitt über zwei Klassifikationssysteme, bevor das Programm, welches als Werkzeug für die komplexe Problemlöseforschung dient, beschrieben wird.

2.6 Taxonomien

Bereits 1976 hat Dörner eine zweidimensionale Klassifikation für Probleme der Denkpsychologie entworfen. Dabei unterscheidet er den Bekanntheitsgrad der *Mittel* und die Klarheit der *Zielkriterien*. Sind die Mittel und die Ziele bekannt, handelt es sich um ein Interpolationsproblem, meist geht es dann um die richtige Reihenfolge der Operatorenwahl, wie z.B. beim Schach. Sind hingegen die Operatoren bekannt, das Ziel allerdings nicht, wie z.B. bei dem Problem „Verschönerung des Zimmers“, so handelt es sich um ein dialektisches Problem. Eine Synthesebarriere liegt den Problemlösenden dann vor, wenn das Ziel bekannt ist, die Mittel jedoch nicht, wie z.B. bei der alchemistischen Fragestellung nach der Herstellung von Gold aus Blei (mehr zu Dörners Taxonomie, siehe Funke, 2003).

Eine Taxonomie speziell zum Umgang mit dynamischen Systemen wurde von Funke vorgeschlagen (Funke, 1990) und zur Zuordnung vieler experimenteller Befunde genutzt

(Funke, 1995). Die drei zu beschreibenden Dimensionen im Bereich des komplexen Problemlösens sind die *Personenmerkmale*, die *Situationsmerkmale* und die *Aufgabenmerkmale*.

- Innerhalb des Personenmerkmale unterscheidet Funke *kognitive Merkmale*, wie Gedächtnisstrukturen und Intelligenz, von *emotionalen und motivationalen Merkmalen* und *Persönlichkeitsmerkmale* im engeren Sinne, also so genannte „traits“ wie Ängstlichkeit, Selbstsicherheit etc.
- Bei den Situationsmerkmalen geht es um Kontextmerkmale des Systems, wie die *Transparenz* und die *Art der Aufgabenstellung*.
- Unter die Aufgabenmerkmale fallen die *formalen* und *inhaltlichen Aspekte* des Systems.

Diese beiden Klassifikationen wurden herausgestellt, da die Klassifikation Dörners eine sehr zentrale für die Anfänge der komplexen Problemlöseforschung war und die Taxonomie von Funke genutzt werden soll, um das Experiment dieser Arbeit im folgenden Abschnitt eindeutig zu beschreiben. Weitere Klassifikationen finden sich bei Funke (2003).

2.7 Der DYNAMIS-Ansatz

Ausgehend von Simulationen wie 'Tanaland' (Dörner, Drewes & Reither, 1975), 'Lohhausen' (Dörner, Kreuzig, Reither & Stäudel, 1983) und dem 'Tailorshop' (Putz-Osterloh & Lüer, 1981) wurden in den Folgejahren eine Vielzahl komplexer Szenarien entwickelt. Sie waren zunächst aufgrund ihrer hohen Realitätsnähe und der damit einhergehenden Augenscheinvalidität sehr beliebt.

Problematisch an der Vielzahl der Szenarien ist jedoch deren mangelnde Vergleichbarkeit und damit auch die Vergleichbarkeit der experimentellen Ergebnisse. Musste man als Bürgermeister von Lohhausen noch mit dem Versuchleiter kommunizieren, um die Ausprägungen der Variablenstände zu erfahren, gibt es heute noch immer eine Vielzahl von Szenarien, die nur über Tastatur gesteuert werden. Nach und nach werden diese von Maussteuerungen und einfachen Menüsteuerungen abgelöst. Neben der Interaktion mit dem System gibt es viele solcher Variablen, die dazu beitragen, dass empirische Studien über Problemlösen mit verschiedenen Systemen nicht miteinander verglichen werden können. Es ergeben sich so genannte *lokale Effekte*, die in einem System auftauchen, aber in anderen Systemen nicht repliziert werden können. Das führte soweit, dass in einer Untersuchung Intelligenz für die Problemlöseleistung keine Rolle spielte (siehe Dörner et al. 1983) und in anderen ein positiver Zusammenhang zur Problemlöseleistung nachgewiesen wurde (siehe

Beckmann & Guthke, 1995; Süß, 1996, 1999).

Bereits die Problemlöseleistung stellt sich in vielen Fällen als äußerst problematisch dar, da aufgrund der gewünschten Polytelie gar nicht genau gesagt werden kann, was eine optimale Lösung darstellt. Was ist für die Zufriedenheit der Bürger von Lohhausen wichtiger, die Bevölkerungszahl, die Einkommenshöhe, die Wohnungssituation oder die Anzahl der Grünflächen?

Zwei Möglichkeiten bieten sich an, diese Probleme einzuschränken. Zum einen bietet eine Taxonomie die Möglichkeit, das verwendete Szenario einzuordnen und viele relevante Variablen zu beschreiben und somit vergleichbarer zu machen. Zum zweiten bieten Simulationen, die verschiedene Szenarien mit gleicher Oberfläche und Steuerung herstellen können, die nötige Vergleichbarkeit.

Die Forderung nach vergleichbaren Systemen, die nicht zu lokalen Effekten führen, formulierte Funke bereits 1986. Funke zeigt Kriterien auf, die von dynamischen Systemen eingehalten werden sollten, darunter z.B. dass das Systemverhalten dem Untersucher bekannt sein muss, dass ein Lösungskriterium vorliegen muss und wie es erreicht wird (1986, S.5). In seinem Bericht werden die Grundlagen für dynamische Systeme sowohl inhaltlich wie auch mathematisch beschrieben. Es wird ein Forschungsprogramm gefordert, welches unter anderem die Entwicklung eines Instruments zur Konstruktion beliebiger dynamischer Systeme mit *wohlbekannten Eigenschaften* zum Ziel hat (Funke, 1986, S.8).

Unter wohlbekannten Eigenschaften wird verstanden, dass die eingesetzten dynamischen Systeme mathematisch exakt zu beschreiben sind. Dies führt dazu, dass zum einen eine Optimallösung angegeben werden kann und zum anderen zu jedem Zeitpunkt der bestmögliche Eingriff im System bestimmt werden kann. Dadurch können Problemlösende in ihrer Steuerleistung genau verglichen werden. Neben Funke fordern viele weitere Autoren, Systeme zu verwenden, die einem formalen Rahmen folgen und so wohlbekannte Eigenschaften haben (Buchner, 1999, Preußler, 1998).

Als formale Systeme für komplexe Problemlöseszenarien bieten sich die Theorie der finiten Automaten (mehr zum Einsatz finiter Automaten beim Problemlösen bei Buchner, 1999) und die linearen Gleichungssysteme an (siehe z.B. Vollmeyer & Funke, 1999).

Aus den Forschungsbestrebungen der Gruppe um Funke ging das Instrument DYNAMIS (Version 2.0 vom 25.07.1986 wurde in Turbo-Pascal 3.01 geschrieben) hervor, welches im Detail von Funke, Fahnenbruck und Müller (1986) beschrieben wird. Es basiert auf einem linearen Gleichungssystem und sorgt dadurch für wohlbekannte Eigenschaften. Obwohl das Programm den damaligen technischen Anforderungen entsprach, wurde es mit den rasanten

Veränderungen im Bereich der Betriebssysteme immer schwieriger, dieses noch auf DOS basierende Programm einzusetzen.

Im Rahmen eines Projekts des Psychologischen Instituts der Universität Heidelberg wurde dieses Programm bei Beibehaltung der Anforderungen aus dem Forschungsprogramm modernisiert. Ziel war es, ein plattformunabhängiges System zu programmieren, das auch eine Nutzung über das Internet ermöglichen sollte. Dies wurde umgesetzt bis Mitte 2002. Nun liegt ein in Java programmiertes Applet vor, das über das Internet auf einem Server angesteuert werden kann und per Client-Socket-Verbindung auf jedem beliebigen Rechner laufen kann. Am Ende einer solchen Verbindung werden die Daten an den Server geschickt und dort gespeichert.

Grundsätzlich können mit dem DYNAMIS-Ansatz ganz verschiedene Systeme (nachfolgend auch Szenarien genannt, um vom DYNAMIS-System abzugrenzen) simuliert werden. Sowohl das Szenario „Sinus“ als auch das Szenario „WITS“ wurden darin umgesetzt. Bei „Sinus“ handelt es sich um einen fiktiven Planeten, auf dessen Lebewesen Einfluss genommen werden kann. Durch den fiktiven Planeten mit fiktiven Lebewesen kann nicht auf Vorwissen zurückgegriffen werden. Das Szenario „WITS“ besteht gänzlich aus abstrakten Variablenbezeichnungen, um nicht einmal das Vorwissen an sich vermehrende Lebewesen zuzulassen. Die zugehörigen Kausaldiagramme sind auf den Seite 39 bzw. 59 zu sehen. Innerhalb von DYNAMIS haben die Szenarien die gleiche Computeroberfläche und lassen sich per Tastatur und Maus (seit der JAVA-Version) steuern. Unterschiede gibt es immer noch viele. Einige Beispiele sind die Anzahl der Variablen, deren Verbindungen, die Anzahl der möglichen Eingriffe pro Durchgang und die Transparenz (wieviele Eingriffe sind gleichzeitig sichtbar). Auch die Wissensdiagnostik ist (noch) nicht einheitlich auf eine Form beschränkt.

Wie das Szenario der vorliegenden Untersuchung aussieht, beschreibt der Abschnitt 'Beschreibung und Einordnung des DYNAMIS-Systems „WITS“' auf Seite 58. Im letzten Abschnitt des Kapitels über das Problemlösen mit dynamischen Systemen soll es um Emotionen und Motivationen gehen, wie sie während der Bearbeitung dynamischer Szenarien vorliegen bzw. wie sie notwendigerweise zur Bearbeitung gebraucht werden.

2.8 Emotion und Motivation

In der Kognitionspsychologie liegen inzwischen theoretische Ansätze vor, welche die Forschungsergebnisse der Motivations- und Emotionsforschung mit einbeziehen. Ein solcher Ansatz findet sich im „Bauplan zu einer Seele“ von Dörner (1999). Leider liegt speziell zum

komplexen Problemlösen noch keine dieser theoretischen Integrationen vor, die es ermöglichen, Vorhersagen daraus abzuleiten.

Die folgende Darstellung bezieht sich daher hauptsächlich auf die schon lange bekannten Forschungsergebnisse zur Motivation beim allgemeinen Problemlösen, die aber durchaus auch auf das komplexe Problemlösen übertragen werden können.

Aktivationsniveau.

Ein inzwischen gut untersuchter Befund in der Motivationspsychologie ist das Yerkes-Dodson-Gesetz, dass für jede Aufgabe einen optimalen Level der physiologischen Erregung, also allgemeiner Aktivierung des Organismus, annimmt (Yerkes & Dodson, 1908; zitiert nach Mietzel, 2001).

Die Funktion zwischen Aktivationsniveau und der Qualität der Problemlöseleistung gleicht einer umgekehrten U-Funktion. Mit zunehmender Aktivierung steigt die Qualität der Problemlöseleistung an, um dann wieder abzufallen, wenn der optimale Aktivitätslevel überschritten ist. Hebb (1955) erläutert diese Glockenfunktion in der Weise, dass Menschen, die sich in einem mittleren Aktivationsniveau befinden, ein vergleichsweise hohes Maß an Aufmerksamkeit auf eine vorliegende Aufgabe richten können. Bei einer zu geringen Aktivierung ist mit einer unverhältnismäßig großen Anzahl von Flüchtighkeitsfehlern zu rechnen, während bei einer zu hoher Aktivierung ein Teil der Aufmerksamkeit auf den Zustand des eigenen Körpers gelenkt wird (Mietzel, 2001, S. 386).

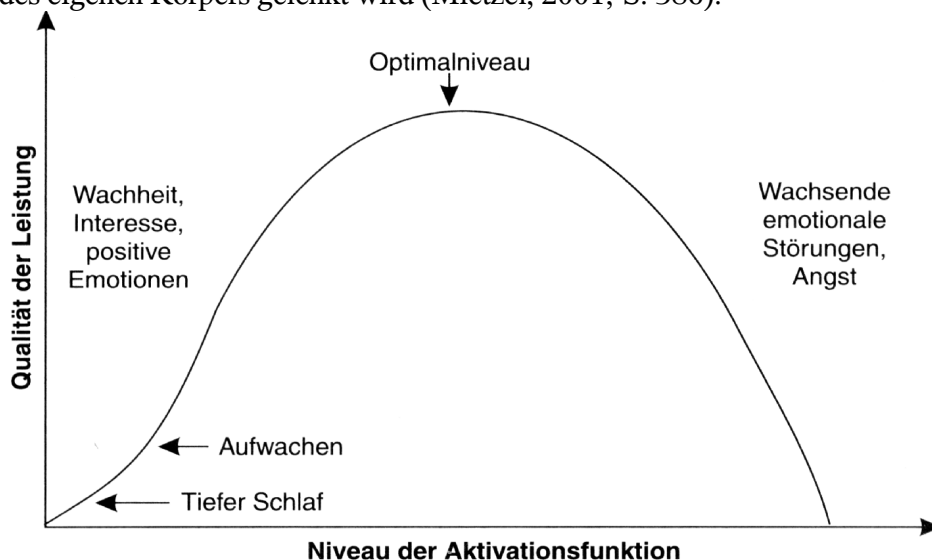


Abbildung 3: Die Aktivationsfunktion nach dem Yerkes-Dodson-Gesetz (aus Mietzel, 2001, S. 386).

Eine so beschriebene Kurve ist auch für das komplexe Problemlösen mit einem DYNAMIS-System anzunehmen. Zum einen wird das Aktivitätsniveau der Versuchsperson von der

allgemeinen Motivation bestimmt, die sowohl von der Hoffnung auf Erfolg als auch der Angst vor Misserfolg (vgl. Atkinson, 1964) und vom Anreiz der Aufgabenstellung abhängt. Zum anderen wird die Aktivationsfunktion durch die Attribute der Aufgabe und deren Wirkung auf den „optimalen Aktivitätslevel“ bestimmt. Eine der wenigen Aussagen zur Wirkung von Aufgabenattributen findet sich bei Bourne, Ekstrand und Dominowski. Sie führen an, dass je komplexer die Aufgabe ist, desto mehr (falsche) Antworten werden im Vergleich zu richtigen erzeugt. Daher sei eine geringere Ängstlichkeit der Versuchsperson besser. Sie zeigten in Experimenten, dass die Neigung, ängstlich zu werden, als Prädiktor für die Qualität der Problemlöseleistung gelten kann (Bourne, Ekstrand & Dominowski, 1971, p. 238 bzw. 101). Aus dem Yerkes-Dodson Gesetz kann auch gefolgert werden, dass großer Druck und empfundener Stress die Leistung schmälert. Ebenso tragen Fehler und Misserfolge im Verständnis des Problems zum Versagen in der Aufgabe bei. Bourne et al. nehmen an, dass dies aufgrund erhöhter Frustration und kognitiver Kapazitätseinbußen erfolgt, da negative Antworten neue Lösungswege fordern und damit weitere Überlegungen auslösen. Dies nennen sie „mitstreitende Prozesse“, die Aufmerksamkeit und Konzentration von der eigentlichen Aufgabe abziehen (Bourne, Ekstrand & Dominowski, 1971, p. 102).

3. Stile, Strategien und Taktiken

In diesem Kapitel werden verschiedene Strategien des komplexen Problemlösens beschrieben. Zuvor soll jedoch der Begriff der „Strategie“ von dem in dieser Arbeit neu eingeführten Begriff der „Taktik“ unterschieden werden.

Was unter „Strategie“ verstanden werden soll, ist schwer zu definieren, und selbst innerhalb der Problemlöseforschung existieren unterschiedlich breite Konzepte. Ein sehr enges Konzept findet sich bei Rollett (2002), der den Begriff für vier „Basisstrategien“ verwendet, die zwei bis neun verschiedene Eingriffsmuster beinhalten. Bei Vollmeyer, Burns und Holyoak (1996) findet sich ein nur z.T. breiteres Konzept, das von einem einzelnen Systemeingriff (VOTAT) bis hin zur „Strategie“ der Reduzierung des jeweils größten Abstandes von Ist- und Soll-Zustand reicht. Beide Konzepte entsprechen nicht dem, was in dieser Arbeit unter dem Begriff „Strategie“ verstanden werden soll. In der hier gewählten Terminologie würden beide Konzepte als „Taktiken“ (siehe Abschnitt 3.3) bezeichnet werden.

Damit ist aber das Problem der Definition des Strategiebegriffs noch nicht gelöst, und auch das psychologische Wörterbuch (Dorsch, Häcker & Stapf, 1996) enthält leider keine eindeutige Definition zum Begriff „Strategie“, sondern verweist auf spezielle Bereiche wie die Denkstrategie, Lernstrategie, Entscheidungsstrategie oder Spielstrategie. Im Duden (Drosdowski, Müller, Scholze-Stubenrecht & Wermke, 1990) findet sich hingegen folgende Definition von Strategie:

„Strategie: genauer Plan des eigenen Vorgehens, der dazu dient, ein militärisches, politisches, psychologisches o. ä. Ziel zu erreichen, u. in dem man diejenigen Faktoren, die in die eigene Aktion hineinspielen könnten, von vornherein einzukalkulieren versucht.“ (S. 745)

Im Gegensatz dazu sind Taktiken als wesentlich kurzfristiger definiert:

„Taktik [gr.-fr.; „Kunst der Anordnung u. Aufstellung“][...] 2. auf genauen Überlegungen basierende, von bestimmten Erwägungen bestimmte Art u. Weise des Vorgehens, berechnendes zweckbestimmtes Verhalten.“(S. 765).

Aufgrund der Uneindeutigkeit, die auch nach diesen Definitionen noch bleibt, soll an dieser Stelle eine Verortung des Begriffs über die Bestimmung von zwei Dimensionen vorgeschlagen werden, zum einen über die Dimension der Bereichsspezifität oder auch Domänenspezifität und zum anderen über den Wirkungszeitpunkt von strategischen Interventionen.

Domänenspezifität.

Mit Domäne oder Bereich ist das jeweilige Wissensgebiet gemeint, in das ein Problem eingebettet ist und in dem nach einer Lösung gesucht wird. Strategien haben entweder einen sehr allgemeinen Charakter und sind über verschiedene Situationen hinweg einsetzbar, oder sie lassen sich nur innerhalb enger Handlungsbereiche einsetzen. Die letztgenannten, meist wesentlich effektiveren Strategien bezeichnet Robertson (2001, p. 37) mit „strong methods“. Sie sind bereichsspezifisch und dadurch gekennzeichnet, dass sie fast immer zu Ergebnissen führen. Die „weak methods“ hingegen zeichnen sich dadurch aus, dass sie auch dann helfen, wenn Problembereiche neu sind und noch kein Wissen vorliegt. Dabei sind sie vielleicht nicht unbedingt besonders effektiv, aber bereichsübergreifend einsetzbar. Für diese beiden Methodenarten werden auch oft die Begriffe Algorithmus versus Heuristik verwendet oder auch domänenspezifisch versus allgemein einsetzbar. Den unterschiedlichen Wert solcher starker bzw. schwacher Strategien beschreibt Nickerson wie folgt:

„Generally speaking, there appears to be a trade-off between range of applicability and power. The more widely applicable – the less domain dependent – a method or strategy is, the less powerful it is likely to be in any particular application.“ (Nickerson, 1998, p. 425).

Sehr ähnlich drückt sich auch Anderson (2000, p. 337) aus, wenn er sagt, dass sich die Produktionsregeln mit zunehmender Erfahrung in speziellen Domänen verfeinern (siehe Anderson & Lebiere, 1998) und damit Expertise gebildet wird. Diese sind aber für andere Einsatzgebiete zu spezifisch, und damit wird immer weniger Transfer auf neue Problemstellungen möglich.

Wirkungszeitpunkt.

Als zweites bietet sich eine zeitliche Dimension an: der Wirkungszeitpunkt. Darunter soll der Zeitpunkt verstanden werden, zu dem die Vermittlung der Strategie eingesetzt werden muss, damit sie ihre Wirkung entfalten kann. So gibt es basale Verhaltensweisen, wie die Lehre der normativen Logik oder auch kognitive Stile, die sich nur sehr langfristig auswirken und auch nur langsam verändert werden können. Dem gegenüber gibt es kurz- und mittelfristig veränderbare Handlungsabläufe wie Pläne und Taktiken, die sich oft und schnell ändern können. Hier reichen bereits kurze Schulungen oder einfache Instruktionsanweisungen, um Verhaltensänderungen zu bewirken.

Diese zeitliche Komponente der verschiedenen Verfahren soll als Einteilungskriterium dienen. Beschrieben werden in Abschnitt 3.1 erst die langfristigen Strategien, die auch im Sinne der Persönlichkeitstheorie von Kelly als zentrale Konstrukte der Persönlichkeit angesehen werden können (siehe Kelly, 1986 oder Pervin, 2000). Die Strategien, so wie sie in dieser Arbeit

verstanden werden sollen, werden im Abschnitt 3.2 dieses Kapitels beschrieben, und die kurzfristigen Verfahren werden im Abschnitt 3.3 unter dem Begriff der Taktiken aufgeführt. Eingeordnet nach der Dimension der Domänenspezifität, ergäbe sich die gleiche Reihenfolge mit den kognitiven Stilen als breites bereichsübergreifendes Konzept auf der einen Seite des Kontinuums bis hin zu den Taktiken als bereichsspezifisches Konzept auf der anderen. Das hier zentrale Konzept der Strategie würde den Übergang zwischen diesen beiden Polen bilden. Im Abschnitt 3.4 werden daher noch einmal abschließend die Begriffe „Strategie“ und „Taktik“ gegenübergestellt und für die Verwendung in dieser Arbeit definiert.

3.1 Stile, Metakognition, basale Operationen

Die folgenden Konstrukte, die in ihrer Vermittlung viel Zeit brauchen und meist langfristig wirken, sollen in diesem Abschnitt nur kurz dargestellt werden, da sie aufgrund ihrer mangelnden Kontrollierbarkeit im Rahmen von Laborexperimenten kaum eine Rolle spielen.

Basale Operationen.

Unbestritten ist wohl, dass grundlegende kognitive Fertigkeiten in all ihren Formen auch Einfluss auf die Problemlöseleistung nehmen. Ganz basale Operationen und Prozesse müssen beherrscht werden. Basale Operationen wie der Umgang mit Zahlen, Kommunikation oder die Fähigkeit zum Klassifizieren sind essentiell, ebenso wie auch das Treffen von Vorhersagen durch reflektiertes Denken und Schlussfolgern. Diese Operationen können sich zwischen Individuen in der Güte ihrer Anwendung qualitativ unterscheiden und dadurch Varianz der Problemlöseleistung aufklären.

Zu diesen basalen Operationen kommen auch Fähigkeiten und Dispositionen hinzu, so z.B. die Schulung im Umgang mit normativer Logik oder verschiedene Überzeugungen und Vorstellungen über die Beeinflussbarkeit von Intelligenz und Fähigkeiten. So wirkt sich die Vorstellung von Intelligenz als stabilem Faktor negativ auf die eigene Anstrengung und den Erwerb neuer Fertigkeiten aus (siehe dazu auch Dweck & Eliot, 1983).

Metakognition.

Unter metakognitiven Fähigkeiten wird verstanden, dass der Mensch zwischen verschiedenen Strategien bewusst wählen kann, dass er seine Aufmerksamkeits- und Konzentrationsfähigkeit einzuschätzen weiß. Dazu gehört auch das Wissen über Beobachten, Kontrollieren und Bewerten von Veränderungen und Eingriffen. Schon zu wissen, dass man mehrere Strategien in einem Szenario anwenden könnte, ist Metakognition. Auch die Fähigkeit, zwischen den Strategien auszuwählen, gehört dazu.

Damit nimmt das Konzept der Metakognition innerhalb psychologischer Betrachtungsweisen

des komplexen Problemlösens eine wichtige Rolle ein. Jede übergeordnete Beschäftigung mit eigenen kognitiven Leistungen fällt unter dieses Konstrukt. Das wichtige Konzept der Selbstreflexion könnte z.B. auch als Metakognition eingeordnet werden. Hier zeigt sich der fließende Übergang der langfristigen Verfahren zu den mittel- und kurzfristigen, denn Dörner hebt gerade die Selbstreflexion als wichtiges Verfahren zur Steigerung der Problemlöseleistung heraus (Dörner, 1992).

Kognitive Stile.

Interessant ist auch der Einfluss von kognitiven Stilen (Sternberg, 1997) auf das Problemlösen. Durch die Beschäftigung mit diesen Stilen werden viele interessante Fragen aufgeworfen: Wählt ein exekutiver Typ, also eher ein „Macher“ und „Ausführender“, andere Strategien und damit Vorgehensweisen als ein judikativer Typ, der eher abwägt, vergleicht und beurteilt? Lässt sich ein kognitiver Stil finden, mit dem das Lösen komplexer Probleme besser gelingt? Ist gar ein effizienter Einsatz einer Strategie von einem bestimmten Typ abhängig?

Leider ist die psychologische Forschung in vielen Gebieten (noch) nicht so weit, als dass sich Antworten auf diese Fragen oder direkte Ableitungen für die vorliegende Arbeit anbieten würden (mehr zu kognitiven Stilen siehe Sternberg, 1997).

All den in diesem Abschnitt dargestellten Bereichen ist jedoch eines gemeinsam: der mangelnde kurzfristige Einfluss auf das Problemlösen in komplexen, dynamischen Systemen. Nur ein sehr großer Forschungsaufwand mit Längsschnittstudien kann zeigen, inwieweit langfristige Verbesserungen im Umgang mit komplexen Problemen durch sie zu erzielen sind. Gemeinsam ist diesen Verfahren auch der Fokus auf den bereichsübergreifenden Einsatz.

3.2 Strategien

Im Folgenden sollen zunächst in 3.2.1 einige wichtige Strategien, wie sie in denkpsychologischen Untersuchungen oftmals beschrieben wurden, dargestellt werden. Besonderes Augenmerk soll dann im Unterabschnitt 3.2.2 auf die Strategie des Analogisierens gerichtet werden.

3.2.1 Beschreibungen verschiedener Strategien aus der Denkpsychologie

Beispiele für Strategien, wie sie in denkpsychologischen Untersuchungen beschrieben wurden, sind Subgoalung, Dekomposition, Forward Chaining, Working Backwards, Means-ends-analysis, Hillclimbing, Selection und Reflexion. Diese werden im folgenden beschrieben.

Subgoalung, Dekomposition.

Beim *Subgoalung* wird ein komplexes Problem in eine Sequenz einfacher Probleme zerlegt, die ihrerseits in Unterziele zerlegt werden können. Diese Strategie kommt bei Programmierertätigkeiten häufig zum Einsatz. Dörner beschreibt sie auch als Dekomposition, bei der aus einem Problem (eventuell mit unklarem Ziel) ein Mehrfachproblem mit vielen Teilkomponenten wird.

Die Zerlegung des Problems in Teilziele bedeutet auch oftmals die Aufstellung eines Plans, der durch die Teilziele Lösungsschritte für das Problem enthält. Tritt ein solcher Lösungsplan durch das Subgoalung nicht direkt hervor, hilft es zumindest, das Problem besser zu verstehen und zu beschreiben (Nickerson, 1998, p. 426).

Das Subgoalung wird schon seit über 30 Jahren im Zusammenhang mit dem „Turm von Hanoi“-Problem untersucht und beschrieben. Anderson betont in diesem Zusammenhang die menschliche Fähigkeit des Werkzeugbaus (Anderson, 2000, p. 314ff). So werden Ziele wie der Häuserbau in Unterziele wie die Beschaffung von Holz aufgeteilt, zu der Werkzeuge wie Sägen oder Äxte benötigt werden. Der Werkzeugbau ist so nur beim Menschen bekannt und in seltenen Fällen auch bei Affen zu beobachten (z.B. Köhler, 1927). Ganz zentrale Prozesse des Produktionssystems ACT-R von Anderson und Lebiere (1998) basieren auf den Forschungsergebnissen zum „Turm von Hanoi“ und der Subgoalung-Strategie.

Das Risiko, dass bei dieser Strategie besteht, ist die Veränderung der „Natur“ des Problems durch die Zerlegung. Oftmals ergibt die Summe der Teile nicht mehr das gesuchte Ganze. Es können Fehler sowohl bei der Zerlegung als auch bei der Zusammensetzung der einzelnen Teilkomponenten auftreten. Gerade beim Prozess des komplexen Problemlösens kann es zu Interaktionen verschiedener Variablen kommen, die in den Teilkomponenten nicht beobachtbar sind.

Forward Chaining und Working Backwards.

Wenn es Problemlösenden möglich ist, sich den gesamten Problemraum vorzustellen, können sie direkt auf die Lösung des Problems hinarbeiten. Dabei verkettet die Problemlösende vorwärtsgerichtet die einzelnen Lösungsschritte vom Ist-Zustand bis hin zum Soll-Zustand. Die mentale Repräsentation des Problems bildet dabei den zentralen Dreh- und Angelpunkt dieser „Forward Chaining“-Methode. Meist ist es nicht möglich, sich zu Beginn einer Aufgabe alle Schritte bis zum Ziel genau zu überlegen. Meist ist dies nur möglich, wenn genug Expertise im Problembereich vorliegt und somit vielfältige Wissensseinheiten mental zusammenfassbar sind, also so genannte Chunks gebildet werden können. Nickerson (1998) weist daher auch ausdrücklich darauf hin, dass diese Strategie meist von Experten verwendet

wird und Novizen eher zur „Means-end“- Strategie greifen (siehe nächster Abschnitt). Dazu verweist er auf Studien von Larkin, McDermeott, Simon und Simon (1980 a,b) und Sweller, Mawer und Ward (1983).

Fehlt der nötige Überblick über den Problemraum und ist das Ziel klar beschrieben, kann man versuchen, den Abstand zwischen Soll- und Ist-Zustand zu verkürzen, indem alle bekannten Lösungsschritte vom Ziel ausgehend zum Anfangspunkt erarbeitet werden. Dies wird als „Working Backwards“- Strategie bezeichnet. Werden Soll- und Ist-Zustand eines Problemraumes in Analogie als Punkte auf einer Karte betrachtet, verkürzt man so den Weg (die Lösung) zwischen den Punkten (Ist- und Soll-Zustand). Gerade bei der Suche nach einer geeigneten Strecke auf einer Landkarte (um bei dem Bild zu bleiben) ist es zumeist sinnvoller, vom Ziel ausgehend einen Weg zum Start zu finden als umgekehrt.

Die „Working Backwards“- Strategie wird häufig in Kombination mit anderen in diesem Abschnitt vorgestellten Strategien verwendet. Man findet sie oft in den Strategiebeschreibungen zu nicht-komplexen Problemlöseaufgaben wie z.B. dem „Turm von Hanoi“ oder den „Kanibalen und Missionaren“.

Means-end analysis.

Die Mittel-Zweck-Analyse wurde von Newell und Simon (1972) genauer beschrieben und für die Entwicklung von Algorithmen zur allgemeinen Problemlösung in Computerprogrammen eingesetzt. Ausgegangen wird von einem klaren Soll-Zustand und von einem aktuellen Ist-Zustand. Es werden genaue Listen der Differenzen aller Komponenten beider Zustände erstellt. Gehandelt wird dann in genau jenem Bereich, in dem eine Differenz dieser Liste reduziert werden kann. Nach und nach wird die Liste der Differenzen abgearbeitet. Es ist dabei egal, ob das vom Ziel ausgehend (Working Backwards) oder vom Ist-Zustand aus in Richtung Ziel erfolgt.

Bei dieser Strategie wird ähnlich der „Hill Climbing“-Strategie niemals der Abstand zwischen Ist- und Soll-Zustand vergrößert. Dies kann bei vielen Problemen in eine Sackgasse führen, aus der kein Weg mehr herausführt.

Hill Climbing.

Wie dem Namen der Strategie bereits zu entnehmen ist, wird in dieser Strategie die Lösung des Problems in Form einer Metapher als Berg bezeichnet. Problemlösende können gemäß dieser Metapher von ihrer Position aus einen Schritt in alle Richtungen tun und sich dann für den Schritt entscheiden, der nach „oben“ geführt hat. Man prüft also nach jedem Schritt, ob man „höher“ gekommen ist. Wenn dies nicht der Fall ist, wird der Schritt zurückgenommen und in eine andere Richtung unternommen, ohne jedoch jemals „abwärts zu laufen“.

Für den Problemlöseprozess heißt das, dass nach jedem Schritt geprüft wird, ob sich der Zielzustand angenähert hat. Ist das nicht der Fall, wird dieser Schritt zurückgenommen und ein anderer ausgeführt. Dabei wird aber die Differenz zum Soll-Zustand niemals vergrößert.

Das Problem dieser Strategie wird bereits in der Metapher deutlich, da es manchmal einfach notwendig ist, zunächst einige Schritte hinabzusteigen, um den Gipfel eines Berges zu erreichen. Es kann passieren, dass der direkte aufwärts führende Weg auf einem Nebenhügel endet, der eben nicht die absolute Spitze des Berges ist.

Selection sensu Bruner, Goodnow und Austin.

Bruner, Goodnow und Austin (1967) beschreiben in ihrem Buch „A Study of Thinking“ zunächst, was eine geeignete Strategie leisten muß. Dabei sind drei Punkte zu beachten (Bruner et al. 1967, p. 82):

- Die Strategie soll die Wahrscheinlichkeit erhöhen, dass Eingriffe (Operatoren) die gesuchten Informationen beinhalten,
- sie erleichtert die Anpassung neuer Informationen und
- sie reguliert das Risiko, das man eingeht, um die richtige Lösung innerhalb einer begrenzten Anzahl von Eingriffen zu finden.

Das Material für die Experimente von Bruner et al. beschränkte sich auf eng umgrenzte Problemräume, die noch dazu recht schnell von den Versuchspersonen überschaut werden konnten. Bei der „Selection“-Strategie handelt es sich daher doch eher um eine recht bereichsspezifische Strategie, die nicht ohne weiteres auf andere Domänen übertragbar ist.

Für besonders geeignet halten die Autoren die von ihnen als „Conservative Focussing“ bezeichnete Vorgehensweise, bei der ausgehend von einer Anfangshypothese stets nur ein Aspekt verändert und überprüft wird. Dieses Vorgehen wird oft auch als „wissenschaftlich“ oder „theoretisch“ bezeichnet (Klahr & Dunbar, 1988; Simon & Lea, 1974).

Selbstreflexion sensu Dörner.

Nach Dörner (1995) legt der Mensch getätigte Lösungsversuche in einer Art „Logbuch“ im Protokollgedächtnis ab. Diese abgelegten Versuche beinhalten auch ganze Strategieansätze, die fehlgeschlagen sind. Je lückenhafter dieser Teil des Gedächtnisses ist, desto schlechter kann Selbstreflexion betrieben werden. Sie ist aber nach Dörner das einfachste Mittel, um die Problemlöseleistung effizient zu steigern. Fragen wie „Sammle ich genug Informationen?“, „Von welchen Annahmen gehe ich eigentlich bei meiner Lösungssuche aus?“, „Ist das Ziel, welches ich anstrebe, eigentlich vernünftig, oder sollte ich vielleicht ein ganz anderes Ziel anstreben?“ beziehen sich auf den Problemlösungsprozess und nicht auf das Problem und können dadurch oftmals weiterhelfen (Dörner, 1995, S. 317).

Das Denken über das eigene Denken ist die Art der Selbstreflexion, die Dörner als hilfreiche Strategie beim Problemlösen ansieht. Bestätigungen seiner Annahme finden sich in Untersuchungen von Reither (1979) und Hesse (1979). Sie gaben ihren Experimentalgruppen nur die Anweisung, über das eigene Denken nachzudenken, ohne eine weitere Anleitung. Bereits diese einfache Instruktion führte zu bedeutsamen Verbesserungen der Problemlöseleistung (Dörner, 1992, S. 302).

3.2.2 Analogien als Strategie zum komplexen Problemlösen

Analogien haben in der Problemlöseforschung bereits eine lange Tradition. Das analoge Schlußfolgern findet sich daher auch in einer Reihe von Intelligenztests als einzelne Aufgaben oder auch als Subskala zum „Analogien bilden“ wieder.

Nach einer kurzen Definition des analogen Schlussfolgerns soll auf Struktur- und Oberflächenmerkmale, den Prozess des Analogisierens und die Einbettung in komplexe Problemlöseuntersuchungen eingegangen werden.

Definition.

Beim Problemlösen durch analogen Transfer wird eine bekannte Wissensdomäne (Quelldomäne) zur Lösung von Problemen in einer neuen Wissensdomäne (der Zieldomäne) gebraucht. Dazu sind zwei Prozesse zentral: der *Abruf* einer analogen Struktur aus dem Gedächtnis und das Herstellen einer *Abbildung* zwischen der abgerufenen Struktur und der Struktur des vorliegenden Problems.

Analoges Schlußfolgern.

Die Aufgaben zum analogen Schlußfolgern liegen meist in der Form $[A : B :: C : ?]$ sprich: „A verhält sich zu B wie C zu ?“ vor. Für das Fragezeichen muss eine passende Antwort gefunden werden, je nach Aufgabenstellung entweder per Multiple-choice-Verfahren oder frei erdacht.

Beispiel:

Hand verhält sich zu Fuß,
wie Finger zu ? [1. Fingernagel, 2. Zehe, 3. Ferse, 4. Schuh]

In der komplexen Problemlöseforschung beschreiben Dörner, Kreuzig, Reither und Stäudel (1983), dass beim Versuch, komplexe Probleme zu lösen, Teilziele über den Analogieschluß gebildet werden. Hussy ergänzt dazu: „*Das analoge Schlußfolgern wird in diesem Fall [...] als Mittel zur Ableitung gesuchter [...] Informationen aus dem vorhandenen Wissensbestand verwendet.*“ (Hussy, 1986, S. 39).

Strukturmerkmale.

Bei der oben beschriebenen Art von Analogiebildung kommt es auf die Strukturmerkmale von *Quelle* und *Ziel* an. So ist in diesem Beispiel die Hand mit ihren Fingern die *Quelle*, von der aus eine Analogie zu Fuß und Zehen gebildet wird. Strukturmerkmale sind in diesem Fall die Grösse, die Funktion und die Zugehörigkeit zum Körper. Diese drei Merkmale sind Strukturmerkmale erster *Ordnung*. Viele weitere Merkmale dieser Art sind denkbar. Ebenso wichtig sind aber auch die Strukturmerkmale höherer Ordnung. Das Prinzip der Ordnung soll nun kurz anhand des obigen Beispiels erklärt werden.

Oftmals kann man Strukturen wie die Regeln eines Produktionssystems verstehen. Eine Produktionsregel einfacher Ordnung wäre z.B. „Teil von (Finger, Hand)“ und würde bedeuten, dass Finger Teile der Hand sind. Auf der Suche nach der passenden Antwort für das Fragezeichen im obigen Beispiel muss die gleiche Produktionsregel auch für das Ziel der Analogie angewandt werden, also „Teil von (?, Fuß)“. Weitere Regeln dieser einfachen Ordnung sind denkbar, z.B. „Kleiner als (Finger, Hand)“. Je mehr solcher Regeln aufgestellt werden können und für Quelle und Ziel der Analogie passen, desto mehr Strukturmerkmale haben Quelle und Ziel gemeinsam. Und je mehr Strukturmerkmale von Quelle und Ziel sich gleichen, desto leichter ist die Analogie für Problemlösende oder Lernende zu erkennen.

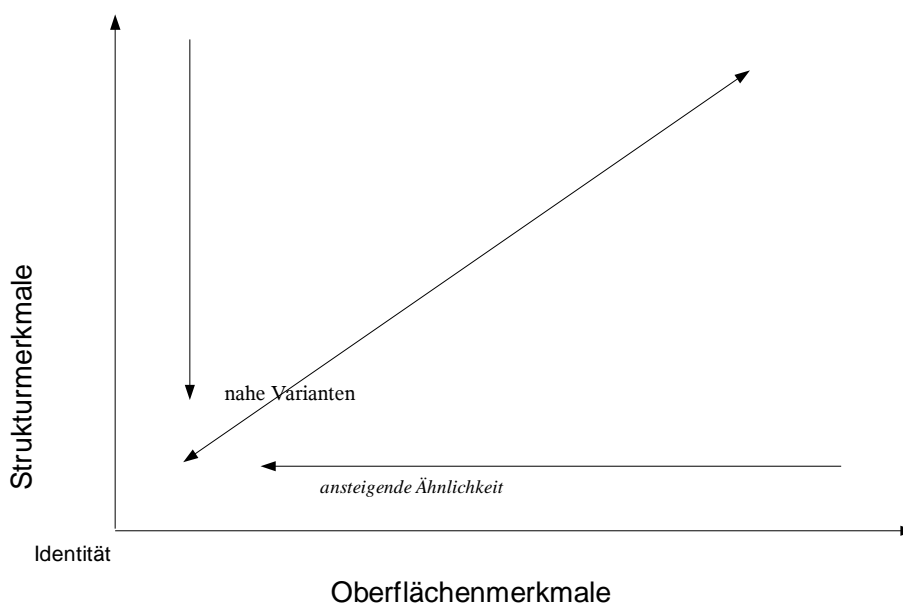


Abbildung 4: Kontinuum der Ähnlichkeit. Je stärker Oberfläche und Struktur von Ziel und Quelle voneinander abweichen, desto unähnlicher sind die Probleme, bis sie nicht mehr vergleichbar sind. Wenn Struktur und Oberfläche gleich sind (links unten), dann sind die Probleme identisch.

Da aber sprichwörtlich „jeder Vergleich hinkt“, birgt auch jede Analogie ihre Fehler, also Strukturmerkmale, die sich unterscheiden. Nur wenn Quelle und Ziel identisch sind, wird es nicht zu diesem Analogiefehler kommen (siehe Abbildung 4). Zu den Strukturmerkmalen einer Analogie gehören auch die abstrahierte Form der Aufgabenstellung und die Lösungsmöglichkeiten. Letztere können oftmals als Regeln höherer Ordnung aufgestellt werden. Produktionsregeln höherer Ordnung sind vergleichbar mit denen einfacher Ordnung, jedoch können sie diese beinhalten und Aussagen machen, die mehr als zwei Attribute (wie Hand und Finger) enthalten (Gentner, 1989, p. 208; Gentner, 1983).

Oberflächenmerkmale.

Ebenso wichtig wie die Strukturmerkmale sind die Oberflächenmerkmale der Analogie. Dabei handelt es sich um Variablen, Objekte, Namen oder auch die Art der Fragestellung, die sowohl in der Quelle als auch im Ziel der Analogie vorkommen. Deutlich werden diese Oberflächenmerkmale in folgendem Beispiel, in dem die Fragen ähnlich klingen und daher zu einem analogen Vergleich leiten, die Strukturmerkmale sich aber unterscheiden.

Beispiel:

Frage 1: Wie weit fährt ein **LKW** mit **Tempo** 80 km/h in 30 min.?

Frage 2: Wie lang braucht ein **PKW** mit **Tempo** 100 km/h für 30 km?

In beiden Fragen ist eine Geschwindigkeit angegeben, die sich nur in ihrem Wert unterscheidet. Auch die Objekte der Fragen sind ähnlich, handelt es sich doch bei LKW und PKW um Verkehrsmittel, die auf der Straße fahren und im Grunde sehr ähnliche Formen haben. Stellt man sich ein oberflächliches Bild zu den beiden Fragen vor, so sieht man darauf sicherlich jeweils ein Fahrzeug auf einer Straße.

Werden einem Schüler diese beiden Aufgaben gestellt und ist er in der Lage, die erste Frage zu lösen, ist es sehr wahrscheinlich, dass der Schüler die zweite Aufgabe zu der ersten in Beziehung setzt und als Analogie betrachtet. Dies kann jedoch in die Irre führen, da der Lösungsweg in der ersten Frage über eine Multiplikation führt, in der zweiten jedoch über eine Division. Die Strukturmerkmale dieser Beispiele werden durch die Formel deutlich. So gilt für die erste Frage die Formel: Strecke s gleich Geschwindigkeit v mal Zeit t ($s = v \cdot t$). Für die zweite Frage gilt die Formel: Zeit t gleich Strecke s durch Geschwindigkeit v ($t = s / v$). Die Oberflächenmerkmale (v , s und t) sind identisch, die Lösungswege und damit die Struktur unterscheiden sich jedoch.

In Tabelle 1 wird noch einmal eine solche typische Schulaufgabe als Quelle mit den

Oberflächenmerkmalen und der Lösungsstruktur dargestellt. In der Tabelle sind dann Variationen für das Ziel angegeben, sowohl für die Struktur als auch für die Oberfläche.

Tabelle 1: Oberflächen- und Strukturmerkmale von Problemen (Tabelle aus Robertson, 2001).

Quelle		
<i>Ein PKW fährt 333 km von Heidelberg nach München mit einer durchschnittlichen Geschwindigkeit von 100 km/h. Wie lang braucht er dorthin?</i>		
<i>Oberflächen-, „Objekte“</i>	PKW, 333km, Heidelberg, München, 100km/h, „Wie lang...“	
<i>Lösungsstruktur</i>	Zeit = Strecke ÷ Geschwindigkeit (a = b ÷ c)	
Ziel		
	<i>Variation der Struktur</i>	
<i>Variation der Oberfläche</i>	<i>Ähnlich</i>	<i>Unähnlich</i>
<i>Ähnlich</i>	Ein LKW fährt 480 Kilometer von Paris nach Lyon mit einer durchschnittlichen Geschwindigkeit von 80 km/h. Wie lang braucht er dorthin?	Ein PKW braucht 4 h 27 min. von Heidelberg nach München mit einer durchschnittlichen Geschwindigkeit von 80 km/h. Welche Strecke legt er zurück?
<i>Unähnlich</i>	Ein Tourist wechselt €300 in US-Dollar zu einem Wechselkurs von €1,17 pro Dollar um. Wie viele US-Dollar bekommt er?	Ein Tourist wechselt \$200 zu einem Kurs von €1,17 pro Dollar um. Wie viele Euro bekommt er?
<i>Lösungsstruktur</i>	a = b ÷ c	b = a · c

Zur Interaktion von Struktur und Oberfläche.

In den hier genannten Beispielen scheint es so, als ob Oberflächenmerkmale mehr zu Missverständnissen führen können, da sie Lösungsstrukturen suggerieren, die nur oberflächlich passen. Dennoch sind sie sehr wichtig für den Umgang mit Analogien. Meist wird die Möglichkeit, eine Analogie zu bilden, erst gar nicht in Betracht gezogen und stattdessen versucht, den schwierigeren Weg der Neuentwicklung einer Lösung zu gehen. Eine interessante Untersuchung dazu liefern Holyoak und Koh (1987). Sie studierten die Wirkungsweise von *Oberflächenähnlichkeit* und *struktureller Ähnlichkeit* und fanden, dass Probleme häufiger gelöst wurden, wenn beides vorlag (70%), und seltener, wenn nur

strukturelle oder Oberflächenähnlichkeit vorlagen (38% bzw. 33%).

Oberflächenmerkmale helfen also, innerhalb des Prozesses der Analogiebildung die Entsprechungen von Quelle und Ziel zu erkennen. Neben diesem Schritt sind eine Reihe weiterer Schritte nötig, um Erfolg beim Analogisieren zu haben, wie der Prozess der Analogiebildung zeigt.

Prozess der Analogiebildung.

Aus verschiedenen Arbeiten von Holyoak über Analogien extrahiert Robertson folgende fünf Schritte zum Prozess des Analogisierens (Catrambone & Holyoak, 1989; Holyoak, 1984, 1985; Holyoak & Thagard, 1989a, b, alle zitiert nach Robertson, 2001, p. 150):

1. Repräsentation bilden,
2. Zugreifen auf eine passende Analogie,
3. Abbilden (mapping) von passenden Merkmalen,
4. Adaptieren von Strukturen (inklusive Lösungswege) und
5. Lernen (für den Transfer bei ähnlichen Aufgaben).

Zu Anfang ist eine möglichst genaue Vorstellung des Problems nötig. Sowohl der Ist- als auch der Soll-Zustand helfen, später eine passende Analogie zu finden. Der zweite Schritt macht deutlich, dass auch das Analogisieren von Gedächtnisleistungen abhängig ist. Je mehr Erfahrung in einem Problembereich vorliegt, desto mehr Analogien liegen vor. Spontan nutzen nur wenige Problemlöser die Strategie des Analogisierens. Erst wenn ein Hinweis gegeben wird, dass bereits ähnliche Probleme bearbeitet wurden, suchen einige nach den im dritten Schritt passenden Merkmalen (Robertson, 2001).

Dabei sind die Oberflächenmerkmale wichtig für das Erkennen der Analogie, während die Strukturmerkmale für das Abbilden von Ziel und Quelle genutzt werden müssen. Wurde das Problem der Analogiequelle bereits gelöst und liegt der entsprechende Lösungsweg noch im Gedächtnis vor, kann dieser auf das Zielproblem übertragen werden. Als letzter Schritt wird hier das Lernen für ähnliche Aufgaben genannt. Dieser Lernschritt baut also auf einen Anwendungsfall auf und ist vergleichbar den „Case based reasoning“-Ansätzen von Schank (1999). Für den nächsten ähnlichen Fall steht dann bereits ein weiterer gelernter Prozess zur Verfügung.

Analogien beim komplexen Problemlösen.

Die Nutzung von Strategien und speziell von Analogien beim komplexen Problemlösen ist weitgehend unerforscht. Allerdings weist Dörner auf die Nützlichkeit des Analogisierens hin (Dörner, 1992, S. 113f). Dabei bezieht er sich auf den *hypothesengenerierenden* Wert von Analogien. Die Hypothesen machen es damit dem Nutzer möglich, Fragen zu stellen und

diesen gezielt nachzugehen. Dabei müssen die Problemlösenden auf eine abstrakte Betrachtungsweise wechseln, also Strukturmerkmale höherer Ordnung erkennen und nutzen. Dörner weist darauf hin, dass das Bilden von Analogien oftmals naheliegend erscheint, Versuchspersonen es aber dennoch nicht nutzen: *„A posteriori mag eine solche Analogisierung sehr primitiv und ganz selbstverständlich erscheinen; viele Versuchspersonen kommen aber auf Derartiges nicht und verharren hilflos in der konkreten Situation.“* (Dörner, 1992, S. 114).

Die Nutzung von Analogien oder auch anderen Strategien ist scheinbar nicht selbstverständlich und erfolgt meist nur auf direktem Hinweis. Ein Grund könnte die kognitive Belastung während der gesamten Problembearbeitung sein, da z.B. für das Analogisieren eine Struktur für das gesamte Problem gefunden werden muss. Demgegenüber sind Taktiken kurzfristiger und wirken sich zumeist direkt auf den Wissenserwerb aus. Nachfolgend wird noch genauer auf die Unterschiede zwischen Taktiken und Strategien im engeren Sinn eingegangen.

3.3 Taktiken

Die Unterscheidung zu treffen, wann ein Handlungsablauf auf einer Strategie beruht und ab wann er „nur“ noch als Taktik zu betrachten ist, ist äußerst schwer. Die Bereichsübergänge sind eher fließend, als durch klar markierbare Grenzen abgetrennt. Der Unterschied zwischen Taktik und Strategie soll im folgenden anhand des beispielhaften Vergleiches der „Difference Reduction“-Taktik mit der „Hill Climbing“-Strategie aufgezeigt werden.

Des weiteren sollen die Taktiken vorgestellt werden, die sich aus Forschungsarbeiten zum komplexen Problemlösen ergeben haben.

3.3.1 Difference Reduction versus Hill Climbing

Im Vergleich zur „Hill Climbing“-Strategie handelt es sich bei „Difference Reduction“ eher um eine Taktik. Bei dieser Taktik wird stets der Operator gewählt, der die Differenz zwischen dem aktuellen Zustand und dem Zielzustand verringert (siehe Anderson, 2000, S. 312f.). Diese Taktik erleichtert die Auswahl des Operators, also des zu wählenden Lösungsschrittes. Innerhalb komplexer Szenarien wird sie häufig dann angewendet, wenn Ziele konkret angegeben werden. Vollmeyer, Burns und Holyoak (1996, p. 93) stellen fest, dass Versuchspersonen im laufenden Experiment zu einer solchen Taktik wechseln und das sogar, wenn ihnen eine andere Taktik (Einzeleingriffe, s.u.) per Instruktionstext aufgetragen wurde (auch Vollmeyer et al. verwenden den Begriff Strategie, für das, was hier als Taktik

bezeichnet wird).

Ein solch spontaner Wechsel hin zu einem differenz-reduzierenden Verhalten findet sich nicht nur beim Bearbeiten komplexer Probleme, sondern auch im Alltag. Das Aufräumen eines Zimmers geschieht laut Anderson beispielsweise in dieser Form, sowie auch das Bearbeiten eines Puzzles. Aber nicht nur Menschen nutzen dieses Verhalten, man findet es auch bei nahezu allen Tierarten. Selbst Fliegen und Läuse gehen diesem Verhalten instinktiv nach (Anderson, 2000, p. 314).

Was unterscheidet nun eine solche Taktik von einer vergleichbaren Strategie? Zunächst klingt das Vorgehen nach dem „Difference Reduction“-Prinzip dem Vorgehen der „Hill Climbing“-Strategie sehr ähnlich. Auch bei ihr werden stets direkt jene Operatoren gewählt, die den Ist-Zustand näher an den Soll-Zustand bringen. Im Vergleich zur benannten Taktik allerdings wird bei der „Hill Climbing“-Strategie der Ist- und der Soll-Zustand weiterhin mental repräsentiert. Um in der Analogie der Bergbesteigung zu bleiben, heißt das: Der gesamte Weg ist zwar noch nicht bekannt, aber die Problemlösende weiß, dass es einen geben wird, der sich aus vielen Schritten zusammensetzt. Bei der schlichteren „Difference Reduction“-Taktik handelt es sich hingegen nur um einzelne Eingriffe, die wesentlich kurzfristiger gedacht sind. Sie sorgen eher für kognitive Entlastung (im Sinne des *cognitive load* nach Sweller, 1988). Wird versucht, Ist- und Soll-Zustand sowie die Lösungsschritte zu repräsentieren, so sorgt dies für einen zusätzlichen kognitiven Energieaufwand. Mit dem Verweis auf den kognitiven Mehraufwand interpretieren auch Vollmeyer et al. (1996, p. 97) in ihren Ergebnissen das Wechseln der Taktiken von sinnvolleren Einzeleingriffen zu solchen der einfachen „Difference Reduction“.

Der nachfolgende Abschnitt beschreibt die für ein wissenschaftliches Vorgehen sinnvollsten Taktiken, die eigentlich in jeder Untersuchung mit komplexen Szenarien zu erwarten sein sollten, da sie den größten Informationsgewinn versprechen.

3.3.2 Nulleingriffe, Einzeleingriffe, starke Eingriffe

Für das sinnvolle Bearbeiten komplexer Szenarien muss zielgerichtet und planvoll vorgegangen werden. Wie angesprochen sind dazu längerfristige Handlungsabläufe in Form von starken oder schwachen Methoden (domänenspezifische oder -unspezifische Strategien) von Nutzen (siehe 'Domänenspezifität' auf Seite 32). Aber auch kurzfristig gibt es einige Möglichkeiten, gezielt Informationen zu sammeln. Solche kurzfristigen Strategien sollen in dieser Arbeit als Taktiken bezeichnet werden, wovon nachfolgend die wichtigsten aufgeführt und kurz erklärt werden. Einige Benennungen entstammen aus der komplexen

Problemlöseforschung und können sogar einzelnen Autoren zugeordnet werden, andere ergeben sich schlicht durch Nachdenken bzw. aus praktischen Erfahrungswerten.

Nulleingriffe.

Bei einem so genannten Nulleingriff werden alle Eingabewerte auf den Basiswert gesetzt, meist also auf die Null. Damit wird das Ziel erreicht, die Dynamik – sowohl Eigendynamik als auch Nebenwirkungen – des Systems zu erkennen. Das heißt, es wird ermittelt, wie die Variablen aufeinander wirken, ohne dass auf das System von außen Einfluss genommen wird. Nur so ist ein eindeutiges Erkennen von Eigendynamiken möglich.

Einzeleingriffe.

Meist besteht bei komplexen Systemen die Möglichkeit, mehr als eine Variable direkt zu beeinflussen. Wird nur eine Variable ausgewählt, die einen Wert erhält, und lässt man die anderen bei Null, so spricht man allgemein von einem Einzeleingriff. Von einigen Autoren wird dies bereits als Strategie bezeichnet (Preußler, 1998; Vollmeyer et al. 1996 u.a.). Ziel der Beschränkung auf eine Variable ist das Erkennen der direkten Einflüsse von Eingabevariablen auf Ausgabevariablen.

Im Englischen wird der Einzeleingriff auch als VOTAT (Vary One Thing At a Time) bezeichnet, so auch von Vollmeyer et al., die sich hierbei auf Tschirgi (1980) beziehen. Dieser identifizierte diese „Strategie“ bei Versuchspersonen, die den Einfluss multipler Faktoren auf eine oder mehrere abhängige Variablen finden sollten (Vollmeyer et al. 1996, p. 79).

Starke Eingriffe.

Werden neben direkten Wirkungen der Eingabevariablen auf eine Ausgabevariable noch weitere so genannte Seiteneffekte vermutet, so sollte die Taktik eines „starken Eingriffs“ gewählt werden. Dabei wird, wenn möglich, ein Eingabewert ausgesucht, der in seiner 10er-Potenz deutlich über dem höchsten Wert im bestehenden System liegt. Ist beispielsweise der höchste Wert in aktuellem System 30, 60 oder auch 100, dann wäre ein starker Eingriff eine 1000 oder auch 10.000. Der Eingriff muss also so stark gewählt werden, dass weder eine Eigendynamik noch eine Nebenwirkung für die Auswirkungen verantwortlich zu machen sind. Eine objektive Definition für starke Eingriffe ist bisher noch nicht gefunden und bleibt wohl bis noch mehr praktische Erfahrungen dazu vorliegen vom jeweiligen System abhängig. Auch in dieser Arbeit wird der starke Eingriff an späterer Stelle abhängig von Systemvariablen festgelegt (siehe 'Analyse der Nutzung der Taktiken' auf Seite 80). Klar ist

jedoch das Ziel, dass mit einem solchen Eingriff ein sichtbarer Effekt der direkten Wirkung und auch des Seiteneffekts auf die Ausgabevariablen erzielt wird. Problematisch kann jedoch bei dieser Taktik sein, dass das System nach diesem Eingriff in einem Zustand ist, der sich nicht mehr kontrollieren lässt. Aufgrund der Dynamiken in einem hoch vernetzten System hat die Umkehrung der Wirkung durch Eingabe des gleichen starken Eingriffs mit umgekehrtem Vorzeichen meist nicht den gewünschten Effekt.

Nullsetzen von Ausgabevariablen.

So wie Wirkungen von Ein- auf Ausgabevariablen nur durch Einzeleingriffe eindeutig auszumachen sind, so lassen sich auch Eigendynamiken (eine Ausgabevariable wirkt auf sich selbst) von Nebenwirkungen (eine Ausgabevariable wirkt auf eine andere Ausgabevariable) nur unterscheiden, wenn in Frage kommende Ausgabevariablen auf Null gesteuert werden. Ziel dabei ist zu erkennen, ob der Einfluss auf sie selbst oder auf eine andere Variable in der Nähe von Null kleiner wird oder sich im negativen Bereich sogar umkehrt. Verändert sich die Variable weiterhin gleichmäßig, so handelt es sich um eine Nebenwirkung einer anderen Variable auf diese ausgewählte Variable.

Problematisch bei dieser Taktik ist, dass Eingriffswissen bereits vorhanden sein muss, bevor so gezielt auf Variablen Einfluss genommen werden kann. Das Nullsetzen ist daher eigentlich eine doppelt sinnvolle Taktik, da sie neben dem beschriebenen Effekt noch das Hypothesenprüfen miteinschließt. Denn das Auf-Null-Steuern ermöglicht die Prüfung des bisherigen Wissens über die Wirkrichtung und Wirkstärke der Eingabevariablen.

Umkehrung von Vorzeichen.

Als sehr einfache Taktik, die dennoch nicht häufig eingesetzt wird, erweist sich das Umkehren der Vorzeichen sowohl bei den Eingaben als auch bei den zu steuernden Ausgabevariablen. Wird eine Ausgabevariable in den Bereich unter Null gesteuert, kann dies Nebenwirkungen und Eigendynamiken aufdecken. Sinkt eine der Ausgabevariablen unter Null und die andere bleibt positiv, so zeigt sich, ob die nun negative Variable durch eine positive Nebenwirkung einer anderen Ausgabevariable beeinflusst ist oder ob sie bei nun sinkenden Werten sich selbst durch eine Eigendynamik verändert.

Ebenfalls zur Aufklärung von Dynamiken geeignet ist die summierte und im Vorzeichen gewechselte Eingabe aller bisherigen Eingaben. Denn ein statisches System würde sich dann wieder im Ausgangszustand befinden, was viele Versuchspersonen implizit auch von komplexen Systemen erwarten.

3.4 Arbeitsdefinitionen von Strategie und Taktik

Zusammenfassend sollen noch einmal die Konzepte der Strategie und der Taktik einander gegenübergestellt werden und eine Definition, wie sie für die vorliegende Arbeit gelten soll, gefunden werden.

Definition: Strategie.

Unter einer Strategie soll ein System von Regeln verstanden werden, das auf vorliegende Informationen angewendet werden kann und zu bestimmten Entscheidungen führen soll. Dabei steht jeweils ein Ziel im Vordergrund, zu dessen Erreichung die Strategie als Mittel dient.

Im Rahmen dieser Arbeit wird die Analogie als bereichsübergreifende Strategie untersucht, wobei der Strategiebegriff auch als domänenspezifisches Prinzip denkbar ist.

Definition: Taktik.

Taktiken sind Muster von Eingriffen und fest verbunden mit einer eingeschränkten spezifischen Domäne, können also nicht einfach in andere Bereiche übertragen werden. Taktiken sind zumeist leicht zu vermittelnde spezifische Anwendungen, die dem Ziel der (mental) Entlastung dienen können. Sie sind kurzfristiger als Strategien und bedürfen keines übergeordneten Handlungsplans.

Tabelle 2: Gegenüberstellung der Konzepte Strategie und Taktik.

Strategie	Taktik
System von Regeln	Muster von Eingriffen
auch domänenübergreifend	domänenspezifisch
Ziel steht im Vordergrund	Verhalten steht im Vordergrund
kein Algorithmus	
übergeordneter Handlungsplan	kurzfristige Aktion
entscheidungssuchend	

Tabelle 2 fasst die Unterschiede zwischen Strategien und Taktiken, wie sie in dieser Arbeit verstanden werden, zusammen.

4. Methoden

Für diese Untersuchung wurde ein experimentelles Setting gewählt, um den Einfluss von Analogien auf den Wissenserwerb und die Steuerleistung von Problemlösenden zu untersuchen. Als Werkzeug dient ein dynamisches Szenario, beschrieben in Unterabschnitt 4.5.1, welches im DYNAMIS-Programm umgesetzt wurde. Neben den Effekten der Analogien, welche in Unterabschnitt 4.5.3 beschrieben werden, ist ebenfalls die Wirkung von Taktiken auf den Wissenserwerb Hauptbestandteil dieser Arbeit.

Dieses Kapitel liefert zunächst einen Integrationsversuch (4.1) über bisher angesprochene Konstrukte der genannten Theorien. Neben diesen theoretischen Grundlagen war auch eine qualitative Vorstudie (4.2) mit lauten Denkprotokollen für die Bildung der Hypothesen wichtig. Die Hypothesen werden in Abschnitt 4.3 beschrieben. Der Versuchsablauf wurde noch durch eine Pilotstudie (4.4) erprobt, deren Ergebnisse (4.4.1) noch kleinere Auswirkungen auf den Versuchsablauf hatten. Der Versuchsablauf wird gemeinsam mit dem experimentellen Design im Abschnitt 4.5 beschrieben. Es schließen sich die Überlegungen zur Stichprobengröße (4.6) und den Datenanalyseverfahren (4.8) an. Derzeit noch untypisch für psychologische Experimente ist die Durchführung einer Internetuntersuchung, weshalb diese in Abschnitt 4.7 genauer beschrieben wird.

4.1 Einflussfaktoren der Problemlösegüte – ein Rahmenmodell

In den vorangegangenen Kapiteln sind die wesentlichen Konstrukte für diese Untersuchung definiert und auf Basis der ausgewählten Theorien diskutiert worden. Dieser Abschnitt soll mit Hilfe der Abbildung 5 einen Überblick über diese Konstrukte geben und damit auch eine Zusammenfassung des bisher Gesagten liefern.

Wichtiger Bestandteil der Untersuchung ist die Wissenskomponente, die in der komplexen Problemlöseforschung seit den Experimenten von Berry und Broadbent (1984) kontrovers diskutiert wird. Sie ist daher auch in der Abbildung 5 zentral dargestellt. Wie im Abschnitt 'Wissensanwendung' auf Seite 21 bereits beschrieben, haben verschiedene Autoren gezeigt, dass das Strukturwissen einen mehr oder weniger starken Einfluss auf die Güte der Systemsteuerung hat. Ob ein solcher Zusammenhang auch in dieser Untersuchung besteht, prüft die Hypothese III.

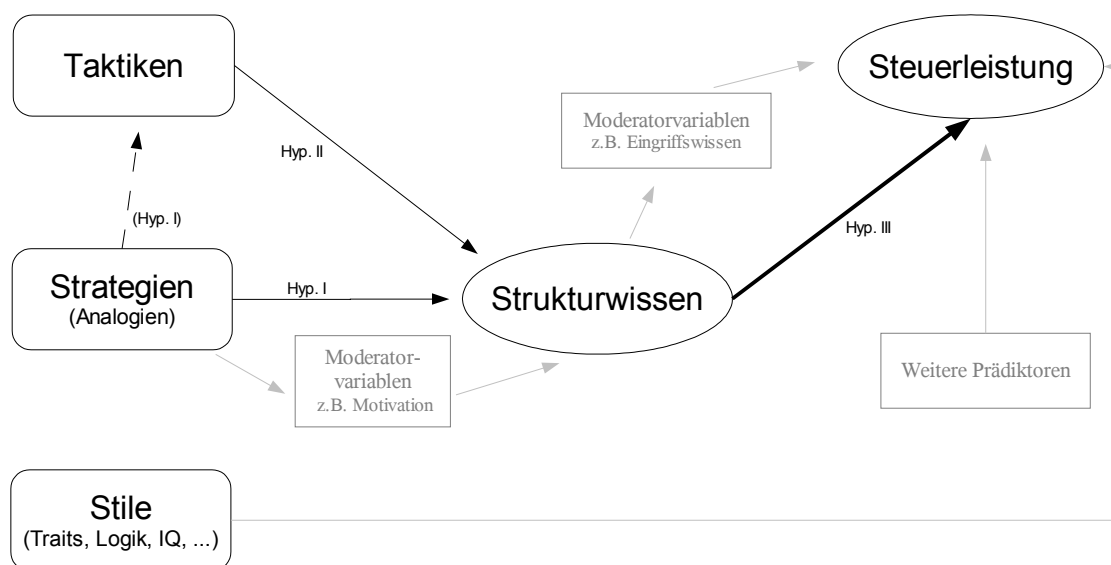


Abbildung 5: Ein einfaches Rahmenmodell der Einflussfaktoren auf die Problemlösequalität (Steuerleistung) mit graphischer Darstellung der Hypothesen (Hyp.).

Dieser Zusammenhang ist auch aus weiteren Gründen für diese Untersuchung wichtig. Es wird davon ausgegangen, dass speziell das Strukturwissen als Moderatorvariable zwischen der Steuerleistung und den handlungsleitenden Strategien wirkt.

Weitere Moderatorvariablen können sowohl zwischen der Beziehung der Strategie zum Strukturwissen als auch vom Strukturwissen zur Steuerleistung vorhanden sein. Vor allem dem Bereich der Motivation und des Interesses sollte große Aufmerksamkeit geschenkt werden, wie dies teilweise durch Aufnahme von Interessensskalen, die bei der Bearbeitung komplexer dynamischer Systeme erhoben werden (z.B. in FSYS 2.0 von Wagener & Conrad, 1997) auch schon geschieht. Die genaueren Analysen müssen aber weitergehenden Untersuchungen überlassen werden, da sie den Rahmen dieser Arbeit sprengen würden. Dies gilt auch für die mögliche Interaktion, die sich durch die kognitiven Stile oder ähnliche Persönlichkeitsmerkmale ergeben könnte. Auch viele weitere Prädiktoren werden untersucht, können hier aber nicht genauer in dieses einfache Modell integriert werden. Die vermuteten Zusammenhänge werden daher im Rahmenmodell hellgrau gezeichnet.

Ziel der Untersuchung ist aber dennoch, in einen Bereich vorzustoßen, der bisher eher gemieden wurde und über den daher wenige Erkenntnisse vorliegen. Dieser Bereich der Strategien zum komplexen Problemlösen ist sehr uneinheitlich in der Literatur beschrieben

(siehe Kapitel 3. Stile, Strategien und Taktiken). Es wird daher die Strategie des Analogisierens ausgewählt, mit der die Hoffnung verbunden ist, dass mit einer Aufklärung der Nutzung von Analogien beim komplexen Problemlösen eine Grundlage für die Analyse weiterer Strategien gelegt werden kann.

Bisherige Untersuchungen widmen sich zumeist den Eingriffsmustern, die in dieser Arbeit als Taktiken bezeichnet werden (siehe Abschnitt Nulleingriffe, Einzeleingriffe, starke Eingriffe auf Seite 44). Mit mehr oder weniger großem Erfolg wird dadurch eine Varianzaufklärung durch verschiedene Taktiken betrieben. In dieser Untersuchung widmet sich die Hypothese II der Frage, welche Taktik in welchem Zusammenhang mit dem Wissenserwerb steht.

Eine Aufklärung bzw. Verbesserung des Problemlöseverhaltens wäre nicht nur für eng umgrenzte dynamische Systeme wünschenswert. Als Vorstoß ins Unbekannte ist der Versuch zu bewerten, ob denn übergeordnete zielführende Strategien den Problemlösenden in dynamischen Szenarien helfen können. Hypothese I widmet sich dieser Frage nach dem Wissenserwerb und gleichzeitig der Frage, ob ein möglicherweise größerer Wissenserwerb über die Vermittlung besserer Taktiken erfolgte (gestrichelte Linie im Rahmenmodell). Dazu wurde nur eine Strategie, die des Analogisierens, ausgewählt. Mit dieser wurde zwar schon in der Transferforschung und Didaktik gearbeitet, bisher aber noch nicht im Bereich des komplexen Problemlösens.

Die Protokolle des lauten Denkens einiger Versuchspersonen in der nachfolgend beschriebenen Vorstudie brachten viele Hinweise darauf, dass Analogien ein Mittel zur Vereinfachung (Komplexbildung) für viele Problemlösende sein könnten.

4.2 Voruntersuchung: Einzelfallstudien – Lautes Denken

Bereits im Juni 2001 wurden vier Versuchspersonen bei der Bearbeitung von DYNAMIS-Szenarien an der neuen Benutzeroberfläche untersucht. Dabei sollte ein möglichst abstraktes System verwendet werden, das keinerlei Vorwissen zuließ. Gewählt wurde zunächst ein einfacher Automat, der aus drei Eingabe- und drei Ausgabevariablen bestand. Es stellte sich aber recht schnell heraus, dass dieser Automat sehr wohl mit inhaltlichen Annahmen besetzt war, so dass die Versuchspersonen versuchten, ihr Vorwissen zu nutzen. Dies war vorher nicht bedacht worden und passte daher auch gar nicht zum System. Hinzu kam, dass der Automat keine Annahmen über Nebenwirkungen und Eigendynamiken zu ließ. Daher wurden die Fallstudien schnell umgestellt und es wurde auf ein bewährtes System zurückgegriffen. Das „Sinus“-System von Funke (1992; siehe auch Abbildung 4 auf Seite 39) wurde leicht abgewandelt und verwendet.

Bei der Beobachtung der vier Versuchspersonen fiel besonders Verhalten auf, das aus Sicht des Versuchsleiters als *Fehler* zu bezeichnen wäre, da es in dem linearen Gleichungssystem gar keinen Sinn machte.

- (1) So versuchten einige bereits den ersten Takt auszuwerten und das System zu interpretieren,
- (2) andere versuchten, Zahlenreihen zu erkennen und diese zu ergänzen, wie es in Intelligenztests oft als Aufgabe gestellt ist.
- (3) Des weiteren wurde davon ausgegangen, dass das System nicht proportional sei und im negativen Bereich anders reagiere als bei positiven Werten.
- (4) Auch wurde angenommen, jeder Durchgang sei anders
- (5) und ebenso würde das System in höheren Takten anderen Gesetzmäßigkeiten folgen als in anfänglichen Takten eines Durchgangs.
- (6) Zumeist wurde auch ein Zusammenhang zwischen den endogenen Variablen nicht in Betracht gezogen.
- (7) Dafür wurde z.T. davon ausgegangen, dass alleine die graphische Anordnung bereits auf Kausalitäten bzw. Zusammenhänge hinweisen würde.

Um Fehler solcher Art in dem Experiment zu umgehen, wurden diese sieben Stolpersteine in den Instruktionstexten berücksichtigt und explizit gemacht. Dies sollte die Transparenz erhöhen und zu geringerer Varianz führen. Als ein weiterer "Fehler" fiel auf, dass der Einsatz von systematischen Variationen der Eingabedaten minimal war. Fast keiner ließ einen Takt mit nur Nullen zu, höchst selten und von einigen gar nicht wurde nur eine Variable verändert, und infolgedessen kam es auch nicht zu starken Eingriffen auf einzelne Variablen, um direkte Zusammenhänge zu erkennen. Um zu überprüfen, ob diese einfachen Taktiken zu mehr Erfolg beim Wissenserwerb und der Steuerleistung führen als das eigentliche Treatment, wurde eine zusätzliche Gruppe in das Untersuchungsdesign aufgenommen und als zweite Kontrollgruppe (KG2) bezeichnet. Sie bekam das abstrakte System ohne Variablenbezeichnungen, wie auch die „normale“ Kontrollgruppe (KG1). Neben den allgemeinen Instruktionen bekam sie aber einen weiteren Text, der diese taktischen Tipps enthielt (siehe Abschnitt 'Experimentelles Design und Versuchsablauf' auf Seite 56 bzw. für den Instruktionstext Anhang C).

Was sich im Umgang mit dem "Sinus"-System auch zeigte, war der zweiseitige Umgang mit den Variablenbezeichnungen. Die frei erfundenen Wörter wie "Olschen", "Gaseln" und "Schmorken" wurden einerseits als "komisch" und "sonderbar" bezeichnet, andererseits führten sie auch direkt dazu, dass einige Versuchspersonen sich die Variablenbezeichnungen als

Käferpopulationen vorstellten, die fressen und gefressen werden – was auf der einen Seite zum leichteren Umgang mit dem System führte und Hemmschwellen abbaute, was zum anderen aber zu großer Mehrbelastung führte, da es schnell passierte, dass Fresser und Gefressene verwechselt wurden und die Überlegungen zu der Käferanalogie neu aufgestellt werden mussten.

Die Analogien für das Hauptexperiment mussten also zum einen frei von Systemhinweisen aufgrund des Inhaltes sein und zum anderen möglichst allgemein zu Überlegungen der Versuchspersonen passen. Auch mussten Variablenbezeichnungen gefunden werden, die der Alltagssprache der Versuchspersonen entsprachen und zu einfacher kognitiver Verarbeitung führten und nicht zu größerer Mehrbelastung durch das Erlernen sinnfreier unbekannter Wörter. Mit den ausgewählten Analogien, die im Unterabschnitt 'Das Treatment der Analogien und der Taktiken' auf Seite 62 beschrieben werden, wurde versucht, diesen verschiedenen Ansprüchen gerecht zu werden.

Durch das Treatment der unterschiedlichen Anzahl von Analogien wird experimentell der Einfluss von Analogien überprüft. Dieser Überprüfung widmet sich Hypothese I. Sie ist durch die kausale Testung als die stärkste und wichtigste Hypothese dieser Untersuchung anzusehen. Nachdem sie im folgenden Abschnitt genauer beschrieben wird, erläutert dieser auch noch zwei schwächere Korrelationshypthesen.

4.3 Hypothesen

Aus den bisherigen Kapiteln und der Voruntersuchung ergeben sich drei Hypothesen, die mit der nachfolgenden Untersuchung überprüft werden sollen. Nachfolgend sind die Hypothesen inhaltlich (I) beschrieben und ausführlich hergeleitet aus den entsprechenden Theorien und Befunden sowie in statistische (S) Hypothesen übersetzt.

Hypothese I

(I) Die Versuchspersonen, die sich intensiv mit dem Abbilden von Strukturen aus den Analogien auf das abstrakte System beschäftigen, sollten einen vergrößerten Hypothesenraum haben und dadurch zu mehr Strukturwissen kommen.

Die Analogien werden in der Untersuchung so vorgegeben, dass ihre Oberflächenmerkmale (die Variablenbezeichnungen) anfangs noch nicht sinnvoll zugeordnet sind. Die psychologische Arbeit des „Mappings“ (siehe 'Prozess der Analogiebildung' auf Seite 42)

muss von den Versuchspersonen selbst vorgenommen werden. Um die Oberflächenmerkmale der Analogien den Variablen zuordnen zu können, muss man sich Gedanken über die Strukturmerkmale der Quelldomäne (die Analogie) und der Zieldomäne (das Variablengefüge des DYNAMIS-Szenarios) machen. Diese Überlegungen über Strukturen beider Domänen sollen helfen, schneller mit dem System zurechtzukommen und Fehler zu vermeiden. Da die Analogien recht alltägliche Inhalte haben, sollten viele Strukturmerkmale recht schnell und einfach generiert werden können. Bei der Zuordnung der Variablenbezeichnungen zu den Variablen gibt es oftmals nicht nur eine korrekte Lösung. Die Analogien wurden absichtlich so offen gehalten, um nicht den Experimentalgruppen durch die semantischen Inhalte Vorteile gegenüber den Kontrollgruppen zu verschaffen. Erst die mentale Auseinandersetzung mit den Strukturmerkmalen sorgt für ein größeres Verständnis und mehr Hypothesen bei den Versuchspersonen der Experimentalgruppen. Es wird erwartet, dass die Personen, die sich gerne eigene Experimente ausdenken und diese prüfen (vgl. „Experimentierer“ bei Klahr & Dunbar, 1988, p. 24) von den Analogien profitieren können, da sie aus den Analogiebereichen solche prüfbaren Experimente übertragen können. Ebenso wird davon ausgegangen, dass auch die anderen (vgl. „Theoretiker“ bei Klahr & Dunbar, 1988) Versuchspersonen von dem Treatment profitieren, da es ihnen erleichtert wird eigene Hypothesen zu generieren.

(S) Erwartet wird, dass die Experimentalgruppe mit drei Analogien (EG2) mehr Strukturwissen erlangt als die mit einer Analogie (EG1) und diese immer noch mehr als die unspezifische Kontrollgruppe (KG1).

Hypothese II

(I) Versuchspersonen, die sinnvolle Taktiken zum Explorieren anwenden (wie z.B. die aus Abschnitt 'Nulleingriffe, Einzeleingriffe, starke Eingriffe' von Seite 44), werden mehr Strukturwissen erzielen.

Bei den Experimenten von Vollmeyer, Burns und Holyoak (1996) zeigte sich ein Zusammenhang zwischen dem Nutzen von Einzeleingriffen (VOTAT vgl. Abschnitt 'Taktiken' auf Seite 43) und Strukturwissen, welches sich wiederum auf die Steuerleistung im Transfer auswirkte. In der vorliegenden Untersuchung wurde daher eine spezielle „taktische“ Kontrollgruppe (KG2) aufgenommen, die direkte Anweisungen für Taktiken bekam. Von dieser zweiten Kontrollgruppe wird daher erwartet, dass sie die Taktiken häufiger anwendet

als die Kontrollgruppe (KG1). Zusätzlich dient diese Gruppe als Vergleich für die Experimentalgruppen. So ist letztlich zu erwarten, dass auch diese beiden Gruppen (EG1 und EG2) angeregt durch die Analogien ein systematischeres Vorgehen wählen (siehe Hypothese I).

(S) Erwartet wird eine positive Korrelation zwischen der aggregierten Variable der Nutzung von Taktiken und den richtigen Antworten im Strukturwissenstest. Eine negative Korrelation wird für den Zusammenhang zwischen der Taktikvariable und den Falschantworten im Strukturwissenstest erwartet.

Hypothese III

(I) Mehr Strukturwissen sollte sich positiv auf die Steuerleistung im letzten Durchgang, dem Zieldurchgang, auswirken.

Angelehnt an Experimente mit einer vergleichsweise langen Explorationsphase und einem einzelnen Zieldurchgang (vgl. Funke, 1986, 2003; Preußler, 1996, 1998; Vollmeyer et al., 1998) ist einer der Hauptbestandteile auch bei dieser Untersuchung der Wissenserwerb, speziell der Erwerb des Strukturwissens. Wie sich in den Experimenten von Preußler, aber auch von Vollmeyer et al. gezeigt hat, wirkt sich erworbenes Strukturwissen vor allem dann aus, wenn beim so genannten Transfer den Probanden neue Ziele vorgegeben werden. In der vorliegenden Untersuchung werden nur zu einem Zeitpunkt, nämlich vor dem vierten Durchgang, Ziele vorgegeben, so dass diese für alle bis dahin unbekannt sind. Damit handelt es sich beim Zieldurchgang um einen der Transferphase vergleichbaren Durchgang. Bei der Steuerleistung handelt es sich um ein logarithmiertes Abstandsmaß (genauer in Ergebnisse zur Steuerleistung' auf Seite 84), bei dem kleinere Werte für eine bessere Steuerleistung stehen. Daher werden negative Korrelationen mit diesem Abstandsmaß erwartet.

(S) Es sollte sich eine negative Korrelation zwischen Daten der Wissensdiagnostik vor dem letzten Durchgang und der Steuerleistung ergeben.

4.4 Pilotstudie

Im Mai 2002 wurden die gesammelten Erfahrungen aus der Vorstudie (wenn man so will, die alpha-Phase) zusammengetragen und in einen Pilottest integriert (die bei Programmierern so genannte beta-Phase). Inzwischen war das zugrundeliegende System von sechs Variablen auf vier reduziert worden. Erfahrungen mit einigen Einzeltestern hatten ergeben, dass diese sich zu stark gefordert sahen. Da im System eine Eigendynamik und eine Nebenwirkung enthalten bleiben sollte und auch eine der beiden Eingabevariablen einen Seiteneffekt neben ihrem Haupteffekt auf die Ausgabevariablen haben sollte, war nun mit einem zwei mal zwei Variablensystem schon fast die minimale Grenze für ein komplexes Problemszenario erreicht. Im Vergleich zum "Sinus"-System von Funke (1992) fiel das dortige Subsystem weg, die übrigen vier Variablen unterschieden sich hingegen in ihren Wirkstärken, nicht jedoch in der Anzahl der Wirkverbindungen.

Neben dem Test des WITS-Szenarios sollte natürlich auch die Internettauglichkeit des neuen DYNAMIS-Programms getestet werden. Zu diesem Zweck wurden daher auch die Pilottester über das Internet via Email informiert und gebeten, sich dem Test zu stellen. Da es sich noch nicht um das endgültige Experiment handelte, wurde die Werbung nicht veröffentlicht und das Feld der Pilottester auf zehn Personen eingeschränkt.

4.4.1 Ergebnisse der Pilotstudie

Die Ergebnisse der Pilotstudie waren ebenso erschreckend wie lehrreich. Obwohl die Testpersonen persönlich bekannt waren und ihren guten Willen äußerten, ist der Aufforderungscharakter von Emails wohl als sehr gering einzuschätzen. Es verstrich einige Zeit, bevor sich die ersten an das Experiment wagten. So kam es, dass neben der ersten Email noch weitere folgten, um die Pilotstudie zu beschleunigen und zu einem Abschluss zu bringen. Zwischen dem 2. Juni und dem 13. Juni 2002 kamen dann immerhin fünf vollständige Datensätze an. Die anderen blieben zunächst aus. Auf persönliches Nachfragen wurden dann verschiedene Gründe genannt. Man habe bisher nur mal so reingeschaut, man wäre so schlecht gewesen, dass mit den Daten eh nichts anzufangen gewesen wäre und ähnliche Aussagen konnten gesammelt werden. Daneben gab es aber auch Verbesserungsvorschläge für die ein oder andere Formulierung der Instruktionstexte, die in die Hauptuntersuchung einfließen konnten. Klar wurde ebenfalls für die Hauptuntersuchung, dass ein größerer Anreiz geschaffen werden mußte, um Versuchspersonen zu vollständigen Dateneingaben zu bringen. Daher wurde ein Gewinnspiel für alle Teilnehmer, die bis zum Ende der Untersuchung durchhielten, angesetzt.

Was die Analyse der Daten aus den fünf Datensätzen anging, fielen die Erkenntnisse eher gering aus. Trotz eher überdurchschnittlich intelligenter Zusammensetzung der Pilottester und -testerinnen (alles erfolgreiche Studierende), war deren Kenntnisstand über die Systemvariablen sehr gering bis gar nicht gegeben. Das meldeten auch die Testerinnen und Tester zurück, die zwar guten Willens starteten, dann aber dennoch vor Ende abbrachen. Nur eine der fünf Personen erwarb die Kenntnis über alle Wirkverbindungen und Wirkrichtungen, tat sich aber über deren Stärke auch schwer. Dennoch gelang es ihr damit, das System etwas zu steuern. Die Bearbeitungszeit mit mehr als drei Stunden war jedoch jenseits des Möglichen für eine Internetstudie.

Aus diesen Kenntnissen heraus musste gefolgert werden, dass das System trotz geringer Variablenzahl zu anspruchsvoll war. Da eine Kürzung weiterer Variablen nicht sinnvoll erschien, die angesprochenen Verbindungen innerhalb der Ausgabevariablen bleiben sollten, schien die naheliegendste Lösung zu sein, die Stärke der Eigendynamik und die der Nebenwirkung deutlich zu verringern, um so das System leichter steuerbar und etwas träger zu gestalten. Die Verringerung der Stärke von Eigendynamik und Nebenwirkung bringt den Effekt, dass Eingaben im Vergleich zu den Dynamiken stärker und damit besser hervortreten. Wie die Hypothesen in ein Design umgesetzt wurden und wie das Experiment ablief, beschreibt der nächste Abschnitt.

4.5 Experimentelles Design und Versuchsablauf

Dem Experiment unterliegt ein Versuchsdesign mit zwei Kontrollgruppen (ohne und mit taktischen Tipps) und zwei Experimentalgruppen, wobei die erste Experimentalgruppe (EG1) nur eine Analogie erhielt und bei der zweiten Experimentalgruppe (EG2) zu jedem Durchgang der Explorationsphase die Analogie wechselte, d.h. dass den Versuchspersonen insgesamt drei Analogien präsentiert wurden (die Beschreibungen zur Explorationsphase und den Analogien folgen jeweils im Unterabschnitt 4.5.2 bzw. 4.5.3). Im letzten Durchgang, dem Zieldurchgang, wurde die dritte Analogie beibehalten. Es handelt sich daher um ein einfaktorielles Design mit drei Stufen (null, eine und drei Analogien). Nach Hager und Westermann (1983) könnte bei den taktischen Tipps auch von einem zweiten Faktor gesprochen werden, der nicht vollständig gekreuzt wurde. Es läge dann ein unvollständig gekreuztes zweifaktorielles Design vor. Der Einfachheit halber soll von einem Faktor gesprochen werden und einer unspezifischen Kontrollgruppe (KG1) sowie einer taktischen Kontrollgruppe (KG2).

Die einzelnen Phasen der Untersuchung können detailliert der Tabelle 3 entnommen werden. Sie gliedern sich grob in einen (1) Voruntersuchungsteil mit Informationen über Dauer der Untersuchung und dem soziodemographischen Fragebogen (2). Im Hauptteil (ab 3) folgt ein langer Instruktionstext für alle, bevor im zweiten Instruktionstext (4) die verschiedenen Treatments an die Gruppen gegeben werden. Danach findet dreimal jeweils ein Explorationsdurchgang gefolgt von einem Test zur Wissensdiagnostik statt. Nur die EG2 bekommt zwischen den Durchgängen neue Analogien mitgeteilt, und sie erhalten neue Variablenbezeichner für die Eingabe- und Ausgabevariablen. Zum Ende des Hauptteils bekommen alle Versuchspersonen die gleichen Ziele genannt (14), und nach einem weiteren Wissenstest folgt eine Eingabemaske für einige Fragen und freie Kommentare.

Tabelle 3: Versuchsdesign des Experiments mit den einzelnen Phasen.

Phase der Untersuchung		KG1	KG2	EG1	EG2
1	HTML-Seiten (Gewinnspiel, Versuchsbeschreibung, Softwaretest)	alle			
2	soziodemographischer Fragebogen	alle			
3	Instruktionstext	alle			
4	Analogie und Analogietext " Mensa "			X	X
	Taktische Tipps		X		
5	Durchgang 1	alle			
6	Wissenstest 1	alle			
7	Analogie und Analogietext " Kleines Ökosystem "				X
8	Durchgang 2	alle			
9	Wissenstest 2	alle			
10	Analogie und Analogietext " Müllsystem "				X
11	Durchgang 3	alle			
12	Wissenstest 3				
13	Zieltext mit den Zielwerten				
14	Durchgang 4				
15	Wissenstest 4				
16	abschließender Kommentar				
17	Danksagungstext				

Der gleiche Ablauf des Experiments wurde auch für die Pilotstudie (siehe Abschnitt 'Pilotstudie' auf Seite 55) verwendet. Was die Versuchspersonen genau zu tun hatten, wurde bisher noch nicht beschrieben und soll nun nachgeholt werden. Außerdem folgt nun eine Zuordnung des verwendeten DYNAMIS-Szenarios in die Taxonomie Funkes aus dem Abschnitt 'Taxonomien' von Seite 25.

4.5.1 Beschreibung und Einordnung des DYNAMIS-Systems „WITS“

Ein DYNAMIS-Szenario wird über die Eingabevariablen (auch Input- oder exogene Variablen genannt) gesteuert. Dies geschieht durch die Eingabe von Zahlenwerten zu einem jeweiligen Zeittakt. Die Eingabe von nahezu beliebigen positiven oder negativen Zahlenwerten wiederum verändert die Ausgabevariablen (auch Output- oder endogene Variablen genannt). Nach Eingabe der Zahlenwerte wird durch Drücken des OK-Buttons der Eingriff des aktuellen Taktes abgeschlossen, und es wird der neue Zeittakt mit den Wirkungen auf die Ausgabevariablen angezeigt. Der Screenshot in Abbildung 7 auf Seite 63 zeigt einen Durchgang, in dem bereits sieben Eingriffe getätigt wurden. In welcher speziellen Weise die Eingaben in den exogenen Variablen modifiziert werden und auf die endogenen Variablen wirken, liegt an der Matrix, die das lineare Gleichungssystem bestimmt. Für das in dieser Untersuchung verwendete System soll sie nun kurz beschrieben werden.

In Anlehnung an einen aufwendig produzierten Film, der 1999 die Kinocharts erstürmte und in dem sich die Akteure immer wieder die Frage stellen „What is the matrix?“, wurde das spezielle DYNAMIS-Szenario für die vorliegende Untersuchung mit dem Kürzel „WITS“ benannt. Die Abkürzung „WITS“ steht für die Frage, die sich die Akteure der Untersuchung immer wieder stellen sollten: „What is the system?“.

Tabelle 4: Die drei Matrizen für das WITS-Szenario aus der Steuerdatei für DYNAMIS.

Axx	Ayx	Ayy
1 0	0 -1.5	1 0.25
0 1	1 -2.0	0 1.1

Tabelle 4 zeigt die drei Matrizen, die für das lineare Gleichungssystem nötig sind. Sie wirken trocken und eher unübersichtlich, beinhalten aber für den geübten DYNAMIS-Anwender alle Information über die Struktur des speziellen Systems. Übersichtlicher wirkt dagegen ein vollständiges Kausaldiagramm wie das in Abbildung 6.

Sowohl in den drei Matrizen als auch im Kausaldiagramm fehlen wichtige Angaben für das DYNAMIS-System wie die Startwerte und die Zielwerte. Alle Angaben werden in einer Steuerdatei an das System übergeben (siehe Anhang A). Die Steuerdatei ist eine einfache Textdatei, aus der das System die Angaben ausliest und dann das entsprechende Szenario am Bildschirm darstellt. Durch Veränderung der Werte und Namen in der Steuerdatei sind nahezu beliebige Szenarien denkbar. Hauptbestandteil bleiben die Matrizen Ayy und Ayx, die die Wirkungsweise der Variablen aufeinander bestimmen. Wirkt wie hier die zweite exogene Variable (II) auf die erste endogene Variable (A), was in der Matrix Ayx mit -1.5 zu sehen ist,

Kausaldiagramm zum WITS-System

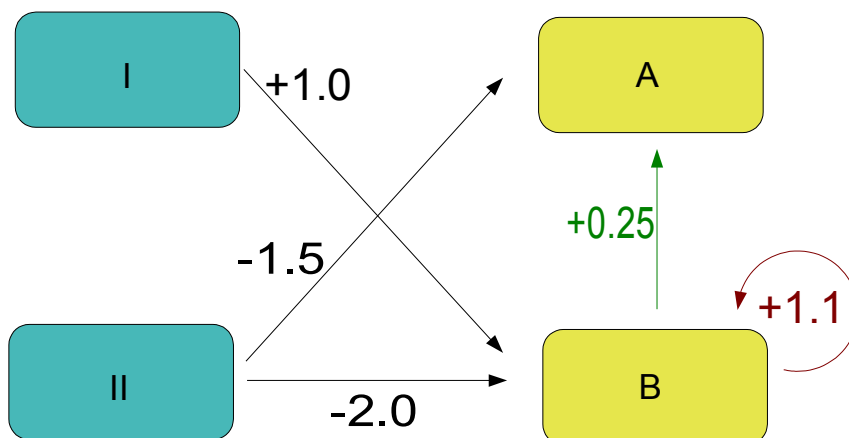


Abbildung 6: Kausaldiagramm des DYNAMIS-Systems "WITS".

so kann diese Wirkung durch jede reelle Zahl ersetzt werden und z.B. durch die Zahl 3.0 verdoppelt und in ihrer Richtung durch den Vorzeichenwechsel umgekehrt werden. Die anderen Systemmerkmale bleiben ansonsten aber unberührt von der Veränderung.

Wie bereits in Abschnitt 'Taxonomien' auf Seite 25 festgestellt wurde, ist eine Zuordnung von Systemmerkmalen für eine bessere Vergleichbarkeit von Forschungsergebnissen äußerst sinnvoll. Daher soll nun die Zuordnung des vorliegenden Systems zur oben genannten Taxonomie Funkes erfolgen.

Die in dieser Arbeit erhobenen Merkmale werden kursiv dargestellt und den Begriffen der Taxonomie untergeordnet. Auf eine ausführliche Erläuterung der Merkmale wird verzichtet, wenn diese an anderer Stelle erfolgt oder aufgrund der Eindeutigkeit und Bekanntheit der Merkmale vorauszusetzen ist.

I. **Persönlichkeitsmerkmale.**

1. Kognitive Merkmale.

Im soziodemographischen Fragebogen zu Anfang der Untersuchung wurden die letzte Mathematik- und Deutschnote der Versuchspersonen erfasst. Zusammen mit der Gesamtnote des letzten Schulzeugnisses kann so indirekt auf die Höhe der *Intelligenz* geschlossen werden. Zumindest ist ein direkter Schluß auf die Schulleistung möglich (vgl. Amelang, 1997, S. 246).

Informationen zu *Computerkenntnissen* und der *Sympathie zum Arbeiten mit dem Computer* (*Computerliking*) wurden ebenfalls erhoben. Letzteres könnte schon eher den emotionalen

Merkmale zugeordnet werden.

2. Emotionale und motivationale Merkmale.

Neben der Computerliking wurde die *Bearbeitungszeit* erfasst sowie die *Anzahl erneuter Aufrufe der Informationstexte*, mit denen indirekt auf das Interesse und die Motivation geschlossen werden kann.

3. Persönlichkeitsmerkmale.

Es wurden keine Persönlichkeitsmerkmale erhoben., da diese für die Fragestellung nicht relevant waren und zunächst ein Persönlichkeitstest für die Online-Untersuchung erstellt werden müsste.

II. Situationsmerkmale.

1. Transparenz.

Das System sollte möglichst transparent dargestellt werden, ohne jedoch eine Analogie vorzugeben, da Analogien ja die Treatmentbedingung dieser Untersuchung darstellen. Daher war das Vorlegen eines Kausaldiagramms ebensowenig möglich, wie das Beschreiben durch ein Beispiel anhand einer möglichen semantischen Einbettung. Ersteres würde zu Verzerrungen aufgrund visueller Analogien führen, und letzteres würde der Kontrollgruppe eine Anregung zum Analogisieren geben, die möglichst zu vermeiden war. Die Instruktionstexte sind dementsprechend ausführlich und detailliert (siehe Anhang C). Sie beinhalten den Hinweis auf eine Vernetzung sowohl von exogenen auf endogene Variablen als auch von endogenen Variablen untereinander und den Hinweis, dass exogene Variablen nicht vernetzt waren. Auch wird darüber informiert, dass die zeitliche Verzögerung der Eingabeauswirkungen auf den Folgetakt eingegrenzt ist.

2. Aufgabenstellung.

Zunächst sollen die Versuchspersonen die Systemmerkmale identifizieren. Dazu bekommen sie erst eine Explorationsphase (drei Durchgänge), bevor ihnen die Ziele für die Experimentalphase (Zieldurchgang) gegeben werden. Für die genaue Beschreibung siehe Abschnitt 'Experimentelles Design und Versuchsablauf' auf Seite 56.

III. Aufgabenmerkmale.

1. Formale Aspekte.

Es handelt sich um ein lineares Gleichungssystem mit zwei exogenen (I und II) und zwei endogenen (A und B) Variablen. Die Vernetztheit des Systems war nahezu maximal. Es gab eine Nebenwirkung und eine Eigendynamik, so dass es sich um ein stark exponentiell wachsendes System handelt. Der autoregressive Prozess war minimal, also eins. Es gab demnach keine Zeitverzögerung über den Folgetakt hinweg.

Variable I hat eine exogene Wirkung, Variable II hat zwei exogene Wirkungen, die beide negativ sind. Berücksichtigt man die Nebenwirkung der endogenen Variable B, waren beide Variablen A und B mehrfachabhängig.

2. Inhaltliche Aspekte.

Das System wurde bewusst völlig abstrakt gehalten, damit durch unterschiedliche Vorwissenskonzepte nicht noch mehr Varianz entsteht. Auch sollte durch die Vorgabe der Analogien kein semantischer Mehrwert gegeben werden.

Mit dieser Zuordnung des „WITS“-Szenarios in die Taxonomie Funkes soll eine Vergleichbarkeit mit anderen komplexen Szenarien erleichtert werden. Wie Vergleichbarkeit von Vorwissen geschaffen werden kann, zeigt der folgende Unterabschnitt auf.

4.5.2 Überlegungen zur Explorationsphase

Komplexe dynamische Systeme wie der Tailorshop (Putz-Osterloh & Lüer, 1981) oder Lohhausen (Dörner et al., 1983), bei denen Zielvorgaben wie Gewinnmaximierung oder Zufriedenheit der Bürger vom ersten Eingriff an gegeben waren, standen in der Kritik, dass sie zu stark vom Vorwissen der Versuchspersonen über Betriebswirtschaft, Marktverhältnisse o.ä. abhängig seien (Funke, 1986).

Ein Ausweg aus der Abhängigkeit des Vorwissens ist, dem System eine Explorationsphase zur Erkundung voran zu stellen. Dadurch kann sich Vorwissen bilden bzw. anpassen, bevor die eigentliche Experimentalphase mit den Zielen beginnt. Mit Explorationsphasen ist es möglich, quasi „unbeobachtet“ Dinge zunächst einfach auszuprobieren. Wissen kann erworben werden, und gravierende Fehler wirken sich nicht bis zum Ende der Simulation aus und zerstören damit die gesamte Steuerleistung, sondern werden am Ende eines Durchgangs rückgängig gemacht, und ein System im Anfangszustand steht für neue Eingriffe bereit.

Diese Vorgehensweise mit einer Explorationsphase ist vor allem bei abstrakten Systemen sinnvoll, da sich diese Systeme ja gerade durch ihre Vorwissensunabhängigkeit auszeichnen. Auch das „WITS“-Szenario enthält nur abstrakte Variablennamen für die Kontrollgruppen. Die Tabelle 3 zeigt den Wechsel von Probedurchgängen (5, 8, 11) mit den Tests der Wissensdiagnostik (6, 9, 12). In Anlehnung an die Untersuchung von Vollmeyer et al. (1996) wurden drei Durchgänge mit jeweils sieben Zeittakten für die Explorationsphase gewählt. Vollmeyer et al. stellten bei ihrem System fest, dass das Wissen im vierten Durchgang nicht mehr zunahm. Da ihr System aus acht Variablen bestand und die Durchgänge einen Takt kürzer waren, dürfte die hier gewählte Explorationsphase für ein Vier-Variablen-System ausreichend lang sein.

4.5.3 Das Treatment der Analogien und der Taktiken

Zunächst sollen die drei verschiedenen Analogien beschrieben werden, die die Versuchspersonen der Experimentalgruppe EG2 bekommen. Die EG1 bekommt nur die erste Analogie, während die Kontrollgruppe (KG2) einen Text mit fünf taktischen Tipps erhält, die weiter unten aufgeführt werden.

Beiden Experimentalgruppen wird im ersten Durchgang das Mensa-System vorgegeben. Dies bleibt bei der EG1 bis zum Zieldurchgang bestehen und wechselt nicht mehr. Die EG2 bekommt im zweiten Durchgang jedoch ein kleines Ökosystem vorgelegt, und für den dritten Durchgang und den Zieldurchgang wird ihnen ein Abfallsystem präsentiert.

Damit eine Analogie wirken kann, muss die Versuchsperson bereits Erfahrung, am besten ein Schema zu der Quelldomäne der Analogie besitzen (Robertson, 2001, p. 140). Daher wurden diese einfachen Analogien aus dem Alltag gewählt. Bei Blanchette und Dunbar (2000) zeigte sich, dass Versuchspersonen, die selber Analogien *erzeugen* sollten, solche erzeugten, deren Strukturen gleich waren, während sie bei *vorgesetzten* Analogien solche wählten, deren Oberflächen ähnlich wirkten. Bei dem hier gewählten Aufbau waren zwar auch Oberflächenmerkmale gegeben, dennoch musste sich die Versuchsperson ihre Analogie noch selbst erzeugen und somit Strukturmerkmale vergleichen.

Die Mensa.

Die Mensa besteht aus den vier Variablenbezeichnungen *Besucher*, *Preis*, *Werbung* und *Zufriedenheit*. Diese sollen in eine analoge Struktur gebracht werden und die abstrakten Namen der Variablen (A, B, I und II) ersetzen. Dabei ist das „WITS“-System die eigentliche Zieldomäne (vgl. Abschnitt 'Analogien als Strategie zum komplexen Problemlösen' auf Seite 38), und aus der Quelldomäne des Mensa-Wissens sollen Ideen für mögliche Strukturen kommen. Um diesen Prozess möglich zu machen, ist ein sinnvolles Mapping nötig. Es wird davon ausgegangen, dass bei einem so alltäglichen Begriff wie dem der Mensa einiges an Wissen über die Strukturmerkmale dieser Quelle vorliegt. Diese bekannten Strukturmerkmale können nun als Hypothesen für die Strukturmerkmale des neuen, unbekanntes Systems dienen. Diese Hypothesen können dann überprüft werden. Erst wenn auch über das Ziel einige Strukturmerkmale bekannt sind, können durch das Mapping die Variablenbezeichnungen zugeordnet werden. Technisch läuft dies so ab, dass die Versuchsperson aus einem Drop-down-Menu eine der vier Variablenbezeichnungen auswählt. Diese bleibt den Durchgang über dort stehen, bis eine andere gewählt wird. In Abbildung 7 sieht man die Darstellung eines Durchgangs des DYNAMIS-Systems, in dem die Ausgangsstellung der Analogiebezeichner

noch nicht verändert wurden. Alle Drop-down-Menüs stehen noch auf dem Startwert *Besucher*.

What Is The System (WITS.prm) (Durchgang 1/4)								
		Takt 1/7	Takt 2/7	Takt 3/7	Takt 4/7	Takt 5/7	Takt 6/7	Takt 7/7
A	Besucher	100	150	230	-7182	-2085	-2228	-2086
B	Besucher	200	320	352	-9612	-574	-631	5
I	Besucher	100	0	0	0	0	300	
II	Besucher	0	0	5000	-5000	0	-200	
								weiter
		Zeige Text 1		Zeige Text 2				

Abbildung 7: Screenshot des "WITS"-Szenarios in Durchgang 1 für die Experimentalgruppen.

Das Ökosystem.

Das Ökosystem hat die Variablenbezeichnungen *kleine Käfer*, *Licht*, *große Käfer* und *Wärme*. Der Screenshot in Abbildung 8² zeigt den zweiten Durchgang. Diese Analogie wird aber nur der EG2 gegeben, mit der Bitte, sie zu nutzen und die Bezeichner für die abstrakten Variablen einzustellen. Für sie ergibt sich also eine ganz neue Quelldomäne mit neuen Strukturen, die den eventuell schon bekannten Strukturen aus Durchgang 1 gleich zugeordnet werden können und/oder für neue Ideen (Hypothesen) über mögliche Strukturen sorgen. Für die EG1 bleibt die Mensa-Analogie, die aber immer noch in den Drop-down-Menüs verstellt werden kann.

What Is The System (WITS.prm) (Durchgang 2/4)								
		Takt 1/7	Takt 2/7	Takt 3/7	Takt 4/7	Takt 5/7	Takt 6/7	Takt 7/7
A	gr. Käfer	100	150	205	290	484	433	
B	kl.Käfer	200	220	342	176	393	233	
I	Wärme	0	100	-200	0	0		
II	Licht	0	0	0	-100	100		
								weiter
		Zeige Text 1		Zeige Text 2		Zeige Text 3		

Abbildung 8: Screenshot des dritten Durchgangs der Experimentalgruppe (EG2).

2 Der Screenshot in Abbildung 8 zeigt den Durchgang des DYNAMIS-Systems unter einem Windows-Betriebssystem, alle anderen Screenshots entstanden unter Linux.

Das Abfallsystem.

Das Abfallsystem beinhaltet die vier Variablenbezeichnungen *Anzahl der Entleerungen*, *Gestank aus den Tonnen*, *Füllstand der Tonnen* und *Menge des Abfalls*. Diese standen ausformuliert im dritten Instruktionstext und in abgekürzter Form in den Drop-down-Menus (siehe Abbildung 9).

What Is The System (WITS.prm) (Durchgang 3/4)								
		Takt 2/7	Takt 3/7	Takt 4/7	Takt 5/7	Takt 6/7	Takt 7/7	Takt 8/7
A	Gestank	150	205	265	282	666	481	502
B	Fülle	220	242	306	336	820	82	90
I	Abfall	0	40	80	50	-300	0	3/4
II	Leerungen	0	0	40	-200	260	0	
								OK !
		Zeige Text 1	Zeige Text 2	Zeige Text 3	Zeige Text 4			

Abbildung 9: Screenshot zum Durchgang 3 der Experimentalgruppe (EG2).

Auch hier galt es für die EG2, noch einmal aktiv über Strukturmerkmale nachzudenken und auf neue Ideen zu kommen.

Die taktische Kontrollgruppe bekam als gesondertes Treatment die folgenden fünf Tipps durch einen Instruktionstext vorgelegt (vgl. Abschnitt 'Taktiken' auf Seite 43):

- i. Hypothesen: Bilden Sie sich immer wieder Hypothesen, was passieren wird, wenn diese oder jene Variable von Ihnen verändert wird.
- ii. Einzeleingriffe: Verändern Sie nicht immer alle Variablen gleichzeitig, sondern lassen Sie auch mal eine Variable leer.
- iii. Nulleingriffe: Um Dynamiken zu erkennen sollten Sie auch mal nichts verändern.
- iv. Starke Eingriffe: Um Beziehungen eindeutig zu erkennen nützen hohe Eingabewerte.
- v. Negative Ausgabevariablen: Steuern Sie die Ausgabevariablen nacheinander auf Null oder in den negativen bzw. positiven Bereich.

Auch die taktische Kontrollgruppe (KG2) wurden nur gebeten über die Taktiken nachzudenken und sie gegebenenfalls sinnvoll einzusetzen.

Auch bei einem einfachen Design spielt die Größe der Stichprobe eine wichtige Rolle. Je nach Datenanalyseverfahren können unterschiedlich große Effektstärken statistisch bedeutsam werden. Daher gilt diesen Überlegungen der nächste Abschnitt.

4.6 Vorüberlegungen zur Stichprobengröße

Ein Internetexperiment hat den Vorteil, dass es erlaubt die Datenerhebung zu automatisieren. Unabhängig vom Aufenthaltsort des Versuchsleiters können Testungen durchgeführt werden und dies zeit- und relativ rechnerunabhängig. Nur eine Internetverbindung muss bestehen, und ein javafähiger Internetbrowser (Mozilla, Netscape Navigator, Internet Explorer, Opera etc.) genutzt werden. Mit Internetexperimenten verbindet sich die Hoffnung auf große Stichproben. Andere Internetexperimente unter ähnlichen Bedingungen zeigen, dass über 40 Versuchspersonen pro Woche möglich sind (Blümke & Friese, 2003, S. 50). Bei solchen Stichprobengrößen stellt sich schnell die Frage, ob nicht eine „optimale“ Stichprobengröße überschritten werden kann. Für kleine Stichproben existieren Wertebereiche, in denen die H_0 und eine spezifische H_1 miteinander vereinbar sind. Eine spezifische H_1 ließe sich z.B. über die Festlegung auf mittlere Effekte berechnen. Umgekehrt gilt für zu große Stichproben, dass es Wertebereiche gibt, in denen sowohl die H_0 als auch die H_1 abgelehnt werden müssten (Bortz, 1993, S. 121).

Zur Festlegung der Effektstärke ε werden im Rahmen der üblichen Konventionen nach Cohen (1988) die Werte für mittlere Effektstärken gewählt. Auch seiner Empfehlung, das β -Fehler-Niveau gegenüber dem α -Fehler-Niveau zu vervierfachen, wird nachgegangen mit $\alpha = .05$ und $\beta = .20$. Die Teststärke $1-\beta$ läge damit bei .80.

Für verschiedene Signifikanztests ergeben sich unterschiedliche, optimale Stichprobengrößen. Sie wurden mit Hilfe des Programms G*Power (Erdfelder, Faul & Buchner, 1996; Buchner, Erdfelder & Faul, 2001) als apriori-Analysen durchgeführt. Einige sollen hier aufgeführt werden. Für alle gelten die genannten α - und β -Niveaus. Die mittlere Effektstärke ε für den Test wird mit angegeben.

Für die Korrelationen gilt eine optimale Stichprobengröße von $n = 64$ bei einer kritischen Fehlergröße von $t(62) = 1.6698$ und mittlerer Effektstärke $\varepsilon = 0.3$.

Für einen mittelwertsvergleichenden t-Test ergibt sich eine Stichprobengröße von $n = 102$ für eine mittlere Effektgröße von $\varepsilon = 0.5$.

Für eine ANOVA mit zwei Gruppen und mittlerer Effektstärke $\varepsilon = 0.25$ (entspricht einer Varianzaufklärung η^2 von ca. 6%) ergibt sich eine Stichprobengröße von $n = 128$ (entspricht $n = 64$ pro Gruppe), die sich auf $n = 180$ (entspricht $n = 45$ pro Gruppe) erhöht, wenn vier statt zwei Gruppen in die ANOVA eingehen.

Eine Regression mit zwei Prädiktoren bei einer mittleren Effektstärke $\varepsilon = 0.15$ benötigt ein n von 68, bei drei Prädiktoren steigt n auf 77.

Sinnvolle Stichprobengrößen schwanken also bei mittleren Effektstärken je nach Verfahren zwischen $n = 64$ und $n = 180$. Bei einer geplanten Laufzeit von 12 bis 14 Wochen und ca. 20 bis 40 Versuchspersonen pro Woche lagen die Schätzungen zwischen 260 und 520 Teilnehmenden. Dass diese Schätzungen theoretischer Natur waren und bleiben, zeigt der nachfolgende Abschnitt, der die Phase der Datenerhebung über das Internet beschreibt, wie sie in der Praxis ablief.

4.7 Untersuchungsdurchführung über das Internet

Die Phase der Datenerhebung erstreckte sich von Anfang Oktober 2002 bis zunächst Weihnachten 2002 und wurde dann einige Male bis letztlich zum 18. März 2003 verlängert. Um das Programm unter Belastung zu testen, wurde zu Anfang eine entsprechende Meldung im Usenet unter 'de.comp.lang.java' gepostet (siehe Anhang B). Dies erbrachte viele Teilnehmer, die das Programm antesteten, dann aber schon im ersten Durchgang abbrachen. Anfang Oktober 2002 war dieses Verhalten zunächst auch noch erwünscht, vor allem von den Programmierern der genannten Newsgroup, da diese auf Mängel im Programmablauf hinweisen konnten. So konnten in dieser Anfangsphase noch einige kleine Änderungen vorgenommen werden.

Weitere Postings erfolgten in deutschsprachigen Newsforen, die sich meist direkt dem Thema Psychologie oder zumindest Umfragen widmeten, darunter

z-netz.wissenschaft.psychologie,
de.sci.psychologie,
de.alt.umfragen und
bln.lv.tub.psychologie.dppp-mp

Es ergaben sich zwar einige Diskussionen, vor allem in de.sci.psychologie, der Zulauf an Versuchspersonen zum Experiment durch diese Gruppen muss aber als sehr gering eingeschätzt werden, da dieser sich nach den jeweiligen Postings nicht steigerte.

Auch die Möglichkeit, auf speziellen Internetseiten auf das Experiment hinzuweisen, wurde angestrebt. Teilweise ist die Aufnahmedauer auf entsprechenden Seiten jedoch länger als sechs Monate, und ob tatsächlich der Link aufgenommen wird, ist sehr fraglich. Weder die Anmeldungen bei yahoo.de noch bei psycube.de wurden jemals geschaltet. Diese hätten für eine bessere Trefferrate in Suchmaschinen gesorgt.

Als nächstes wurden Versuchspersonen über Plakate am Heidelberger Psychologischen Institut geworben. Hinzu kamen Emails, die über den internen Mailverteiler des Instituts an die verschiedenen Semester gesendet wurden, zunächst an das erste, später dann auch an das

dritte Semester.

In diesen Mails war der Aufruf enthalten, die Mail an Freunde und Bekannte weiterzuleiten, so dass sich per Schneeball-Verfahren eine große Menge an potentiell interessierten Teilnehmern bilden sollte. Einen solchen Aufruf enthielt auch die Email, die von mir privat an Freunde und Bekannte versandt wurde. Es ist davon auszugehen, dass diese Art der Werbung für die größte Teilnehmerzahl am Experiment sorgte. Evaluiert wurde die Werbewirksamkeit der verschiedenen Verfahren jedoch nicht. Sie kann nur anhand der Eingangsdaten auf dem Server abgeschätzt werden.

Als gegen Ende der Untersuchung die Zelle der Kontrollgruppe immer noch nur sehr spärlich besetzt war, wurden einige Versuchspersonen direkt angesprochen und nahmen in Anwesenheit des Versuchsleiters am Experiment teil. Dies aber auch nicht in Laboratmosphäre sondern, in privaten Wohnungen mit Internetverbindung. Zum Abschluss dieses Kapitels soll auf die verschiedenen Datenanalyseverfahren eingegangen werden, um auf detaillierte Ausführungen im Ergebnisteil verzichten zu können.

4.8 Datenanalyseverfahren

Die Untersuchung lief vollständig mit dem in Java programmierten DYNAMIS-System über das Internet. Die von den Versuchsperson erzeugten Daten wurden vom System in einzelnen Logfiles auf dem Server abgelegt. Der Dateiname besteht dabei aus der Gruppenzugehörigkeit plus einer Ganzzahl, die die Millisekunden seit dem 01.01.1970 angibt. Die Logfiles setzen sich aus vier Teilen zusammen. Der erste Teil beinhaltet die soziodemographischen Daten sowie vom Computer erfragte Daten über das Computersystem der Versuchsperson. Der zweite Teil umfasst die Prozessdaten der vier Durchgänge inklusive der Analogiebezeichnungen aller Takte. Der dritte Teil beinhaltet die Daten aus der Wissensdiagnostik, und der letzte Teil umfasst Antworten der Versuchspersonen auf offene Fragen, die freiwillig ausgefüllt werden konnten.

Die Daten wurden in SPSS 11.0 für Windows ausgewertet. Dazu wurde ein Java-Programmpaket geschrieben, das die einzelnen Logfiles einliest und jede Zeile in ihre Inhalte zerlegt. Anschließend erfolgen bereits in diesem Java-Programmpaket einige Aggregationen, wie z.B. die Auszählung der verwendeten Taktiken oder die Veränderungen der Analogiebezeichner. Die relevanten Daten aus den ersten drei Teilen der Dateien werden daraufhin in eine reine Textdatei umgewandelt, die dann über Syntax-Befehle in SPSS eingelesen werden kann. Der jeweilige vierte Teil der Logfiles enthält auch die Email-Adresse derer, die an der Gewinnauslosung teilnehmen wollten. Diese Teile werden, um Anonymität

wahren zu können, nicht mit in die Datendateien aufgenommen.

Die statistische Analyse umfasst sowohl deskriptive als auch inferenzstatistische Auswertungen. Bei den deskriptiven Angaben handelt es sich um Kennwerte, wie arithmetisches Mittel und Standardabweichung bei den intervallskalierten Variablen. Für die nominalen Daten werden Häufigkeiten und Median angegeben. An manchen Stellen wird für die intervallskalierten Daten sowohl Mittelwert als auch Median angegeben, dies geschieht, um die Schiefe der Verteilung abschätzen zu können, wenn keine anderen Werte der Schiefe vorliegen (Bortz, 1993, S. 45).

Für die statistische Darstellung des Zusammenhangs werden für die intervallskalierten Variablen die Produkt-Moment-Korrelationskoeffizienten (r) nach Pearson angegeben (z.B. in Hypothese III). Sollten die Voraussetzungen für die statistische Absicherung der Korrelationskoeffizienten nicht erfüllt sein, so wird der Korrelationskoeffizient nach Spearman (r_s) berechnet (z.B. in Hypothese II). Im Text wird an der entsprechenden Stelle noch einmal darauf eingegangen.

Für die Mittelwertsvergleiche (z.B. Hypothese I) werden einfaktorielle Varianzanalysen (ANOVA) bzw. t-Tests berechnet. Des Weiteren werden hierarchische Regressionsanalysen eingesetzt, um den Anteil der Varianzaufklärung durch einzelne Variablen bestimmen zu können.

Drei Bedingungen müssen zur Verwendung einer ANOVA gegeben sein: die Normalverteilung der Fehlerkomponenten, die Varianzhomogenität und die Unabhängigkeit der Fehlerkomponenten (Bortz, 1999, S. 274). Eine Überprüfung auf Normalverteilung der Fehlervarianzen wird in der Praxis selten vorgenommen (Bortz, 1999, S. 276), da Signifikanztests bei größeren Stichproben ($n > 30$) robust reagieren.

Abweichungen der Normalverteilung oder Auffälligkeiten im Levene-Test werden an den entsprechenden Stellen des Ergebnisteils angesprochen.

5. Ergebnisse

Neben den Ergebnissen zu den drei Hypothesen enthält dieser Teil der Diplomarbeit zusätzliche Abschnitte mit Analysen über weitere abhängige Variablen. Zur besseren Verständlichkeit werden diese Abschnitte denen der Hypothesen jeweils vorgeordnet. Die zentrale Variable für die jeweilige Hypothese ist dadurch bereits bekannt und genauer definiert, so dass die Ergebnisse der Hypothesen leichter verständlich sind.

Zunächst wird eine ausführliche Beschreibung der Stichprobe (5.1) mit den soziodemographischen Daten aus dem Fragebogen erfolgen. Dort soll auch die große Zahl der Abbrecher dargestellt werden. Es folgt ein Abschnitt, der die Verwendung der Analogien durch die Versuchspersonen prüft und darstellt. Dieser Abschnitt über die Nutzung der Analogien (5.2) dient sowohl als Treatmentcheck als auch zur Vorbereitung auf die Ergebnisse zur Hypothese I (5.3). Diesen folgt dann eine genaue Analyse über die Nutzung der Taktiken (5.4), bevor die Ergebnisse der Hypothese II (5.5) erläutert werden. Die Bewertung der Steuerleistung wird im Abschnitt 5.6 dargestellt und über ein erstes Regressionsmodell beschrieben, bevor sich Abschnitt 5.7 mit den Ergebnissen zur Hypothese III über die Auswirkung von Strukturwissen auf die Steuerleistung befasst. Der abschließende Abschnitt widmet sich der Analyse der Bearbeitungszeiten (5.8) der einzelnen Gruppen und Phasen im Experiment.

5.1 Beschreibung der Stichprobe

Im Zeitraum zwischen dem 19.10.2002 und dem 18.03.2003 aktivierten insgesamt 368 Internetsurfer das Programm und erzeugten damit eine Ausgabedatei auf dem Server. Von diesen $N=368$ Teilnehmern absolvierten jedoch nur ca. $n = 66$ das Experiment vollständig bis zum Ende. Im folgenden wird zunächst auf die Abbrecher eingegangen, bevor die Verteilungen für Geschlecht, Alter und Bildung beschrieben werden. Zum Abschluss des Abschnittes werden die Angaben der Versuchspersonen über ihr Arbeiten mit Computern und die vom Computer abgefragten Daten beschrieben.

Die Abbrecher.

Von den 302 Teilnehmern, die das Experiment abbrachen, haben nur zwei das Programm bis zum zweiten Durchgang bearbeitet, nur diese beiden haben also einen vollständigen Durchgang und die erste Wissensdiagnostik gesehen. 155 brachen noch vor dem Ende des ersten Durchganges ab, so dass von diesen nur Daten über ihr Computersystem und der demographische Fragebogen vorliegen. 66 Teilnehmer brachen sogar schon während des demographischen Fragebogens ab, so dass nur einige Informationen über die

Computersysteme vorliegen.

Über die weiteren 79 Datensätze kann nur gesagt werden, dass für sie zwar vom Server eine Datei angelegt wurde, diese aber nicht ein Byte an Information enthält. Gegeben ist nur das Datum und die Zuordnung zur experimentellen Bedingung, beides liegt im Dateinamen vor. Zunächst wäre ganz einfach davon auszugehen, dass es sich bei diesen 79 Teilnehmern um Abbrecher handelt, die noch vor allen anderen den Versuch beendeten. Dies kann, muss aber nicht der Fall sein, da es sich hier eher um Datenverlust aufgrund technischer Inkompatibilitäten handelt, die nicht ohne weiteres geklärt werden können. Denn die Informationen über die Computersysteme werden gleich zu Anfang, ohne ein weiteres Zutun des Teilnehmers abgefragt und bereits beim Anlegen der Datei in eben diese geschrieben. Bei einigen liegt sogar die vollständige Bearbeitung des Experimentes vor und dennoch blieb die Datei leer. Dies ist zumindest bei vier Heidelberger Studierenden sicher, die einen Code am Ende des Versuches erhielten und diesen später beim Versuchsleiter für Versuchsstunden einreichten. Rechnet man diese 79 Fehlerdateien heraus, dann kann man aus den verbleibenden 289 den Prozentsatz der 66 vollständigen Datensätze mit 22.84% bestimmen, d.h. nicht einmal jeder vierte hat das Experiment vollständig bearbeitet. Da nur wenig über die Abbrecher bekannt ist, kann eben auch eine systematische Verzerrung nicht ausgeschlossen werden. Die Zufälligkeit und Repräsentativität der Stichprobe ist daher schon an dieser Stelle deutlich in Frage zu stellen und wird Thema im Diskussionsteil sein.

Von den 66 Teilnehmenden gehen in die statistischen Untersuchungen nur $N = 53$ Fälle ein, da fünf Personen auf die Frage der Erstteilnahme mit „Nein“ antworteten, d.h. mehrfach teilgenommen hatten, und eine den demographischen Fragebogen nicht korrekt ausfüllte. Des weiteren wurden sieben Abbrecher aussortiert, die das System zwar bis zum Ende „bearbeiteten“, jedoch keine Zahlen mehr eingaben. Das Kriterium für die Zuordnung zum Abbrecher wurde auf 13 oder mehr Nulleingriffe innerhalb der ersten 21 Takte, also der ersten drei Durchgänge, festgelegt. Dies wurde z.B. erreicht, wenn jemand nach dem ersten Durchgang nur noch den „Weiter“-Button betätigte. Andere Abbrecher-Kriterien sind sicherlich denkbar und ein wichtiges Thema, welches im Diskussionsteil dieser Arbeit angesprochen werden wird.

Die aufgrund der Abbruchrate geringe Stichprobengröße von $N = 53$ hat entsprechende Auswirkungen auf die Teststärke aller Signifikanztests dieses Kapitels. So ergibt sich beispielsweise im Extremfall für eine ANOVA mit vier Gruppen bei angenommenen mittleren Effekten von $\varepsilon = .25$ und einem angenommenen Signifikanzniveau von $\alpha = .05$ eine Teststärke von $1 - \beta = .28$ und damit ein β -Fehler von $.72$. Um bei dem wie oben angenommenen β -

Niveau von .20 (vgl. Abschnitt 'Vorüberlegungen zur Stichprobengröße' auf Seite 65) eine Signifikanz auszuweisen, müsste ein sehr großer Effekt von $\varepsilon = .475$ vorliegen. Ähnlich verhält sich der β -Fehler und die mindestens vorliegende Effektgröße für die t-Tests ($\beta = .64$). Für die Korrelationen hingegen ist das Kriterium von 64 Versuchspersonen fast erreicht, so dass dort β nur auf .27 für mittlere Effekte steigt, was einer Teststärke von immerhin 73% entspricht.

Viele anderen psychologischen Untersuchungen weisen geringe Teststärken auf, so fanden z.B. Sedlmeier und Gigerenzer (1989) bei einer Durchsicht der Zeitschrift *Journal of Abnormal Psychology* (Jahrgang 1984) in 64 Experimenten eine durchschnittliche Teststärke von 37% für die Aufdeckung eines mittleren Effektes. Daher sollen auch hier die Signifikanztests erfolgen, jedoch mit dem Hinweis auf die geringen Teststärken.

Geschlecht, Alter und Bildung.

Die 53 Versuchspersonen der vollständigen Datensätze verteilen sich gleichmäßig auf die vier Gruppen. Die unspezifische Kontrollgruppe hat $n = 12$, die taktische Kontrollgruppe $n = 13$ und die beiden Experimentalgruppen je $n = 14$ Versuchspersonen. Die Geschlechter verteilen sich mit $n = 25$ Frauen und $n = 28$ Männern etwas untypisch für eine psychologische Laboruntersuchung. Hier sieht man bereits einen Effekt des Internets auf die Stichprobe, da bei einer vergleichbaren Untersuchung, die nur am Psychologischen Institut Heidelberg erfolgt wäre, eher ein Geschlechterverhältnis von 70% zu 30%, also 37 Frauen zu 16 Männern zu erwarten gewesen wäre.

Das Alter der Teilnehmenden lag zwischen 19 und 56 Jahren mit einem Mittelwert von $M = 27.3$, einem Median von $Md = 27$ und einer Standardabweichung von $SD = 6.62$. Die

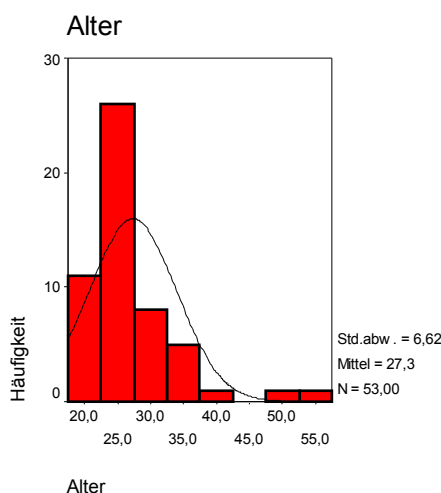


Abbildung 10: Altersverteilung in der Stichprobe.

Altersverteilung ist deutlich rechtsschief mit einem Wert von 2.218. Bei den Darstellungen der Häufigkeitsverteilungen (z.B. Abbildung 18) ist jeweils die Kurve für die Normalverteilung eingearbeitet, so dass auch ein optisches Urteil über die Verwendbarkeit der jeweiligen Variable für parametrische Testverfahren möglich ist. Zudem werden auch in den Tabellen die Werte für die Schiefe eingetragen, die bei Werten größer +1 oder kleiner -1 auf ein zu großes Abweichen von der Normalverteilung hinweisen.

Die schulische Bildung der Versuchspersonen ist im Vergleich zum Bundesdurchschnitt sehr hoch. Abgeschlossenes Abitur haben immerhin $n = 41$, ein abgeschlossenes Studium $n = 10$, und bei zwei Teilnehmenden fehlte die Angabe. Die letzte Schulabschlußnote reicht von 3.6 bis 1.0 mit einem Mittelwert von $M = 2.09$ und einem Median von $Md = 1.9$ bei einer Standardabweichung von $SD = 0.67$.

Tabelle 5: Angaben der Versuchspersonen zu ihrem Berufsstand

	Häufigkeit	Prozent	kumulierte Prozente
Angestellte(r)	9	17,0%	17,0%
Beamte(r)	1	1,9%	18,9%
Doktorand(in)	1	1,9%	20,8%
selbständig	5	9,4%	30,2%
Student(in)	36	67,9%	98,1%
Umschüler(in)	1	1,9%	100,0%

Mehr als 2/3 der Teilnehmenden waren Studierende. Die anderen Berufsgruppen sind der Tabelle 5 zu entnehmen.

Computerangaben.

Innerhalb des Fragebogens wurden drei Fragen zum Computerumgang gestellt (siehe Tabelle 6 bzw. die Fußnoten 3, 4 und 5). Zunächst wurde gefragt, wie gerne mit dem Computer gearbeitet und über wieviel Erfahrung verfügt wird. Beides sind subjektive Einschätzungen, die auf einer fünfstufigen Skala anzugeben waren. Sie wurden erhoben, um die Einstellung zu Computern und die subjektive Expertise abzufragen. Die dritte Frage zur Dauer der Computernutzung zielt auf das konkret beobachtbare Verhalten. Diese Angabe sollte als Indikator für die tatsächliche Nutzung von Computern dienen. Die Angabe konnte in Stunden eingegeben werden und erzielte einen Mittelwert von $M = 3.72$ bei einer Standardabweichung von $SD = 3.98$. Die hohe Standardabweichung kommt durch wenige hohe Werte zustande. Daher mag der Median dieser Verteilung mit $Md = 3.0$ etwas aussagekräftiger über die

Computerzeit pro Woche sein. Die anderen beiden Fragen verteilen sich annähernd normal um den Mittelwert $M = 3.58$ bzw. $M = 3.89$ bei Standardabweichungen von $SD = 0.87$ bzw. $SD = 0.87$ (siehe Tabelle 6 bzw. Abbildung 11). Erwähnenswert dabei ist nur das Auslassen des minimalen Wertes („1=gar nicht gerne“ bzw. „1=gar keine Erfahrung“). Alle Versuchspersonen verfügen demnach über zumindest geringe Kenntnisse, und kein Teilnehmer lehnt das Arbeiten mit Computern völlig ab.

Tabelle 6: Subjektive Einschätzung des Computerumgangs und Dauer der Nutzung.

	N	Minimum	Maximum	M	Md	SD	Schiefe
Computererfahrung ³	53	2	5	3.58	4	0.87	0.06
Computerliking ⁴	53	2	5	3.89	4	0.87	-0.50
Computerzeit ⁵	52	0.5	24	3.72	3.0	3.98	3.31

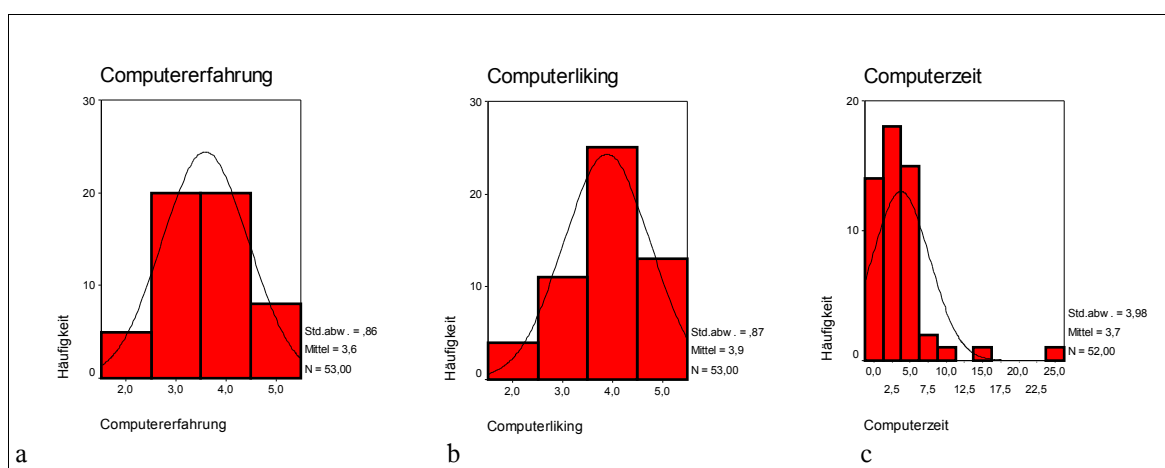


Abbildung 11: Häufigkeiten der Befragung nach Erfahrung, Neigung und verbrachter Zeit mit Computern.

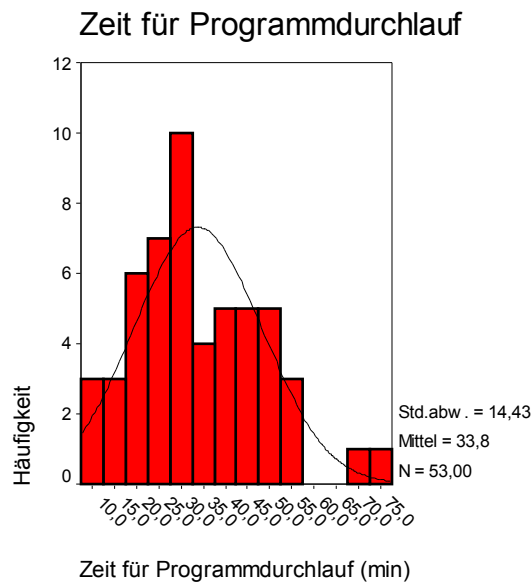
Des Weiteren wurde die Art des Internetzugangs erfragt. Diese Frage zielte auf einen möglichen Zeitdruck, dem Versuchspersonen mit Modem- oder ISDN-Verbindungen ($n = 28$) im Vergleich zu Nutzern von Standleitungen ($n = 25$), die nicht pro Minute abgerechnet werden, ausgesetzt sein könnten. Es zeigten sich aber keine Unterschiede in der Bearbeitungszeit oder anderen abhängigen Variablen zwischen diesen Gruppen.

Von den vom Computer automatisch abgefragten Daten sollen nur einige erwähnt werden, da sich hier nichts Außergewöhnliches abzeichnet. So gab es nur drei Teilnehmende, die nicht unter einem der Windows-Derivate als Betriebssystem arbeiteten. Die Bildschirmauflösungen bewegten sich zwischen SVGA (800x600) und SXGA+ (1400x1050), für die die Darstellung des Programms durchaus geeignet war.

3 Wieviel Erfahrung haben Sie mit Computern? [1= gar keine - 2 - 3 - 4 - 5= sehr viel]

4 Wie gerne benutzen Sie Computer? [1= gar nicht gerne - 2 - 3 - 4 - 5= sehr gerne]

5 Wie häufig benutzen Sie einen Computer? Stunden pro Tag: []



Kurz erläutert wurden die Analogien in den Instruktionstexten für die Experimentalgruppen jeweils direkt vor dem neuen Durchgang. Ein Hinweis auf die Nutzung könnte daher der Abruf dieser Texte sein. Zu erwarten wäre ein häufigerer Abruf dieser Texte von der Experimentalgruppe EG2 gegenüber den anderen Gruppen, da deren Instruktionstexte ab dem zweiten Durchgang nur noch aus einem einfachen Hinweissatz⁶ auf den nächsten Durchgang bestanden.

Wie die Abbildung 13 zeigt, ergibt sich kein einheitliches Bild der Informationsnachfragen. Zudem liegen die Mittelwerte dieser Abfrage so dicht beieinander und das bei sehr großen Standardabweichungen (siehe Tabelle 7), dass hier nicht von einem Trend oder einer statistischen Bedeutsamkeit ausgegangen werden kann ($F(3,52) = 0.11$; $p = .95$).

*Tabelle 7: Mittelwerte und Standardabweichungen
der Informationsaufrufe.*

Gruppe	Abrufe aller Instruktionstexte	
KG1	<i>M</i>	29,42
	<i>SD</i>	21,32
KG2	<i>M</i>	28,31
	<i>SD</i>	39,91
EG1	<i>M</i>	30,36
	<i>SD</i>	14,94
EG2	<i>M</i>	24,79
	<i>SD</i>	26,97

Zum Vergleich ist in Abbildung 14 noch die Abfrage des Zieltextes angegeben. Den Zieltext gab es nur im letzten Durchgang, und er enthielt die genauen Zielwerte, die erreicht werden sollten, angegeben für die Variablen A und B. Die Tendenzen kehren sich hier scheinbar um, die Experimentalgruppe EG1 ruft im Durchschnitt nur 0.86 mal zusätzlich den Text mit den Zielen ab, die Gruppe EG2 hingegen mit $M = 3.71$ deutlich öfter. Aber auch hier liegen keine statistischen Bedeutsamkeiten ($F(3,52) = 2.167$; $p = .104$) vor, und auf spekulative inhaltliche Interpretationen soll hier verzichtet werden.

Die Nutzung der Variablenbezeichner mit Hilfe der Drop-Down-Menüs wird in Abbildung 15 veranschaulicht. Es zeigt sich, dass jeweils zum ersten Takt eines jeden Durchgangs die Bezeichner verändert wurden. Für die Gruppe mit nur einer Analogie (EG1 in der oberen Reihe der Graphik) verringert sich dieser Wert deutlich vom ersten Durchgang zu den

⁶ „Dieser Probedurchgang ist nun zu Ende. Es folgt ein weiterer Durchgang zum Üben.“

Für die anderen Instruktionstexte siehe Anhang C.

folgenden, da die Variablenbezeichnungen in der gewählten Einstellung blieben. Bei der Gruppe mit drei Analogien (EG2 in der unteren Reihe der Graphik) ist ein deutlicher Abfall des über die 14 Versuchspersonen gemittelten Wertes erst zum Zieldurchgang hin zu beobachten. Auch hier blieben die Bezeichner in der Stellung, die ihnen in Durchgang 3 gegeben wurde.

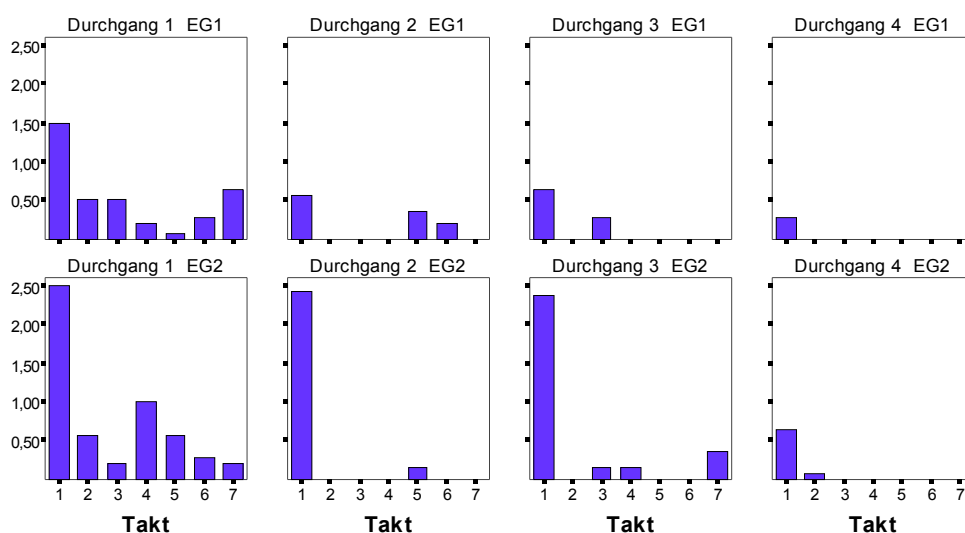


Abbildung 15: Anzahl der Wechsel der Variablenbezeichner je Takt für Gruppen und Durchgänge getrennt. Gemittelt über die je $n = 14$ Versuchspersonen pro Gruppe.

Wenn alle 14 Versuchspersonen einer Experimentalgruppe die jeweilige Analogie einmal für sich passend mit den Variablenbezeichnern umstellen, so müsste sich eine Summe von mindestens 42 Wechseln ergeben (3×14 , da einer der vier Bezeichner stehen bleiben kann). In den jeweils ersten Durchgängen stellte die EG1 52 Bezeichner um und die EG2 sogar 75. In Durchgang 2 und 3 sank diese Zahl für die EG2 auf 36 bzw. 42.

Es ist erfreulich zu sehen, dass die Analogien genutzt wurden und das von der EG2 stärker als von der EG1, was so auch zu erwarten war. Ob nun die Bezeichner nur umgestellt wurden, um alle vier lesen zu können oder ob ein tatsächliches Mapping mit den Strukturmerkmalen der Analogie und des WITS-Szenarios bei den Versuchspersonen vorliegt, lässt sich nicht mehr bestimmen. Insgesamt wurden aber höhere Zahlen erwartet als hier vorliegen. Dies kann aber auch daher kommen, dass einige sich gar nicht auf die Analogien einließen, da es in der Untersuchung auch keine notwendige Bedingung war.

Nachfolgend soll in den Ergebnissen zur Hypothese I der Kausalzusammenhang zwischen dem Wissenserwerb und den Analogiegruppen (EG1 und EG2) dargestellt werden.

5.3 Ergebnisse zur Hypothese I

Erwartet wird, dass die Experimentalgruppen (EG1 und EG2) mehr Strukturwissen erlangen als die Kontrollgruppen (KG1 und KG2) und dies unabhängig vom Einsatz der Taktiken. Genauer betrachtet ergeben sich dadurch drei Fragen:

Konnten die Experimentalgruppen ($EG2 > EG1$) mehr Wissen erwerben als die Kontrollgruppen (1) und im besonderen mehr als die unspezifische Kontrollgruppe (2)? Ist der Mehrerwerb an Strukturwissen auf die vermittelnde Nutzung von Taktiken zurückzuführen, oder besteht er auch unabhängig davon (3)?

Aus dem Wissenstest stehen drei Variablen zur Verfügung: eine über die Wirkzusammenhänge, eine über die Wirkrichtung und eine über die Wirkstärke (siehe 'Wissensanwendung' auf Seite 21). Dabei wurde ein Zähler für jede dieser Variablen erhöht, wenn eine richtige Antwort vorlag (Treffer). Nichts veränderte sich, wenn Versuchspersonen die Grundeinstellung (-- ? --) nicht veränderten. Für falsche Angaben wurde ebenfalls ein Zähler erhöht (Fehler), so dass auch drei Wissensvariablen mit Fehlern vorliegen. Die Ergebnisse dieser drei Variablen mit den Treffern bzw. Fehlern wurden dann zu je einer Wissensvariablen aggregiert.

Für die erste Frage sind die Mittelwerte der Gruppe mit versus ohne Treatment zu betrachten. Sie sollten für den Fall der Treffer beim Wissenstest für die Experimentalgruppen ($M = 5.75$) größer sein als für die Kontrollgruppen ($M = 4.84$). Auch wenn ein Trend zu sehen ist, wird dieser nicht signifikant ($F(1,52) = 0.472$; $p = .495$). Im Falle der Fehler im Wissenstest müsste der Mittelwert der Experimentalgruppen ($M = 4.86$) kleiner sein als der der Kontrollgruppen ($M = 6.96$). Dieser deutliche Trend wird nur knapp nicht signifikant ($F(1,52) = 3.063$; $p = .086$; siehe Tabelle 9).

Da beide Wissensvariablen (Treffer und Fehler) der Tendenz nach in der vorhergesagten Richtung liegen, böte sich auch die Bildung eines gemeinsamen Wissensindex an, bei dem analog dem Verfahren des Signal-Entdeckungs-Paradigmas (vgl. Bortz, 1995, S. 152ff) die Fehler von den Treffern abgezogen werden. Der einfacheren Beschreibung wegen wird an späteren Stellen dieser Arbeit auch auf diesen Wissensindex Bezug genommen. Für die Darstellung der Ergebnisse zu Hypothese I wird aber noch die etwas informativere Beschreibung beider Wissensvariablen genutzt.

Kontrastiert man für den zweiten Teil der Frage die beiden Experimentalgruppen gegen die unspezifische Kontrollgruppe, so erhöht sich zwar noch einmal der Mittelwertsunterschied in die durch die Hypothese vorhergesagte Richtung für die korrekten Antworten, doch weder

dieser Mittelwertsunterschied ($t(49) = 1.40$; $p = .166$) noch der in den Falschangaben ($t(49) = -0.817$; $p = .418$) wird bei angenommener Varianzgleichheit statistisch bedeutsam. Besser sehen die Werte der geplanten Kontraste vor allem für die Treffer aus, wenn aufgrund von nicht gleicher Varianz korrigiert wird ($t(30.934) = 1.824$; $p = .075$). Im Fall von nicht gleicher Varianz zeigt sich für diesen Kontrast für den Wissensindex, der stets etwas höhere t- oder F-Werte erreicht, ein signifikanter Mittelwertsunterschied ($t(26.396) = 2.230$; $p = .034$).

Tabelle 8: Anzahl korrekter Angaben im Wissenstest vor dem Zieldurchgang

		<i>N</i>	<i>Mittelwert</i>	<i>Standard- abweichung</i>	<i>Standardfehle- r</i>
Strukturwissen (Treffer)	KG	25	4.84	4.41	0.88
	KG1	12	3.42	2.75	0.79
	KG2	13	6.15	5.30	1.47
	EG	28	5.75	5.14	0.97
	EG1	14	6.07	6.49	1.73
	EG2	14	5.43	3.55	0.95
	Gesamt	53	5.32	4.79	0.66

Tabelle 9: Anzahl falscher Angaben im Wissenstest vor dem Zieldurchgang

		<i>N</i>	<i>Mittelwert</i>	<i>Standard- abweichung</i>	<i>Standardfehle- r</i>
Strukturwissen (Fehler)	KG	25	6.96	4.69	0.94
	KG1	12	6.08	4.34	1.25
	KG2	13	7.77	5.02	1.39
	EG	28	4.86	4.06	0.77
	EG1	14	3.86	3.88	1.04
	EG2	14	5.86	4.13	1.10
	Gesamt	53	5.85	4.45	0.61

In der ersten Frage steckt zusätzlich noch die Behauptung, dass die Experimentalgruppe mit drei Analogien mehr Strukturwissen erlangt als die mit nur einer Analogie. Ein erster Eindruck von dem erworbenen Wissen mit Treffern und Fehlern wird bereits durch die Darstellung der Mittelwerte über die Treatment-Gruppen in Abbildung 16 vermittelt (siehe auch die kleingedruckten Zeilen in Tabelle 8 und 9).

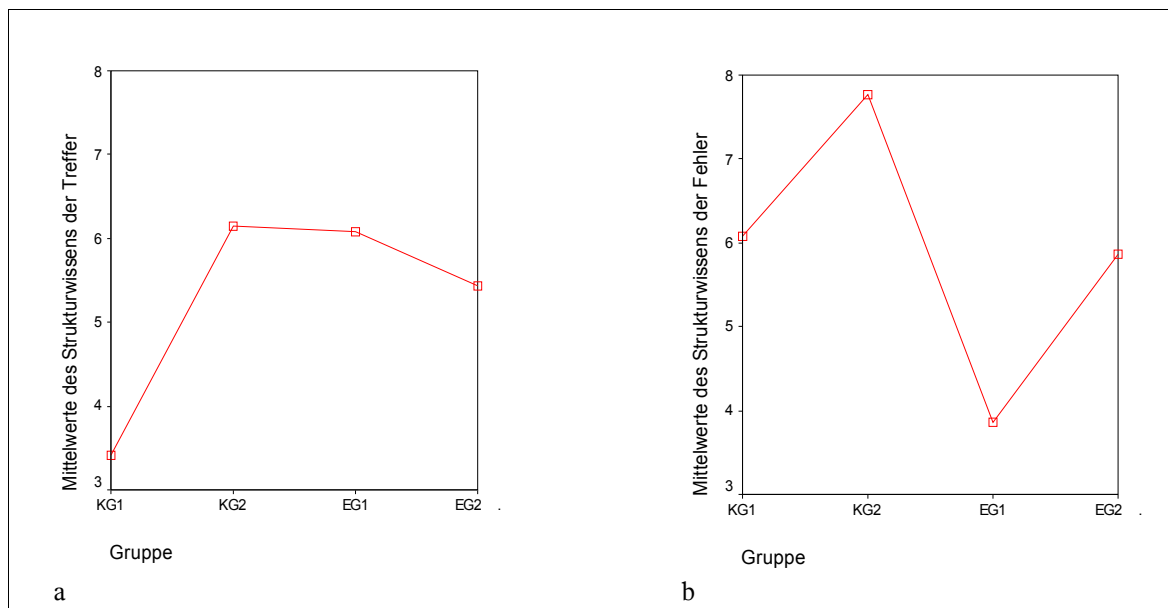


Abbildung 16: Mittelwerte des Strukturwissens für Treffer (a) bzw. Fehler (b) in den Gruppen.

Deutlich wird in dieser Abbildung, dass die unspezifische Kontrollgruppe (KG1) tatsächlich weniger Wissen erwirbt (Abbildung 16a) und dabei immer noch mehr Fehler macht (Abbildung 16b) als die Experimentalgruppen (sowohl EG1 als auch EG2). Was sich bereits abzeichnet, ist, dass die Experimentalgruppe mit nur einer Analogie (EG1) sowohl mehr Wissen erwirbt als auch weniger falsche Angaben macht als die Experimentalgruppe mit drei Analogien (EG2), was nicht unbedingt der Vorhersage entspricht. Der in den Abbildungen sichtbare Unterschied der Mittelwerte wird jedoch nicht statistisch signifikant. Die Unterschiede der einzelnen Gruppen werden bei einer ANOVA statistisch weder für die Treffer ($F(3,49) = 0.875$; $p = .461$) noch für die Fehler ($F(3,49) = 1.836$; $p = .153$) bedeutsam.

Offen ist noch die dritte Frage zu dieser Hypothese, ob der Wissenserwerb der Experimentalgruppen durch den Gebrauch von sinnvollen Taktiken modifiziert wird.

Zu diesem Zweck wurde je eine hierarchische Regressionsanalyse berechnet, die die Varianz der Strukturwissensvariablen, einmal für Treffer und einmal für Fehler, aufklären soll. In der ersten Ebene der jeweiligen Regressionsanalyse fließt nur die aggregierte Taktikvariable als Prädiktor ein. Als zweites folgt dann ein Modell, in dem zusätzlich zur Taktikvariable noch die nominale Variable der Gruppenzugehörigkeit unterschieden in Gruppe ohne Analogie (KG) und Gruppe mit Analogie (EG) einfließt.

Den Erwartungen der Hypothese I entsprechend, zeigt sich ein Gewinn an Varianzaufklärung vom einfachen Modell (1) zum Modell mit der Gruppenvariable (2). Die Steigerung der Varianzaufklärung ($R^2_{\text{Modell-2}} - R^2_{\text{Modell-1}} = .058 - .042 = .016$) für die Strukturwissensvariable

der Treffer beträgt nur 1.6% , und weder das erste Modell ($F(1,52) = 2.278$; $p = .137$) noch das zweite Modell ($F(1,52) = 1.557$; $p = .221$) werden statistisch bedeutsam.

Bedeutsamer wird die Varianzaufklärung für die Strukturwissensvariable der Fehler. Hier werden sowohl das erste ($F(1,52) = 4.510$; $p = .039$) als auch das zweite Modell signifikant ($F(1,52) = 4.784$; $p = .013$), und die Steigerung der Varianzaufklärung durch die Zuordnung der Gruppen beträgt immerhin 8.0% ($R^2_{\text{Modell-1}} = .081$; $R^2_{\text{Modell-2}} = .161$). Die Steigerung der Varianzaufklärung um 8.0% entspricht einer Effektgröße von $\varepsilon = 0.29$, welche zwischen den mittleren und großen Effekten für Regressionsanalysen nach Cohen (1988) einzuordnen wäre. Die Steigerung für die richtigen Antworten kommen mit einer Varianzaufklärung von 1.6% nur zu einem $\varepsilon = 0.13$, was noch unter den mittleren Effekten laut Cohen liegt.

Damit konnte gezeigt werden, dass das Treatment der Analogien beim Lösen komplexer Probleme zu einem tendenziellen Gewinn an Strukturwissen führt und das auch noch bei herauspartialisiertem Taktiknutzen. In Abschnitt 'Analyse der Zeiten' auf Seite 88 finden sich noch zusätzliche Hinweise auf eine Erleichterung beim Problemlösen durch Analogien.

Bevor nun die Ergebnisse zur Hypothese II dargestellt werden, erfolgt eine allgemeine Analyse der Taktiknutzungen in den verschiedenen Gruppen.

5.4 Analyse der Nutzung der Taktiken

Es wird erwartet, dass die taktische Kontrollgruppe (KG2) häufiger die sinnvollen Taktiken anwendet als andere Gruppen, da dies den Versuchspersonen in den Instruktionstexten geraten wurde. Der Instruktionstext riet zu Nulleingriffen, zu Einzeleingriffen und zu starken Eingriffen, ebenso wie zum Hypothesenbilden, und damit verbunden riet der Instruktionstext dazu, die Ausgabevariablen einmal nah an Null zu steuern bzw. sie auch in den negativen Bereich zu bringen (siehe auch 'Nulleingriffe, Einzeleingriffe, starke Eingriffe' auf Seite 44).

Tatsächlich gab es einige deutliche Unterschiede in der Verwendung einiger Taktiken (vor allem der Einzeleingriffe), andere wiederum konnten von keiner Gruppe genutzt werden (das gezielte Steuern der Ausgabevariablen), so dass sich keine signifikanten Unterschiede ergaben. Tendenziell zeigen die Mittelwertsunterschiede aber über alle Taktiken hinweg in die erwartete Richtung. Nachfolgend soll auf einige bedeutsame Unterschiede eingegangen werden. Eine Übersicht gibt Tabelle 10.

Tabelle 10: Mittelwerte (M) und Standardabweichung (s) aller Taktiken für die unspezifische (KG1) und die taktische (KG2) Kontrollgruppe sowie für die Experimentalgruppen mit einer (EG1) und drei (EG2) Analogien

Gruppe	Taktiken aggregiert	Null-eingriffe	Einzel-eingriffe	All-Eingriffe	starke Eingriffe	Ausgabevariablen nahe Null	Ausgabevariablen unter Null	
KG1	M	15.92	3.25	6.67	11.08	1.33	0.33	4.33
	SD	6.92	2.83	5.07	5.96	1.92	0.65	4.48
KG2	M	21.15	5.08	9.54	6.38	1.92	1.0	3.62
	SD	9.78	3.73	4.07	5.9	2.66	3.32	4.5
EG1	M	18.57	3.07	6.86	11.07	1.57	1.07	6.0
	SD	7.77	3.2	5.26	5.57	2.95	1.49	6.24
EG2	M	13.57	2.79	4.64	13.57	1.36	0.43	4.36
	SD	11.01	2.91	4.53	6.72	2.76	1.09	4.73
Gesamt	M	17.28	3.53	6.89	10.58	1.55	0.72	4.6
	SD	9.27	3.23	4.94	6.44	2.56	1.9	5.01

Das gezielte Steuern von Ausgabevariablen auf Werte unter Null, also in den negativen Bereich, wurde von allen Gruppen nahezu gleich genutzt, und die kleinen Unterschiede liegen hier im Bereich des Zufälligen. Ebenso wurde das Ansteuern einer Ausgabevariable in die Nähe von Null nur sehr selten erreicht, obwohl auch Werte von -10 bis +10 für eine Wertung zählten.

Deutliche Unterschiede treten bei der Nutzung der Eingabevariablen auf. Der als schlecht zu bewertende gleichzeitige Eingriff auf alle Variablen kommt bei der taktisch instruierten Kontrollgruppe (KG2) am seltensten vor ($M = 6.38$) im Unterschied zur Gruppe mit drei Analogien (EG2 mit $M = 13.57$) und den beiden anderen Gruppen (KG1 mit $M = 11.08$; EG1 mit $M = 11.07$). Dieser Unterschied ist bei einer Varianzanalyse überzufällig ($F(3, 52) = 3.274$; $p = .029$). Der Unterschied im Einsatz von Einzeleingriffen wird knapp nicht signifikant ($F(3,52) = 2.394$; $p = .080$). Auch hier unterscheidet sich die KG2 mit einer häufigen Nutzung von der EG2, die diese Taktik seltener anwendet. Ein ähnliches Bild ergibt sich für den Einsatz der Nulleingriffe, die wichtig sind, um die Eigendynamiken und Nebenwirkungen zu erkennen. Sie werden am häufigsten von Kontrollgruppe KG2 eingesetzt ($M = 5.08$), am seltensten von EG2 ($M = 2.79$). Jedoch werden diese Unterschiede nicht signifikant ($t(25) = 1.786$; $p = .086$). Das gleiche gilt für die Summe der verschiedenen Taktiken. Für diese aggregierte Variable wurden die Null-, Einzel- und starken Eingriffe addiert, ebenso wie die beiden Taktiken zu den Ausgangsvariablen. Der größte Mittelwertsunterschied der KG2 ($M = 21.15$) wird jedoch knapp nicht signifikant

($t(25) = 1.886$; $p = .071$) gegenüber dem der EG2 ($M = 13.57$).

Bei den wichtigen Einzeleingriffen fällt eine hohe Korrelation zwischen den Einzeleingriffen auf Variable I und den Einzeleingriffen auf Variable II auf. Sie beträgt $r = .654$ und ist hoch signifikant ($p < .001$). Das lässt sich wohl dahingehend interpretieren, dass Einzeleingriffe wenn überhaupt, dann meist bei beiden Variablen eingesetzt wurden, was für einen gezielten Einsatz dieser Taktik spricht.

Als starker Eingriff wird eine Eingabe dann gewertet, wenn ein Wert einer Eingabevariable über 200 gewählt wurde und gleichzeitig der andere Wert nicht 10 überschritt. Mit diesen Grenzen soll gewährleistet werden, dass bei einem solchen Eingriff die direkten Wirkzusammenhänge deutlich im System sichtbar werden. Dies ist bei diesem Abstand und der unteren Grenze von 200 im System WITS durchaus gegeben. Selbst bei diesen recht leicht zu erreichenden Kriterien kommen die starken Eingriffe selten vor und unterscheiden sich statistisch nicht bedeutsam in den Gruppen ($F(3,52) = 1.010$; $p = .397$). Angedacht waren auch andere Definitionen für starke Eingriffe. Die Möglichkeit, von allen Eingabewerten das obere Quartil als starke Eingriffe zu bezeichnen, bringt den Vorteil mit sich, dass es eine gewisse Mindestzahl an starken Eingriffen gibt. Dementsprechend lägen von den 2968 Eingaben 742 über der Quartilsgrenze, die im vorliegenden Experiment einen sehr hohen Wert von 20 000 erreichte. Die Eingabe solch eines hohen Wertes ist jedoch nur dann als starker Eingriff zu werten, wenn die andere Eingabevariable nicht ebenfalls so hoch gewählt wird. Die Auszählung der starken Eingriffe nach dem Kriterium eines Wertes von mindestens 20 000 und einem Größenverhältnis von 100 zu 1 für die zweite Eingabe erbrachte eine noch geringere Anzahl, so dass auch bei dem Quartils-Kriterium nicht von einer zufriedenstellenden Definition für einen starken Eingriff gesprochen werden kann.

Wie in diesem Abschnitt gezeigt, liegen aber mit der aggregierten Taktikvariable und den deutlichen Unterschieden in der Nutzung der Null-, Einzel- und All-Eingriffe gute Möglichkeiten vor, das taktische Verhalten der Versuchspersonen zu differenzieren. Welchen Einfluss dieses taktische Verhalten auf den Wissenserwerb hat, zeigen die Ergebnisse zu Hypothese II auf.

5.5 Ergebnisse zur Hypothese II

Wie sieht es nun aber mit dem Zusammenhang zwischen dem Nutzen der verschiedenen Taktiken und anderen Variablen wie dem Wissenserwerb und den Steuerungsleistungen aus? Aggregiert man die Anzahl der Null-, Einzel- und starken Eingriffe sowie die Anzahl der auf Null bzw. unter Null gesteuerten Ausgabevariablen, erhält man eine Variable mit der

Gesamtnutzung der Taktiken. Nachfolgend sollen diese Variablen sowohl in der aggregierten Form als auch als einzelne Variablen mit Wissensvariablen korreliert werden. Da nicht alle Variablen normalverteilt sind, wird die Korrelation nach der Formel von Spearman verwendet, um Artefakte zu vermeiden.

Korreliert man die aggregierte Taktikvariable mit den Wissensvariablen, so fällt zunächst auf, dass fünf der sechs Korrelationen in die vorhergesagte Richtung ausfallen. Nur die kleinste, nicht signifikante Korrelation mit den gemachten Fehlern bei der Wirkstärkenangabe sollte negativ und nicht positiv ($r_s = .071$) sein.

Tabelle 11: Rangkorrelation (r_s) nach Spearman für die Taktikvariablen mit den Strukturwissensvariablen

	Wissensangaben (Treffer)			Wissensangaben (Fehler)			
	Zusammenhang	Richtung	Stärke	Zusammenhang	Richtung	Stärke	
Taktiken aggregiert	r_s	.033	.084	.313*	-.241*	-.274*	.071
	p	.406	.275	.011	.041	.024	.306
Nulleingriffe	r_s	.007	-.015	.160	-.248*	-.242*	.112
	p	.479	.458	.126	.036	.041	.211
Einzeleingriffe	r_s	.008	.038	.284*	-.374**	-.385**	.160
	p	.477	.394	.020	.003	.002	.126
All-Eingriffe	r_s	.000	-.012	-.278*	.357**	.366**	-.156
	p	.500	.465	.022	.004	.004	.132
starken Eingriffe	r_s	-.098	-.024	.210	-.077	-.132	.104
	p	.242	.432	.065	.291	.173	.229
Ausgabevariablen nahe Null	r_s	.127	.130	.127	.032	.009	-.098
	p	.182	.177	.183	.411	.474	.243
Ausgabevariablen unter Null	r_s	.025	.049	.110	.027	.015	-.095
	p	.428	.363	.217	.424	.457	.249

** Die Korrelation ist auf dem Niveau von .01 (1-seitig) signifikant.

* Die Korrelation ist auf dem Niveau von .05 (1-seitig) signifikant.

Drei der fünf Korrelationen werden auf dem 5% α -Niveau signifikant. Am deutlichsten korrelieren die Taktiken mit dem Wissen über die Wirkstärken ($r_s = .313$). Wer also viele Taktiken nutzte, wusste auch gut über die Stärken der Verbindungen Bescheid. Um Wirkstärken zu kennen und korrekt angeben zu können, müssen Zusammenhang und Richtung der Vernetzung eigentlich bekannt sein, daher scheint es auf den ersten Blick paradox, dass zu diesen Wissenswerten ein geringerer Zusammenhang mit den Taktiken besteht (siehe Tabelle 11). Interessant ist daher, zu sehen, dass deutliche Korrelationen

aufzutreten, wenn bei eben diesen beiden Variablen die Anzahl der Fehler betrachtet wird. Dabei ist es so, dass wenn mehr Taktiken verwendet werden, es zu weniger falschen Angaben beim Wirkzusammenhang ($r_s = -.241$) und bei der Wirkrichtung ($r_s = -.274$) kommt.

Nutzer von Taktiken erlangen also zum einen tendenziell mehr Wissen und haben zum anderen weniger falsche Annahmen über die Wirkverbindungen. Welche Taktik sich dabei wie auswirkt, ist der Tabelle 11 zu entnehmen. Nachfolgend sollen nur noch einige bedeutende Zusammenhänge herausgestellt werden.

In der Literatur zum komplexen Problemlösen wird der Nulleingriff zur Systembeobachtung als wichtige Taktik zur Erkennung der Dynamiken in einem System betrachtet. In dieser Untersuchung scheint es keine überzufälligen Zusammenhänge zu geben. Dies mag daran liegen, dass der Nulleingriff sicherlich nicht nur informationssuchend eingesetzt wird, sondern auch von solchen Versuchspersonen, die überbelastet oder mit geringer Motivation am Experiment teilnehmen. Auch sie lassen Eingabetakte verstreichen, ohne dabei gezielt auf die Dynamiken des Systems zu achten.

Dafür wirkt sich der Einsatz von Einzeleingriffen wie erwartet aus. Dies zeigt sich in einer Korrelation zum Wissen über die Stärke ($r_s = .284$) auf der einen und den hohen negativen Korrelationen ($r_s = -.374$ bzw. $r_s = -.385$) dieser Taktik mit den Fehlern auf der anderen Seite. Deutlich wird der Zusammenhang auch, schaut man auf die Personen, die häufiger als andere keine Taktik einsetzten und zumeist gleichzeitig alle Eingabevariablen veränderten (All-Eingriff). Wer dies tat, erlangte nicht nur weniger Wissen über die Wirkstärken ($r_s = -.0278$), sondern dazu noch viel falsches Wissen, was die Korrelationen ($r_s = .357$ bzw. $r_s = .366$) mit den All-Eingriffen verdeutlichen.

Mit diesen Ergebnissen aus den Korrelationen kann der Zusammenhang von Taktiken und Wissen bestätigt werden. Damit liegen differenzierte Zusammenhangsmaße für den Einfluss von Taktiken auf Wissen vor.

5.6 Allgemeine Ergebnisse zur Steuerungsleistung

Aufgabe der Versuchspersonen war es, zu einem möglichst frühen Zeittakt im Zieldurchgang die Zielwerte für die Variablen A und B möglichst exakt zu erreichen und zu halten.

Zur Bewertung der Zielerreichung werden die Abstandsmessungen zu den Zielwerten ($A = 20$ und $B = 500$) durch Logarithmierung transformiert. Diese Transformation hat unter anderem den Vorteil, dass die Daten sich der Normalverteilung, die Voraussetzung für fast alle der nachfolgenden Inferenzstatistiken ist, annähern. Weitere Informationen zu der bei komplexen Problemlöseuntersuchungen üblichen Transformation finden sich bei Müller (1993; siehe

auch Vollmeyer, Burns, Holyoak, 1996, S. 85). Von ihm stammt auch das Gütemaß, in dem die Transformation steckt und welches für dynamische Problemlöseszenarien häufig verwendet wird. Nach Müller (1993) wird die *Güte der Systemsteuerung* (GdS) nach folgender Formel gebildet:

$$GdS = \frac{\sum_{i=1}^{n_y} \sum_{j=1}^{n_t} \ln|y_{ij} - z_i|}{n_y \cdot n_t} \quad (1)$$

wobei GdS = Güte der Systemsteuerung,

n_y = Zahl der Zielvariablen, (hier 2 für A und B)

n_t = Zahl der Takte, (hier 7 für den Durchgang
4)

y_{ij} = Wert der Zielvariablen i in Takt j und

z_i = Zielwert der Zielvariablen (hier 20 für A; 500 für B)

Bei $|y_{ij} - z_i| = 0$ würde nicht logarithmiert, da $\ln(0)$ nicht definiert ist. Die GdS-Werte gleich Null bedeuteten, dass keine Abweichungen zwischen Ausgabevariable und Zielvorgaben bestünde, d.h., das Ziel wäre dann perfekt erreicht. Je niedriger die GdS-Werte sind, desto besser ist also die Steuerleistung. Daher werden meist negative Korrelationen erwartet, wenn die GdS-Werte mit anderen Skalen in Zusammenhang gesetzt werden.

Berechnet man die Abweichung nur für den letzten Takt des Zieldurchgangs und teilt dementsprechend nur durch n_y , erhält man die Güte der Steuerleistung für den Endzustand, in den das System von den Versuchspersonen gebracht wurde. Dieser Wert wird in der vorliegenden Arbeit mitberechnet und mit GdS_{T7} bezeichnet. Sowohl GdS als auch GdS_{T7} sind für beide Zielvariablen getrennt für A bzw. B berechenbar.

Für eine erste Regressionsuntersuchung soll zunächst aber nur der einfache GdS-Wert als Kriterium aus verschiedenen Variablen vorhergesagt werden. Ein bereits recht gutes Modell ergibt sich aus den Prädiktoren der Noten für den Abschluß und für Mathematik, der Computererfahrung, der Bearbeitungszeit der vier Durchgänge ohne die Wissenstests und dem Wissenindex aus der Wissensdiagnostik (siehe Tabelle 12). Dieses Modell erreicht eine bemerkenswerte Varianzaufklärung von $R^2 = .445$, und selbst die aufgrund von geringer Stichprobengröße und der Prädiktorenanzahl korrigierte Varianzaufklärung ergibt $R^2 = .384$.

Tabelle 12: Regressionsmodell für die Güte der Systemsteuerung (GdS) mit und ohne die Note für Mathematik mit standardisierten und nicht standardisierten Koeffizienten und T-Werten. Achtung: Für GdS und die Noten gilt, je kleiner der Wert, desto besser die Leistung.

Mo dell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	Signifikanz	
		<i>b</i>	Standardfehler	β	<i>t</i>	<i>p</i>
1	Note - Abschluß	1.541	.535	.421	2.880	.006
	Computererfahrung	-1.097	.346	-.390	-3.174	.003
	Zeit der Durchgänge ohne Wissenstest	-0.089	.036	-.314	-2.458	.018
	Wissensindex Gesamt	-0.111	.057	-.276	-1.965	.055
2	Note - Mathematik	-1.096	.324	-.505	-3.382	.001
	Note - Abschluß	0.409	.461	.112	0.887	.379
	Computererfahrung	-0.859	.374	-.305	-2.297	.026
	Zeit der Durchgänge ohne Wissenstest	-0.095	.040	-.335	-2.375	.022
	Wissensindex Gesamt	-0.057	.060	-.142	-0.953	.346

Verwunderlich ist das Vorzeichen des β -Koeffizienten für den Prädiktor Note-Mathematik. Die Erwartung, das numerische System sei von guten Schülern in Mathematik besser zu steuern, scheint sich hier umzukehren. Zur weiteren Analyse wurde daher ein zweites Modell ohne die Note für Mathematik berechnet. Aus den Differenzen der R^2 -Werte läßt sich auf eine Varianzaufklärung von 13,8% (13,6% für korrigierte Modelle) schließen, die das Modell durch diesen einzelnen Prädiktor erhält.

Tabelle 13: Zusammenfassung der Regressionsmodelle für GdS.

Modell	<i>R</i>	<i>R</i> ²	Korrigiertes <i>R</i> ²	Standardfehler des Schätzers
1	,667	,445	,384	1.92
2	,554	,307	,248	2.13

Für beide Modelle sind in Tabelle 13 die Kennwerte für multiple Korrelation und Varianzaufklärung aufgeführt. Den T-Werten der Tabelle 12 ist des weiteren für das Modell 2 zu entnehmen, dass ohne die Zugabe der Mathematiknote die Prädiktoren Note- Abschluss und Wissensindex entfallen würden, da ihr Beitrag zur Varianzaufklärung einzeln für das Aufnahmekriterium einer schrittweisen Regression nicht ausreichen würde. Es bliebe dann ein einfaches Modell aus der Bearbeitungszeit der vier Durchgänge und der Computererfahrung mit einer Varianzaufklärung von $R^2 = .278$ bzw. korrigiert $R^2 = .248$. Die Mathematiknote liefert damit einen zunächst verwirrenden, aber dennoch wichtigen Beitrag zur Varianzaufklärung des Modells. Welche Rolle sie genau spielt, ist hier nicht eindeutig zu klären und wird noch einmal Thema im Diskussionsteil werden.

Nachfolgend soll aber in den Ergebnissen zu Hypothese III geklärt werden, welchen Zusammenhang der Wissenserwerb mit der Steuerungsleistung hat.

5.7 Ergebnisse zur Hypothese III

Erwartet wurde eine negative Korrelation zwischen den abhängigen Variablen des Wissenserwerbs und der Steuerungsleistung im Zieldurchgang (GdS), da für die Variablen des Wissenserwerbs gilt, je größer der Wert, desto mehr Wissen liegt vor, und da der GdS-Wert ein Abstandsmaß ist, also je kleiner, desto besser.

Bei den in Tabelle 14 vorliegenden Korrelationen handelt es sich um Produkt-Moment-Korrelationen nach Pearson, die hier verwendet werden können, da fast alle Variablen in der Tabelle normalverteilt sind. Die beiden Ausnahmen sind die Gütemaße für den letzten Takt. Diese verfehlen nur knapp die Zielkriterien im Kolmogorov-Smirnov-Anpassungstest. Die Variablennamen dieser beiden Gütemaße sind in der Tabelle kursiv dargestellt und ihre Korrelationen etwas mit Vorsicht zu betrachten. An dieser Stelle sei jedoch auf Bortz (1993, S. 198) hingewiesen. Er weist wiederum auf die Arbeiten von Havlicek und Peterson (1977) hin, die zeigen, dass der Signifikanztest für Korrelationskoeffizienten äußerst robust sowohl gegenüber Verletzungen der Verteilungsannahme als auch gegenüber Verletzungen des vorausgesetzten Intervallskalenniveaus ist.

Tabelle 14: Produkt-Moment-Korrelation nach Pearson für den Zusammenhang zwischen den Wissensvariablen nach Durchgang 3 und den Zielvariablen mit den logarithmierten Abweichungen angegeben als Güte der Steuerungsleistung (GdS) nach Müller (1993)

	Wissensvariable (Treffer)				Wissensvariable (Fehler)		
	Zusammenhang	Richtung	Gesamt		Zusammenhang	Richtung	Gesamt
GdS	<i>r</i>	-.172	-.270*	-.289*	.241*	.303*	.243*
	<i>p</i>	.110	.025	.018	.041	.014	.040
GdS für A	<i>r</i>	-.164	-.266*	-.278*	.229*	.294*	.226
	<i>p</i>	.120	.027	.022	.050	.016	.052
GdS für B	<i>r</i>	-.172	-.263*	-.288*	.245*	.300*	.249*
	<i>p</i>	.109	.028	.018	.039	.015	.036
<i>GdS_{T7}</i>	<i>r</i>	-.203	-.261*	-.277*	.145	.172	.133
	<i>p</i>	.072	.030	.022	.150	.109	.171
<i>GdS_{T7} für A</i>	<i>r</i>	-.221	-.297*	-.309*	.121	.166	.113
	<i>p</i>	.056	.015	.012	.194	.118	.209
GdS _{T7} für B	<i>r</i>	-.172	-.206	-.227	.161	.169	.145
	<i>p</i>	.109	.069	.051	.125	.113	.150

* Die Korrelation ist auf dem Niveau von 0.05 (1-seitig) signifikant.

Neben den aggregierten Wissensvariablen (Gesamt für Treffer bzw. Fehler) sind für den interessierten Leser zusätzlich noch die Zusammenhänge für die Wirkzusammenhänge und die Wirkrichtung aufgeführt. Die Wirkstärke wurde hier wegen der allzu großen Verletzungen der Normalverteilung bewusst weggelassen. Sie fließt aber in die beiden Gesamtvariablen mit ein.

Beide Gesamtvariablen, sowohl die der Treffer ($r = -.289$; $p = .018$) als auch die der Fehler ($r = .243$; $p = .040$), weisen eine deutliche Korrelation mit der Güte der Systemsteuerung (GdS) auf. Des weiteren fällt auf, dass alle Korrelationen der Tabelle 14 in die vorhergesagte Richtung ausfallen. Wenn auch nicht alle statistische Bedeutsamkeit erlangen, kann dies doch als deutlicher Hinweis auf die Gültigkeit der Hypothese gewertet werden.

5.8 Analyse der Zeiten

Dieser Abschnitt widmet sich der Analyse der Bearbeitungsdauer einzelner Teile der Untersuchung, da bei einem Experiment von durchschnittlich mehr als einer halben Stunde Dauer und einer Standardabweichung von fast 15 Minuten mit großen Unterschieden in der Bearbeitungszeit zu rechnen ist. Sollten diese Unterschiede systematisch sein, könnte dies auf Effekte wie z.B. den einfacheren Zugang zum Problem durch das Treatment hinweisen. Zudem soll hier auch ein zusätzlicher Überblick über die Phasen des Experimentes (vgl. Seite 56) durch die Zeitangaben gegeben werden.

Tabelle 15: Mittelwerte und Standardabweichung der Bearbeitungsdauer (in Minuten) der vier Untersuchungsgruppen.

	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Minimum</i>	<i>Maximum</i>
KG1	12	35.88	9.11	19.43	47.95
KG2	13	39.66	18.64	16.05	72.93
EG1	14	31.02	13.84	10.83	56.23
EG2	14	29.25	13.47	9.32	53.92
Gesamt	53	33.77	14.43	9.32	72.93

Betrachtet man die Zeiten der vier Treatment-Gruppen aus Tabelle 15 genauer, so zeigt sich, dass die taktische Kontrollgruppe (KG2) am meisten Zeit benötigte, gefolgt von der unspezifischen Kontrollgruppe (KG1), der Experimentalgruppe mit einer (EG1) und der mit drei (EG2) Analogien. Graphisch veranschaulicht zeigt sich stets ein gleiches Muster ähnlich einer Bergspitze (siehe Abbildung 17 und 10). Dieses Muster zeigt sich genauso auch für jeden einzelnen Durchgang. Die Mittelwertsunterschiede werden jedoch nicht signifikant,

außer in Durchgang zwei ($F(3,53) = 3.504$; $p = 0.022$), in dem der Unterschied zwischen KG2 ($M = 5.82$; $SD = 5.40$) und EG2 ($M = 2.11$; $SD = 1.26$) signifikant wird. Da der größte Teil der Zeit mit den Durchgängen verbracht wurde (42.4%) – im Gegensatz zu den Wissenstests (32.8%) und den abschließenden Bemerkungen bzw. dem soziodemographischen Fragebogen zu Anfang (24.8%) –, ergibt sich auch für den gesamten Programmablauf ein vergleichbares Muster (siehe Abbildung 17).

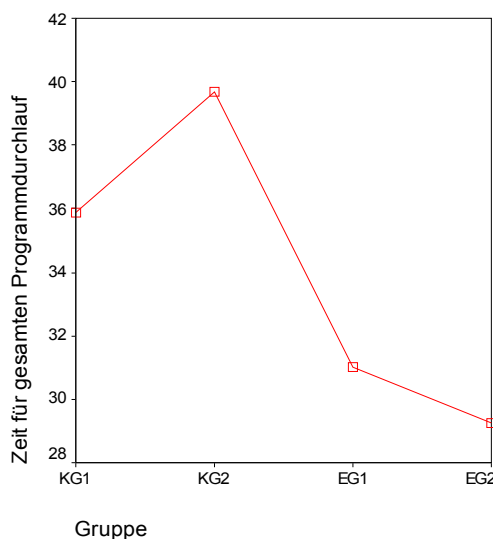


Abbildung 17: Gemittelte Zeiten der Gruppen für das gesamte Experiment in Minuten.

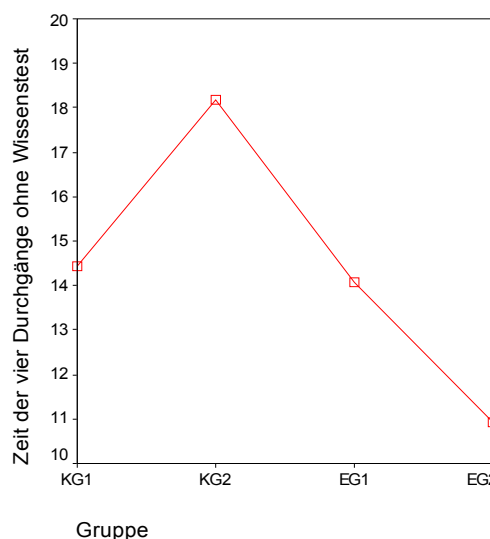


Abbildung 18: Gemittelte Zeiten der Gruppen für die vier Durchgänge in Minuten.

Die Mittelwerte und Standardabweichungen der Zeiten für die einzelnen Phasen der Untersuchung sind der Tabelle 16 zu entnehmen. Sie entsprechen weitgehend den Erwartungen, dass der erste Durchgang sowie der erste Wissenstest aufgrund der Zeit zum Lesen der Instruktionstexte am längsten dauern. Die Bearbeitungszeiten nehmen dann von Durchgang zu Durchgang ab. Im Zieldurchgang erhöht sich erwartungsgemäß die Bearbeitungszeit etwas.

Tabelle 16: Mittelwerte und Standardabweichung der Zeiten einzelner Phasen der Untersuchung.

<i>Phase</i>	<i>Zeit in min.</i>	<i>Phase</i>	<i>Zeit in min.</i>
Durchgang 1	$M=4,19$ $SD=2,09$	Wissenstest 1	$M=5,46$ $SD=3,99$
Durchgang 2	$M=3,56$ $SD=3,29$	Wissenstest 2	$M=1,97$ $SD=1,96$
Durchgang 3	$M=2,78$ $SD=1,95$	Wissenstest 3	$M=1,73$ $SD=2,33$
Durchgang 4 (Ziel)	$M=3,81$ $SD=4,26$	Wissenstest 4	$M=1,00$ $SD=1,49$
Gesamte Durchgänge	$M=14,33$ $SD=8,68$	Gesamte Wissenstests	$M=11,06$ $SD=6,91$
Fragebogen & Kommentar	$M=8,38$ $SD=3,37$		
Gesamter Programmablauf	$M=33,8$ $SD=14,4$		

Die Tatsache, dass die Versuchspersonen der Experimentalgruppen weniger Zeit brauchen als die der Kontrollgruppen, könnte auf einen leichteren Umgang mit dem komplexen, dynamischen Problem hindeuten. Demnach wäre es leichter, das Problem zu bearbeiten, wenn es mit einem bekannten analogen Bereich verglichen werden kann. Kontraintuitiv bleibt aber, dass die EG2 weniger Zeit braucht als die EG1, obwohl deren Versuchspersonen mehr Texte zu lesen haben und für die zweite und dritte Analogie neue Überlegungen anstellen müssten. Wie bereits erwähnt, ist aber selbst bei den größeren Abständen zwischen der Bearbeitungszeit der KG2 zur EG2 der Mittelwertsunterschied nicht signifikant, so dass auch beim Abstand von EG1 zu EG2 von einem zufälligen ausgegangen werden muss.

6. Diskussion

Nachfolgend sollen die Ergebnisse aller Hypothesen diskutiert und soweit möglich in den Forschungskontext eingeordnet werden (siehe Abschnitt 6.1). Da es bisher noch keine Vergleichsstudien von DYNAMIS-Experimenten im Internet gibt, werden mit Abschnitt 6.2 ausführlich die Vor- und Nachteile von Online-Erhebungen diskutiert und auf das vorliegende Experiment bezogen. Dabei wird auch auf die große Zahl der Abbrecher noch einmal eingegangen. Im Anschluss an den Abschnitt über Internetuntersuchungen soll in Abschnitt 6.3 das neue Online-DYNAMIS-System als Untersuchungsinstrument kritisch betrachtet und seine Tauglichkeit geprüft werden.

Viele Experimente mit komplexen dynamischen Szenarien zeigen deutliche Geschlechterunterschiede in der Bearbeitung auf. Abschnitt 6.4 diskutiert deshalb kurz diese Unterschiede, die in der vorliegenden Untersuchung doch eher gering ausfallen. Abschnitt 6.5 widmet sich dann dem nicht ganz leicht zu interpretierenden Effekt der Intelligenz und der mathematischen Fertigkeiten auf den Wissenserwerb und die Steuerungsleistung.

6.1 Die Hypothesen

6.1.1 Hypothese I – Die Analogien

Drei Dinge sollten sich in dieser Hypothese zeigen: Zum ersten wurde erwartet, dass die Experimentalgruppen mehr Strukturwissen erwerben als die Kontrollgruppen. Diese Erwartung wurde für die Vermeidung von Fehlern beim Strukturwissen nur knapp nicht signifikant ($F(1,52) = 3.063$; $p = .086$). Es zeigte sich aber ein deutlicher Trend in den Mittelwerten der Experimentalgruppen ($M = 4.86$) gegenüber denen der Kontrollgruppen ($M = 6.96$).

Zum zweiten sollte sich das Strukturwissen vor allem dann zeigen, als die Experimentalgruppen gegen die unspezifische Kontrollgruppe (KG1) kontrastiert wurden, da durch die Taktikvorgabe auch bei Kontrollgruppe KG2 einer größerer Strukturwissenserwerb zu erwarten war. Für den aggregierten Wissensindex (aus Treffer - Fehler) wurde der Test für geplante Kontraste für den Fall von nicht gleicher Varianz eingesetzt, und es zeigte sich tatsächlich ein signifikanter Mittelwertsunterschied ($t(26.396) = 2.230$; $p = .034$).

Zum dritten sollte geklärt werden, ob der Wissenserwerb der Experimentalgruppen durch den Gebrauch von sinnvollen Taktiken modifiziert wurde oder ob ein Anteil der Varianz sich auf das Treatment der Analogien zurückführen lassen konnte. Dazu wurden die Varianzanteile berücksichtigt, die durch Hinzunahme der Gruppenzugehörigkeit in ein Regressionsmodell

aufgeklärt werden konnten. Im Falle des Strukturwissens ergab sich ein geringer Effekt von 1,6% Varianzaufklärung für die Treffer, ein recht starker aber mit 8% Varianzaufklärung für die Fehler (für die genauere Beschreibung des Vorgehens siehe Abschnitt 'Ergebnisse zur Hypothese I' auf Seite 77).

Neben diesen dreien stellt sich natürlich auch noch die Frage nach den allgemeinen Erleichterungen durch die Analogien. Es zeigt sich im Gegensatz zu den expliziten Taktiken, die der KG2 präsentiert wurden, dass die Präsentation einer Analogie die Bearbeitungszeit zu senken scheint, wie der Abschnitt 'Analyse der Zeiten' auf Seite 88 darstellt. Da die Güte der Systemsteuerung (GdS) auch von der Bearbeitungszeit abhängt, wäre ein Nachweis wünschenswert, dass die Gruppen mit Analogien trotz geringerer Zeit zu besseren Ergebnissen kommen. Der Effekt der Gruppen auf die Steuerungsleistung (GdS) sollte sich daher bei herauspartialisierter Zeit erhöhen. Dies zeigt sich aber in den vorliegenden Daten nicht.

Tabelle 17: Mittelwert, Standardabweichung der Zielabweichungen (GdS) für die verschiedenen Treatmentgruppen.

	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Minimum</i>	<i>Maximum</i>
KG1	12	6.43	2.20	4.78	12.95
KG2	13	5.72	2.04	0.96	9.22
EG1	14	5.80	2.58	-0.37	10.26
EG2	14	6.73	2.90	4.18	15.27
Gesamt	53	6.17	2.44	-0.37	15.27

Trotz vieler gezeigter Unterschiede der Gruppen und guter Korrelationskoeffizienten zum Strukturwissen unterscheiden sich die Mittelwerte der Steuerungsleistung statistisch nicht bedeutsam ($F(3,52) = .527$; $p = .666$) (siehe Tabelle 17). Genau bei der Steuerungsleistung wäre jedoch eine Auswirkung der Analogien ebenfalls wünschenswert gewesen.

Wodurch sich nun die tendenzielle Steigerung im Wissenserwerb für die Experimentalgruppen ergibt, kann letztlich nicht ganz aufgeklärt werden.

So haben Studien von Halpern, Hanson und Riefer (1990) gezeigt, dass beim Verstehen und Lernen von Konzepten Analogien aus entfernteren Bereichen (fremden Domänen) zu besseren Verständnis- und Behaltensleistungen führen als Analogien aus gleichen Domänen. Dies wird mit der aufgebrauchten kognitiven Arbeit erklärt, die höher ist, wenn bereichsfremde Analogien herangezogen werden. Mit anderen Worten kann sich der Effekt der Analogien nur zeigen, wenn mit ihnen intensiv gearbeitet wird. Werden sie nicht betrachtet oder wird nur oberflächlich mit ihnen gearbeitet, dann können sie als Treatment nicht wirken.

Daraus ergibt sich, dass es für die Beobachtung von Effekte des Treatments nötig ist, zunächst auch die Nutzung der vorgegebenen Strategien während des Experiments zu prüfen. In der aktuellen Untersuchung erwies sich der Treatment-Check als wenig aussagekräftig, was daran lag, dass die Anzahl der Nachfragen nach den Instruktionstexten, wie sie in Tabelle 7 und Abbildung 13 auf Seite 75 beschrieben sind, kaum einen Hinweis geben, ob tatsächlich intensiv mit den Analogien gearbeitet wurde oder nicht. Zwar war ein weiterer Treatment-Check geplant, der die zum Lesen jener Texte benötigte Zeit hätte vergleichen können, dieser konnte aber aufgrund eines Programmfehlers⁷ nicht sinnvoll genutzt werden.

Das einigermaßen häufige Wechseln der Variablenbezeichnungen spricht zwar für eine Beschäftigung mit den Analogien, jedoch zeigt sich schnell auch die deutliche Zusatzbelastung vor allem in der Experimentalgruppe mit wechselnden Analogien. So nutzten nur wenige Versuchspersonen die Variablenbezeichnungen für einen Wechsel der Analogiestruktur auch nach dem ersten Takt noch. In der EG2 sind es beispielsweise vier von vierzehn Personen, die nur im ersten Takt Umstellungen vornahmen. Hinzu kommen weitere sechs Personen, die nach dem ersten Takt nur zwei oder weniger weitere Wechsel der Bezeichner vornahmen. Ein ähnliches Bild ergibt sich im ersten Durchgang für die EG1. Dort gab es zwei Versuchspersonen, die die Bezeichner gar nicht wechselten, und weitere vier, die dies nur im ersten Takt taten.

Die Studien von Halpern, Hanson und Riefer (1990) lassen einen Zusammenhang von intensive Nutzung der Analogien und dem Effekt auf den Wissenserwerb erwarten. Dieser Zusammenhang zwischen den Vielnutzern der Variablenbezeichnungen und besserem Wissenserwerb oder besserer Steuerungsleistung konnte jedoch nicht bestätigt werden. Es bleibt die Vermutung, dass sich die Versuchspersonen der Experimentalgruppen mit den Analogien im Sinne des Treatments auseinandersetzen, was sich aber nicht in einem starken Gebrauch der Drop-Down-Menüs auswirkte, sondern durch die Beschäftigung mit den

⁷ Statt die Dauer der Texte am Bildschirm zu addieren, überschrieb die letzte genommene Zeit jeweils die vorhergehenden.

Variablenbezeichnungen und daraus folgenden gedanklichen Strukturen und Beispielen. Möglich wäre die Auswirkung des Treatments auf den Wissenserwerb, da es beim Umgang mit komplexen Problemen weniger auf das langzeitliche Lernen ankommt, wie in den Studien von Halpern et al. (1990), sondern zunächst ein hypothesengenerierendes mentales Modell aufgebaut werden muss, das es dann zu überprüfen gilt. Dörner (1992, S. 114f.) schreibt, dass sich bereits durch das Lesen und Nachdenken über die Quelldomäne neue Hypothesen- und Experimentalräume für die Versuchspersonen aufbauen können. Davon dürften vor allem die nach Klahr und Dunbar (1988) als „Theoretiker“ bezeichneten Problemlöser betroffen sein, da sich durch die Analogien für sie prüfbare Strukturen ergeben. Die „Experimentierer“ hingegen müssen zunächst die Strukturen per Test entdecken, bevor sie ihnen dann auffallen (vgl. Klahr & Dunbar, 1988, p. 27) und damit zum Wissenserwerb dienen können.

Um den Effekt der Analogien besser aufklären zu können, ergeben sich aus dem bisher Gesagten zwei ableitbare Punkte. Zum einen muss eine differenziertere Treatmentkontrolle Verwendung finden, zum anderen sollte das Treatment in seiner Wirkung verstärkt werden. Für die Treatmentkontrolle müsste neben der Dauer, also wie lange sich die Versuchspersonen mit den Texten und damit auch mit den Analogien aufhalten, auch die Anzahl der Mausklicks auf die Drop-Down-Menüs in der Logdatei festgehalten werden. Der Wechsel der Variablenbezeichnungen von Takt zu Takt allein reicht nicht aus.

Möglichkeiten, den Effekt des Treatments zu erhöhen, ergäben sich durch die Einführung einer Proberunde, um das Verständnis der Nutzung von Analogien zu prüfen bzw. es von Beginn an auf das gleiche Niveau zu setzen. Dadurch würde auch die Nutzung der Analogien sichergestellt werden können. Weitere Möglichkeiten bestünden in der Verbesserung der Darstellung und Vermittlung der Instruktionen (siehe dazu Abschnitt 'Untersuchungsinstrument DYNAMIS' auf Seite 98).

6.1.2 Hypothese II – Die Taktiken

Vorhergesagt war ein deutlicher Zusammenhang zwischen dem Nutzen von sinnvollen Taktiken und dem Wissen, das sich Versuchspersonen über das System aneignen. Die Korrelationen für diesen Zusammenhang finden sich in Tabelle 11 auf Seite 83 und zeigen den positiven Zusammenhang deutlich. Damit werden vergleichbare Untersuchungen wie die von Vollmeyer, Burns und Holyoak (1996) bestätigt, die diesen Zusammenhang bereits postulierten.

Vollmeyer et al. (1996, p. 86) nutzten in einem vier Eingabe- auf vier Ausgabevariablen

DYNAMIS-Szenario einen „systematicity score“, in den für jeden Takt mit einem Einzeleingriff 2 Punkte, für All-Eingriffe 0 und alle anderen Eingriffe 1 Punkt einfließen. Dieser Wert korrelierte zu $r = .76$ mit dem Strukturwissen, welches durch Ausfüllen eines Kausaldiagramms diagnostiziert wurde und zu $r = .43$ mit dem Wissen aus Paarvergleichsfragen (siehe 'Paarvergleich' bzw. 'Kausaldiagramme' ab Seite 23).

Wenn auch die Korrelationen zum Zusammenhang von Taktiken und Steuerleistung in dieser Untersuchung nicht so hoch ausfallen wie die von Vollmeyer et al., sondern maximal bei .313 bzw. -.385 liegen, so ergeben sie doch ein differenzierteres Bild. So ist z.B. auffällig, dass die Wirkstärke hohe Korrelationen mit den Taktiken nur bei den korrekten Wissensangaben aufweist und sich bei den Fehlern in der Wissensdiagnostik ein fast umgekehrtes Bild ergibt. Bei den Fehlern der Wissensdiagnostik korrelieren vor allem das Wissen über den Wirkzusammenhang und die Wirkrichtung mit den Taktiken. Diejenigen, die Taktiken intensiv nutzen, machen also weniger falsche Angaben in der Wissensdiagnostik für Zusammenhang und Richtung und geben mehr korrekte Antworten bei der Wirkstärke.

Ob dieser Zusammenhang häufiger so auftritt oder ein Artefakt dieser einen Untersuchung ist, müsste in Folgeexperimenten geklärt werden. Dann wäre auch mittels Regression zu klären, welche Variable zu welcher Korrelation ihre Varianz beiträgt, um ein noch differenziertes Bild für die drei Wissenskomponenten Zusammenhang, Richtung und Stärke zeichnen zu können.

6.1.3 Hypothese III – Der Wissenserwerb

Vorhergesagt war eine deutliche Korrelation zwischen den Variablen des Strukturwissens und der Steuerungsleistung (GdS). Diese Vorhersage konnte bestätigt werden. Es ergaben sich Korrelationen zwischen der Steuerleistung und dem Strukturwissen (Treffer) von $r = -.289$ ($p = .018$) und dem Strukturwissen (Fehler) von $r = -.243$ ($p = .040$) (siehe Abschnitt 'Ergebnisse zur Hypothese III' Seite 87).

Die Frage nach dem Zusammenhang zwischen Wissen und umsetzbarer Steuerungsleistung stellt sich, seit Berry und Broadbent (1984) Ergebnisse zu Tage förderten, in denen das Wissen, welches über das System vermittelt wurde, nicht oder sogar negativ wirkte. Wie im Theorieteil (unter 'Wissensanwendung' auf Seite 21) dargestellt, konnten verschiedene neuere Untersuchungen jedoch den positiven Zusammenhang stärken und auf Ursachen für Berry und Broadbents Ergebnisse hinweisen.

Bei Vollmeyer et al. (1996) korreliert der Strukturwissenswert, der sich aus der Vervollständigung eines Kausaldiagramms ergab, zu $r = -.57$ ($n = 60$) mit dem Abstand der

Steuerungsleistung für den Zieldurchgang und sogar zu $r = -.65$ ($n = 60$) für den Transferdurchgang, der mit dem Zieldurchgang des Experiments für diese Diplomarbeit vergleichbar ist. Ebenso zeigt sich bei Preußler (1998) im Transferdurchgang eine Korrelation von $r = -.42$ ($n = 27$) für GdS und Strukturwissen. Bei ihr zeigen sich in einem zweiten Experiment auch noch etwas geringere Korrelationen für zwei Untergruppen von $r = -.39$ ($n = 26$) bzw. $r = -.30$ ($n = 27$).

Die Korrelationen in dieser Untersuchung, z.B. zwischen GdS und dem Strukturwissen (der Treffer) mit $r = -.29$, liegen etwas niedriger als die von Vollmeyer et al. und Preußler berichteten. Dennoch können sie als Bestätigung des Zusammenhangs angesehen werden. Dass die Korrelationen etwas geringer ausfallen, könnte auf einen eingeschränkten Varianzbereich hinweisen. Dies erscheint in Anbetracht der Schwierigkeit des Systems und der hohen Abbrecherquote durchaus im Bereich des Möglichen. Zudem ist auch die Bearbeitung des Tests der Wissensdiagnostik als nicht sehr intuitiv einzuschätzen, was eine weitere Varianzreduktion mit sich bringen dürfte. Dass der Wissenstest von mathematischen Fähigkeiten abhängen könnte, wird innerhalb des folgenden Abschnitts diskutiert.

Es soll aber noch einmal der Befund herausgehoben werden, dass alle Korrelationen der Tabelle 14 auf Seite 87 in die erwartete Richtung weisen und wenn auch nicht alle signifikant, dann doch immerhin mit relativ geringen α -Fehler-Wahrscheinlichkeiten. Darin zeigt sich, dass sich die erwünschten Effekte im Online-DYNAMIS-Programm nachweisen lassen. Bei den Versuchspersonen erfolgt ein erwünschter Wissenserwerb, und das Online-DYNAMIS-System ermöglicht eine Diagnose des Wissens und der Steuerleistung.

6.2 Die Internetuntersuchungen und ihre Abbrecher

Das am häufigsten vorgebrachte Argument gegen Online-Experimente im Internet ist die mangelnde Repräsentativität der verfügbaren Stichproben, da nicht alle bundesdeutschen Haushalte an das Internet angeschlossen sind (Bandilla, 1999; Hauptmanns, 1999; Janetzko, 1999; Reips, 2000). In der Natur des Mediums Internet liegt es zudem, dass die Stichproben selbstselektierend sind, was zu einer Reihe von Verzerrungen führen kann. Allein die Wahrscheinlichkeit, den Hinweis auf ein Web-Experiment zu finden, hängt stark von der Intensität ab, mit der man das Internet nutzt (Bandilla, 1999). In der vorliegenden Untersuchung wurde zudem dadurch eine Selektion erzeugt, dass die meisten Teilnehmer über Email-Kettenbriefe erreicht oder per Email-Verteiler benachrichtigt wurden. Dazu müssen die Empfänger zunächst aber bei den entsprechenden Diensten eingetragen sein oder in einer Adressliste eines Empfängers, der den Kettenbrief bereits erhalten hat. Als weiteres Argument

wird vorgebracht, dass keine Informationen über die Personen vorliegen, die über die Untersuchung zwar informiert waren, sich jedoch gegen eine Teilnahme entschieden (Hauptmanns, 1999; Reips, 2000). Dieses Problem besteht aber bei Laborexperimenten an Universitäten ebenso wie bei Internetexperimenten.

Wie im Abschnitt 'Beschreibung der Stichprobe' auf Seite 69 gab es auch in dieser Untersuchung eine große Zahl von Abbrechern, wie sie für Internetexperimente im Vergleich zu Laborexperimenten nicht ungewöhnlich ist (Reips, 2000). Typisch ist laut Reips (2000) auch, dass viele mit einem Experiment beginnen, dieses aber aufgrund der Freiwilligkeit auch wieder abbrechen.

Diese Freiwilligkeit, die im Gegensatz zur Pseudo-Freiwilligkeit bei der Ableistung von Versuchspersonenstunden an psychologischen Instituten steht (Reips, 1999), bringt aber auch Vorteile mit sich. Ein Beispiel ist die hohe Motivation, die für gewöhnlich zu einer geringeren Fehlervarianz beiträgt (Reips, 2000). Da im vorliegenden Experiment dennoch eine recht hohe nicht erklärbare Varianz in der Steuerleistung vorliegt, wird vermutet, dass das Experiment trotz Pilotstudie und darauf erfolgter Vereinfachung immer noch sehr schwer war. Dies würde auch einen Teil der Abbrecherrate erklären, die trotz Internetexperiment immer noch recht hoch erscheint (nur 17,5% absolvierten das Experiment bis zum Ende).

Reips (2000) zeigt auf, dass hohe Abbrecherraten in Internetexperimenten auf Mängel der Untersuchungsmaterialien hinweisen können. Eine genaue Evaluierung der Online-DYNAMIS-Version wäre daher notwendig und sinnvoll. Dabei sollte getestet werden, bis zu welchem Grad an Komplexität und Schwierigkeit ein Experiment mit einer annehmbaren Abbrecherrate durchgeführt werden kann.

Trotz mehrfachen Gegenlesens und Verbesserung der Verständlichkeit könnte ein Mangel auch in den Instruktionstexten liegen. Da keine Rückfragen der Teilnehmenden an den Versuchsleiter möglich sind, könnte dies zu Verständnisproblemen führen und stellt daher besonders hohe Anforderungen an das Untersuchungsmaterial. Dass keine Rückfragen möglich sind, erweist sich auf der anderen Seite wiederum als Vorteil, da keine Versuchsleitereffekte auftreten können (Bandilla, 1999).

Ein deutlicher Vorteil der Rekrutierung von Versuchspersonen aus dem Internet gegenüber denen von zumeist psychologischen Instituten ist neben der Quantität auch die Qualität der erreichbaren Stichprobe. Über das Internet sind ganz andere soziale Gruppen zu erreichen (Janetzko, 1999), und damit erhöht sich auch die Heterogenität der Teilnehmenden.

In der vorliegenden Untersuchung zeigt sich der von Janetzko beschriebene Vorteil allerdings nur sehr begrenzt. Von den $n = 53$ Versuchspersonen waren fast 68% Studierende. Scheinbar

brauchte es doch ein gewisses Commitment dem Versuchsleiter oder zumindest dem psychologischen Institut Heidelberg gegenüber, um das Experiment bis zum Ende durchzuarbeiten. Umgekehrt kann allerdings auch argumentiert werden, dass immerhin über 30% der Teilnehmenden keine Studierenden waren, was eine recht hohe Prozentzahl im Vergleich zu lokal durchgeführten Experimenten darstellt.

Der Einsatz eines Internetexperimentes wie der Online-DYNAMIS-Version bringt auch den weiteren Vorteil einer von vielen geforderten Erhöhung der ökologischen Validität psychologischer Experimente (Holzkamp, 1972) mit sich. Diese könnte das bestehende Problem der nicht vollständigen Standardisierung wettmachen. In psychologischen Experimenten ist diese Standardisierung im Gegensatz zu psychologischen Untersuchungen ohnehin nicht zwingend notwendig, wie Westermann (1987) feststellte.

Krantz und Dalal (2000) geben als einen weiteren Nachteil von Internetexperimenten die Gefahr der bewussten Ergebnismanipulation durch Teilnehmer an. Sie kann auf verschiedene Weisen geschehen, etwa durch bewusste Verfälschungen oder Mehrfachteilnahme. Reips (2000) sieht diese Gefahr allerdings nicht und hält die multiple Teilnahme von Versuchspersonen für selten und unwahrscheinlich (siehe auch Reips, 1999, für empirische Evidenz dazu). Dennoch müssen sie für das Untersuchungsverfahren DYNAMIS aus besonderen Gründen diskutiert werden, was im folgenden Abschnitt getan wird.

Abschließend soll aber zum Thema der Internetexperimente gesagt werden, dass sich insgesamt betrachtet eine Menge von Argumenten für die Durchführung von Online-Experimenten finden, die in vielen Fällen die Nachteile überwiegen (Reips, 1997).

6.3 Untersuchungsinstrument DYNAMIS

Der Kritikpunkt der Mehrfachteilnahme von Versuchsteilnehmern bei Internetexperimenten ist für das Online-DYNAMIS-System wichtig, schließlich erhalten die Versuchspersonen erst im letzten Durchgang die Ziele, die es zu erreichen gilt. Ein Mehrfachteilnehmer, der die Ziele schon kennt und dadurch die Explorationsphase als Übungsphase verwenden könnte, unterläge ganz anderen Versuchsbedingungen und würde Daten produzieren, die nicht mit den Daten eines Erstteilnehmers vergleichbar wären. So wäre er nicht nur besser in der Zielerreichung, sondern vermutlich auch schlechter beim Wissenserwerb (vgl. Vollmeyer et al. 1996 zur goal-specificity).

Wenn es besondere Anreize für eine Teilnahme gibt, wie z.B. das Gewinnspiel in dieser Untersuchung, sind Mehrfachteilnahmen sogar zu erwarten. Im vorliegenden Experiment sollte daher eine Frage nach der Erstteilnahme beantwortet werden. Da es beim Gewinnspiel

nicht um Leistungswerte, sondern um die einfache aber vollständige Teilnahme ging, ist nicht von absichtlicher Verfälschung der Frage nach der Erstteilnahme auszugehen. Der Anreiz für eine Teilnahme ist mit Bedacht auszuwählen, will man keine Verfälscher unnötig in Versuchung führen. Dennoch könnte gerade eine Änderung des Gewinnspiels in eine Preisvergabe an die besten Teilnehmenden die Motivation, bis zum Ende gute Leistungen bringen zu wollen, erhöhen.

Auch nicht monetäre Anreize sind für reine Internetexperimente ohne Versuchsleiter besonders wichtig und für zukünftige DYNAMIS-Einsätze zu beachten. So würde schon eine ansprechendere Graphik den „Spielspaß“ des Experimentes erhöhen. Eine Art High-Score-Liste, die über Web-Browser einzusehen wäre, würde als kompetitiver Reiz viele Internetnutzer ansprechen. Auch über eine Verkürzung des Experimentes auf 30 Minuten oder weniger müsste nachgedacht werden. Noch immer ist die Qualität der Bildschirme nicht so gut, dass Texte darauf so gut verstanden werden wie wenn sie auf Papier präsentiert würden. Gerade für die langen Instruktionstexte aber muss sichergestellt werden, dass Versuchspersonen sie lesen und auch verstehen. Die Aufsplitterung der Texte in kurze quizähnliche Bausteine könnte den Anreiz erhöhen und gleichzeitig das Problem verringern, welches durch gescrollte Texte für das differenzierte Lesen entsteht (siehe Ballstaedt, 1997, S. 91). Dies sind z.T. kleine Veränderungen, die die Abbruchrate des Experimentes aber dennoch erheblich senken könnten.

Das Problem der absichtlichen Verfälscher zeigt sich wohl auch aufgrund der Länge des Experimentes nicht nur zu Anfang des Programms, sondern auch innerhalb der zu bearbeitenden Durchgänge. In der vorliegenden Untersuchung gab es z.B. einige Fälle, die aufgrund von mangelnden Eingaben ausgesondert wurden. Das Kriterium war, dass innerhalb der ersten 21 Zeittakte 13 oder mehr Takte leer gelassen wurden. Bei solchen Kriterien muss natürlich die Frage gestellt werden, ob diese zu rechtfertigen sind oder nur der Datenbeschönigung dienen. Zahlreiche andere Kriterien wären denkbar, sowohl in der negativen Form, wie z.B. es dürfen keine zwei Durchgänge in weniger als zwei Minuten bearbeitet worden sein, als auch in der Form einer Positiv-Liste mit Kriterien, die erfüllt sein müssen z.B. Bearbeitungszeit des Zieldurchgangs mindestens eine Minute, die Hälfte aller Takteingaben müssen sich unterscheiden etc. Weitere Internetexperimente mit der Online-DYNAMIS-Version könnten für die Erfahrung sorgen, die nötig scheint, um ein einheitliches Kriterium zum Aussortieren der Verfälscher finden zu können.

Neben diesen Kritikpunkten und Verbesserungsvorschlägen für das Online-DYNAMIS-System sollte aber der wichtige positive Punkt nicht vernachlässigt werden, dass das neue

System auch unter realer Belastung eines Internetexperiments seine Lauffähigkeit unter Beweis gestellt hat. Neben dem Auffinden von kleineren Schwächen und Fehlern, die es jetzt noch zu verbessern gilt, sind nun auch psychologische Ziele der Testkonstruktion ins Auge zu fassen.

Teile wie der soziodemographische Fragebogen oder auch der Test der Wissensdiagnostik (siehe Abschnitt 'Wissensdiagnostik' auf Seite 22) wurden ganz neu entwickelt und müssten daher einer genaueren Evaluation unterzogen werden. In dieser Untersuchung wurden letztlich je drei Skalen für das Strukturwissen der richtigen und falschen Antworten verwendet. Für diese Skalen der Wirkzusammenhänge, Wirkrichtung und Wirkstärke müssten nun genaue Itemanalysen für Trennschärfe und Schwierigkeit erfolgen. Jede mögliche Verbindung zwischen Eingabe- und Ausgabevariablen wäre dann als Item zu betrachten. Jede Zeile der Wissensdiagnostik in Abbildung 2 auf Seite 25 entspricht einem Item. Auch die Verbindungen der Ausgabevariablen auf andere Ausgabevariablen oder im Falle der Eigendynamik auf sich selbst sind Items. Gerade diese letztgenannten Items könnten aber auch eine eigene Skala bilden, die sich in ihrer Dimensionalität leicht von den erstgenannten Verbindungsitens unterscheidet.. Eine Aggregation der Items z.B. für die Wirkzusammenhänge sollte aber nur bei gegebener Homogenität der Skala erfolgen (Amelang und Zielinski, 1997).

Die nachfolgend beschriebenen Geschlechterunterschiede können aufgrund der bisher noch mangelnden Vergleichsdaten nicht eindeutig zugewiesen werden, ob sie systemspezifisch für die Online-DYNAMIS-Version sind oder ob sie sich für Internetstichproben generalisieren lassen. Gerade deshalb sollen sie nun in diesem Diskussionsteil besprochen werden.

6.4 Geschlechtseffekte

Großes Augenmerk wird bei Untersuchungen zum komplexen Problemlösen gerne auf die Geschlechterunterschiede gelegt. In dieser Untersuchung sind es vermehrt die soziodemographischen Daten, in denen sich Frauen und Männer unterscheiden. Diese Unterschiede werden hier der Vollständigkeit halber genannt, sollen aber nicht alle näher interpretiert werden. Für eine Übersicht siehe auch Tabelle 18. Statistisch bedeutsam wird der Unterschied zwischen Frauen und Männern beim Alter, der Computererfahrung, den Schulnoten zur Gesamtabchlussnote und der Deutschnote. Interessanterweise ergibt sich kein bedeutender Mittelwertsunterschied für die Mathematiknote oder den Wert für das Computerliking, also wie gerne mit Computern gearbeitet wird. Daher scheint der Unterschied

in der Computererfahrung eher auf das Alter zurückzuführen zu sein als auf das Geschlecht, da die Frauen dieser Untersuchung sogar noch etwas lieber mit Computern arbeiten ($M = 3.96$) als die Männer ($M = 3.82$).

Tabelle 18: Mittelwerte und Standardabweichungen einiger soziodemographischer und abhängiger Variablen nach Geschlecht mit den Ergebnissen einer ANOVA und Angaben zur Effektgröße.

	Männer		Frauen		F-Test		Effektgröße
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>F</i> (1,52)	<i>p</i>	ϵ
Alter	29.07	7.97	25.36	3.98	4.432	.040*	.56
Note - Mathematik	2.37	0.96	2.11	1.33	0.691	.410	.23
Note - Deutsch	2.32	0.98	1.72	0.66	6.720	.012*	.68
Note - Abschluss	2.29	0.73	1.88	0.54	5.175	.027*	.61
Computerliking	3.82	0.86	3.96	0.89	0.331	.568	.16
Computererfahrung	3.82	0.95	3.32	0.69	4.764	.034*	.58
Ziel - GdS	6.03	2.98	6.32	1.68	0.183	.671	.12
Zeit - Gesamt (min)	34.03	16.76	33.48	11.63	0.019	.892	.04
Taktiken aggregiert	18.89	8.31	13.68	8.15	5.290	.026*	.61
Zeit - Wissenstests in min.	10.11	6.76	12.13	7.06	1.130	.293	.29
Wissen nach Durchgang 3 (Treffer)	5.96	5.87	4.6	3.14	1.074	.305	.29
Wissen nach Durchgang 3 (Fehler)	4.64	4.06	7.2	4.56	4.664	.036*	.57
Wissensindex	1.32	6.1	-2.6	5.33	6.143	.017*	.65

* $p < 0.05$

Interessant ist, dass Männer scheinbar systematischer vorgehen. Sie nutzen im Schnitt 5.21 Taktiken mehr bis zum Zieldurchgang als die Frauen. Dies führt wohl im Sinne der Hypothese II auch zu mehr richtigen Antworten im Wissenstest und der geringeren Fehleranzahl. Auffällig dabei ist, dass Frauen mehr Falschangaben im Wissenstest machen als korrekte Angaben und dennoch zu gleich guten Steuerleistungen (σ : $M = 6.03$; φ : $M = 6.32$) kommen. Vielleicht gibt es neben dem systematischen Vorgehen der Männer doch auch so etwas wie die Intuition der Frauen. Dies wäre dann in dieser Untersuchung als eher ganzheitlicher Überblick zu verstehen, der alle vier Variablen und ihre Einflüsse untereinander im Auge behält, allerdings nicht näher aufgeklärt werden kann. Diese nicht vorhandenen Unterschiede könnten aber auch auf die mangelnde Differenziertheit der Zielwerte zurückzuführen sein. Aufgrund der bereits angesprochenen Schwierigkeit des Experiments liegt im Falle der Steuerungsleistung (GdS) eine Art Bodeneffekt vor.

6.5 Einfluss von Intelligenz und mathematischen Fertigkeiten

Gerade bei den DYNAMIS-Szenarien, in denen mit Zahlen operiert wird, wird allgemein angenommen, dass mathematische Fertigkeiten für einen Großteil der Varianzaufklärung der Steuerungsleistung verantwortlich sind. Ebenfalls wird dies inzwischen von der allgemeinen Intelligenz angenommen. Süß (1996) konnte zeigen, dass gerade die Subskale „Verarbeitungskapazität“ (K) des Berliner Intelligenzstruktur-Tests (Jäger, Süß & Beauducel, 1997) ein guter Prädiktor für die Leistung beim komplexen Problemlösen ist.

Dieser Untersuchung wurde kein expliziter Intelligenztest vorgeschaltet, sondern lediglich die Schulabschlussnote sowie die letzte Mathematik- bzw. Deutschnote erhoben. Sie sollten aufgrund der hohen Korrelation mit IQ-Tests (vgl. Amelang, 1997, S. 246) als Indikatoren der allgemeinen Intelligenz dienen. Überraschenderweise zeigt sich bei keiner der drei Schulnoten eine Korrelation zur Steuerungsleistung. Weder bei der Abschlussnote ($r = .021$; $p = .883$) noch bei der Deutschnote ($r = -.107$; $p = .445$) und - besonders verwunderlich - auch nicht bei der Mathematiknote ($r = -.091$; $p = .518$) ergibt sich eine signifikante Korrelation.

Wird allerdings die Varianz der Mathematiknote aus den Korrelationen der Steuerungsleistung mit der Abschlussnote herauspartialisiert, erhöhen sich die Koeffizienten deutlich. In Tabelle 19 sind die Korrelationskoeffizienten für die normale bivariate Korrelation angezeigt und darunter für die partielle Korrelation, kontrolliert für die Mathematiknote.

Tabelle 19: Korrelationen der Abschlussnote mit Prädiktoren für die Steuerungsleistung, bivariate und für die Mathematiknote partialisierte Korrelation.

		Ziel - GdS	Computererfahrung	Zeit der Durchgänge ohne Wissenstest	Wissensindex Gesamt
Bivariate	r	.021	.213	.084	-.017
Korrelation	p	.883	.130	.552	.907
Partielle	$r_{\text{part-mathe}}$.107	.351	.139	.213
Korrelation	p	.455	.012	.331	.134

Es zeigt sich ein indirekter Einfluss der Mathematiknote auf alle angegebenen Variablen. Diese Variablen der drei rechten Spalten sind die drei besten Prädiktoren der Regression aus Abschnitt 'Ergebnisse zur Steuerungsleistung' neben der Abschluss- und der Mathematiknote für die Steuerungsleistung GdS (siehe Modell 1 in Tabelle 12 auf Seite 86).

Wie in Tabelle 12 auf Seite 86 zu sehen, ist die Abschlussnote mit einem standardisierten Steigungskoeffizienten von $\beta = .421$ der wichtigste Prädiktor. Die einfache Korrelation ergibt

jedoch wie bereits erwähnt nur $r = .021$. Es müssen daher hohe Kollinearitäten in diesem Regressionsmodell vorliegen, die nicht oder zumindest nur schwer zu interpretieren sind.

Für das in dieser Untersuchung verwendete DYNAMIS-Szenario scheint es so zu sein, dass die Steuerungsleistung am besten durch die Prädiktoren der Computererfahrung, der Bearbeitungszeit und des Strukturwissens vorhergesagt werden kann. Überlagert werden diese Prädiktoren jedoch durch den Anteil der allgemeinen Intelligenz (erschlossen aus der Abschlussnote), der auf die mathematischen Fertigkeiten zurückgeht (erschlossen aus der Mathematiknote). Dies führt dazu, dass zwar nicht das System besser gesteuert wird, wohl aber die Bearbeitung des Wissenstests verbessert und die Bearbeitungszeit verkürzt wird, jeweils bei gleich guter Steuerleistung. Zudem scheint die Erfahrung mit Computern durch eine höhere mathematische Fertigkeit nivelliert zu werden. Diese Überlagerungseffekte zeigen sich auch in den veränderten Korrelationskoeffizienten in Tabelle 19.

Die Regressionsgleichung ist damit folgendermaßen zu interpretieren: Wer eine gute Abschlussnote hat, viel Computererfahrung mitbringt, sich Zeit für das Experiment nimmt und Strukturwissen erwirbt und dies trotz und gerade mit einer schlechten Mathematiknote, wird voraussichtlich erfolgreich das System steuern.

Zusammenfassend wird das Modell noch einmal an dieser Stelle dargestellt (Tabelle 20).

Tabelle 20: Regressionsmodell zur Steuerungsleistung (GdS) mit standardisierten Koeffizienten..

Modell	Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	Signifikanz	
	b	Standardfehler	β	t	p
1 Note - Abschluss	1.541	.535	.421	2.880	.006
Computererfahrung	-1.097	.346	-.390	-3.174	.003
Zeit der Durchgänge ohne Wissenstest	-0.089	.036	-.314	-2.458	.018
Wissensindex Gesamt	-0.111	.057	-.276	-1.965	.055
Note - Mathematik	-1.096	.324	-.505	-3.382	.001

Ein solches Modell muss aufgrund der hohen Kollinearitäten mit Vorsicht interpretiert werden. Gerade in diesem recht numerischen DYNAMIS-System mit einer Wissensdiagnostik, die ohne Graphiken auskommen musste und online zu bedienen war, bleibt durchaus vorstellbar, dass gute mathematische Fertigkeiten halfen, bessere Werte im Strukturwissen zu erlangen, ohne wirklich das System besser steuern zu können oder verstanden zu haben. Versuchsteilnehmenden mit schwachen mathematischen Fähigkeiten war es vielleicht aufgrund der schwierigen Wissensdiagnostik nicht möglich, ihr Wissen vollständig in den Wissenstest zu übertragen.

7. Ausblick und Fazit

In diesem Ausblick soll noch einmal kurz und prägnant dargestellt werden, was die zukünftige Forschung mit DYNAMIS-Szenarien für den Bereich der Taktiken (7.1) und den der Analogien (7.2) bereithält, bevor abschließend ein Fazit gezogen wird.

7.1 Prozessauswertung

Als großer Vorteil der Problemlöseforschung mit komplexen dynamischen Systemen wird die Tatsache angesehen, dass computergesteuerte Systeme nahezu den gesamten Bearbeitungsprozess der Problemlösenden protokollieren und damit psychologischen Analysen zugänglich machen können. Dagegen steht in der Intelligenzdiagnostik meist nur ein einzelner Gesamtwert oder eine geringe Anzahl Skalenwerte, die aber alle das Ergebnis der Testbearbeitung sind.

Mit den in dieser Arbeit beschriebenen Taktiken und ihrer differenzierten Darstellung in Bezug auf den Wissenserwerb ist es möglich, für einzelne Takte verschiedener Durchgänge Auszählungen zu machen und so ein Bild über den Bearbeitungsprozess zu bekommen. Für genauere Prozessanalysen sollte aber noch differenzierter vorgegangen werden. Einen solch differenzierten Ansatz verfolgt Rollett (2002), indem er Eingriffe in dynamische Systeme nach verschiedenen Kriterien kategorisiert und so strukturelle Eigenschaften auszählen kann.

Dadurch entstehen vier Basisstrategien⁸, aus denen sich das strategische Verhalten der Versuchsperson ableiten lässt. Jede Basisstrategie besteht aus einer unterschiedlichen Anzahl von Variablen. Die Basisstrategie *Einsatz von Nullen* z.B. würde in einem System mit zwei Eingabevariablen wie „WITS“ die zwei Variablen „Nulleingriffe“ und „Einzeleingriffe“ zählen. Mit drei Variablen würde die Basisstrategie *Einsatz von Größenunterschieden* neben den „starken Eingriffen“ auch gleiche Eingriffe und solche, die sich nicht deutlich unterscheiden, zählen. Die zwei restlichen Basisstrategien sind noch die Basisstrategie *Einsatz von Vorzeichen* mit fünf Variablen und die Basisstrategie zu *Wiederholungen* innerhalb eines Durchgangs mit insgesamt neun verschiedenen Variablen, die ausgezählt werden. Die Anzahl der Variablen zu jeder Basisstrategie bezieht sich jeweils auf ein System mit zwei Eingabevariablen. Bei Systemen mit drei oder mehr Eingabevariablen steigt die Anzahl der Variablen zu den Basisstrategien z.T. erheblich.

Der Ansatz von Rollett könnte auch im DYNAMIS-System effektiv eingesetzt werden. Da die Variablen der vier Basisstrategien von Rollett genau beschrieben sind, wäre es m.E. möglich,

⁸ In diesem Ausblick wird die Terminologie Rolletts verwendet, auch wenn diese Strategien sonst in dieser Arbeit eher als Taktiken bezeichnet werden müssten (vgl. Abschnitt 'Stile, Strategien und Taktiken' Seite 31).

durch entsprechende Programmierung eine automatische Auswertung nach seinem Ansatz zu integrieren. Alle Szenarien dieses erweiterten DYNAMIS-Systems bekämen eine differenzierte Diagnostik zur Prozessauswertung, die es ermöglichen würde, das (strategische) Vorgehen der Teilnehmer genauer zu analysieren. Rollett schätzt die Aufklärung der Leistungsvarianz von GdS durch seine Basisstrategien für ein 2x2 Variablensystem auf 50-60% (Rollett, persönliche Mitteilung, 31.07.03). Aufgrund des erheblichen Programmieraufwands zur Auszählung der vier Basisstrategien aus den Logdateien des neuen Online-DYNAMIS-Systems und in Anbetracht des begrenzten Zeitrahmens einer Diplomarbeit konnte dies in der vorliegenden Arbeit jedoch nicht nachgeprüft werden.

7.2 Analogie und komplexe dynamischen Szenarien

Mit dem Wissen über die Wirkungsweise von Analogien beim Lösen komplexer Probleme ist der Transfer von Wissen über spezifische Problemsituationen hinaus eng verknüpft. Weiß man erst einmal über die Auswirkungen von Analogien Bescheid, können aus diesem Wissen auch Schlussfolgerungen zur Förderung des Wissenstransfers gezogen werden.

Auch für die Kreativitätsforschung ergeben sich Möglichkeiten. So spricht Mayer (1997) der Analogieforschung das Potential zu „...for addressing some of the most fundamental issues in the psychology of problem solving, such as where creative solutions come from.“ (Mayer, 1997, p. 454).

Die Forschung zu Analogien und Transfer ist daher noch lange nicht am Ende. Ob der Einsatz komplexer dynamischer Szenarien dabei eine Rolle spielen kann, wird sich zeigen müssen. Die Vorteile liegen jedoch auf der Hand. Die Möglichkeiten, die sich daraus ergeben, strukturell gleichen Problemen verschiedene semantische Einkleidungen zu geben und damit ihre Oberflächenstruktur zu ändern, bieten ein enormes Potential. Zudem können die Strukturmerkmale nahezu stufenlos verändert werden. Der größte Vorteil, den ein dynamisches System wie DYNAMIS jedoch mitbringt, ist die vollständige Speicherung des Bearbeitungsprozesses. Zum Beispiel kann durch Protokollierung der Veränderungen in den Analogiebezeichnungen der Zeitpunkt der Einsicht über die analogen Strukturen genau festgehalten werden. Auch die Reihenfolge, welche Strukturen wann erkannt werden, ist damit einer weiteren psychologischen Untersuchung zugänglich.

7.3 Fazit ⁹

Die Diskussion der Ergebnisse hat gezeigt, dass das Treatment der Analogien eine Wirkung hat, wenn diese auch deutlich geringer ausfällt als z.B. die Wirkung von Taktiken, wenn beide ausschließlich über Instruktionstexte vermittelt werden. Mit dieser Untersuchung ist der erste Schritt getan, um die Wirkung von Analogien als Strategien beim komplexen Problemlösen aufzuklären, viele weitere müssten folgen, um einen besseren Einblick in die Funktionsweise von Analogien zu bekommen und praktischen Nutzen daraus ziehen zu können.

Die differenzierte Darstellung einzelner Taktiken als „strong methods“ des komplexen Problemlösens und die Analyse von deren Zusammenhängen mit dem Erwerb von Strukturwissen weisen auf einen unterschiedlichen Erklärungswert der einzelnen Taktiken hin. Zeigen zukünftige Untersuchungen ähnliche Ausprägungen, so kann die Liste der „strong methods“ eventuell auf einige wesentliche von Menschen beim komplexen Problemlösen eingesetzte Taktiken reduziert werden.

Letztlich bleibt noch festzustellen, dass das Online-DYNAMIS-Programm funktioniert und einsetzbar ist. Auch hier muss noch viel Arbeit investiert werden, unter anderem müsste als nächster Schritt eine Wissensdiagnostik für den Online-Bereich entwickelt und evaluiert werden. Dennoch sollten Internetexperimente auch mit diesem Instrument weitergeführt werden, schon allein um den Erhebungsort wechseln und vom psychologischen Institut auslagern zu können z.B. in die Interneträume von Schulen oder anderen universitären Instituten.

⁹ Eine solche Arbeit kann nicht ohne die Hilfe von Freunden fertig gestellt werden. Besonderer Dank gilt, in der Reihenfolge ihres Erscheinens als sie gebraucht wurden und dann da waren Elisa Franetzki, Matthias Blümke, insbesondere Jan Hollerbach und Bernadette Ohmer, Melany Richter und Silvia Natale.

Literaturverzeichnis

- Ahrens, H.J. (1984). *Methoden der Persönlichkeitsforschung: Werkzeug- oder Modellfunktion*. In M. Amelang, & H. J. Ahrens (1984). *Brennpunkte der Persönlichkeitsforschung. Band 1*. Göttingen: Hogrefe
- Amelang, M. & Bartussek, D. (1997). *Differentielle Psychologie und Persönlichkeitsforschung (4. überarbeitete und erweiterte Auflage)*. Stuttgart: Kohlhammer.
- Amelang, M. & Zielinski, W. (1997). *Psychologische Diagnostik und Intervention (2., korrigierte, aktualisierte und überarbeitete Auflage)*. Heidelberg: Springer.
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Anderson, J. R. (1989). The analogical origins of errors in problem solving. In D. Klahr & D. Kotovsky (Eds.). *Complex information processing: The impact of Herbert A. Simon* (pp. 343-371). Hillsdale, NJ: Erlbaum.
- Anderson, J. R. (2000). *Learning and memory. An integrated approach*. 2nd ed. New York: John Wiley & Sons.
- Anderson, J. R. & Lebiere, C. (Eds.). (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Anderson, J. R. & Thompson, R. (1989). Use of analogy in a production system architecture. In S. Vosniadou & A. Ortony (Eds.). *Similarity and analogical reasoning* (pp. 267-297). Cambridge, MA: Cambridge University Press.
- Atkinson, J. W. (1964). *An introduction to motivation*. Princeton, NJ: Van Nostrand.
- Ballstaedt, S.-P. (1997). *Wissensvermittlung. Die Gestaltung von Lernmaterial*. Weinheim: Beltz.
- Bandilla, W. (1999). WWW-Umfragen - Eine alternative Datenerhebungstechnik für die empirische Sozialforschung?. In B. Batinic, A. Werner, L. Gräf und W. Bandilla (Hrsg.), *Online Research. Methoden, Anwendungen und Ergebnisse* (S. 9-19). Göttingen: Hogrefe.
- Beckmann, J. F. (1994). Lernen und komplexes Problemlösen. Ein Beitrag zur Konstruktvalidierung von Lerntests. Bonn: Holos.
- Beckmann, J. F. & Guthke, J. (1995). Complex problem solving intelligence and learning ability. In P. A. Frensch & J. Funke (Eds.). *Complex problem solving: The European perspective* (pp. 177-200). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

- Berry, D. C. & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *The Quarterly Journal of Psychology*, 34A, 209-231.
- Blanchette, U. & Dunbar, K. (2000). How analogies are generated: The roles of structural and superficial similarity. *Memory and Cognition*, 28(1), 108-124.
- Blümke, M. & Friese, M. (2003). *Der Einfluss von Redundanz im Implicit Association Test. Relevante und irrelevante Eigenschaften des Stimulusmaterials*. Unveröffentlichte Diplomarbeit, Ruprecht-Karls-Universität Heidelberg.
- Bortz, J. (1993). *Statistik für Sozialwissenschaftler* (4. Auflage). Heidelberg: Springer.
- Bortz, J. (1995). *Forschungsmethoden und Evaluation für Sozialwissenschaftler* (2. Aufl.). Heidelberg: Springer.
- Bortz, J. (1999). *Statistik für Sozialwissenschaftler* (5. Auflage). Heidelberg: Springer.
- Bruner, S. B., Goodnow, J. J. & Austin, A. G. (1967). *A study of thinking*. (7th ed.). New York: John Wiley & Sons.
- Buchner, A. (1999). Komplexes Problemlösen vor dem Hintergrund der Theorie finiter Automaten. *Psychologische Rundschau*, 50, 198-205.
- Buchner, A., Erdfelder, E., & Faul, F. (2001). *How to Use G*Power*. [WWW-Dokument]. Verfügbar unter: www.psych.uni-duesseldorf.de/aap/projects/gpower/how_to_use_gpower.html [16.07.2003].
- Catrambone, R. & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(6), 1147-1156. [zitiert nach Robertson, 2001]
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (rev. ed.)*. Hillsdale, NJ: Erlbaum.
- Dörner, D. (1976). *Problemlösen als Informationsverarbeitung*. Stuttgart: Kohlhammer.
- Dörner, D. (1979). *Problemlösen als Informationsverarbeitung* (2. Aufl.). Stuttgart: Kohlhammer.
- Dörner, D. (1992). *Die Logik des Mißlingens. Strategisches Denken in komplexen Situationen*. Reinbek: Rowohlt
- Dörner, D. (1995). Problemlösen und Gedächtnis. In D. Dörner & E. Van der Meer (Hrsg.). (1995). *Das Gedächtnis. Probleme – Trends – Perspektiven* (S. 295-320). Göttingen: Hogrefe.
- Dörner, D. (1999). *Bauplan für eine Seele*. Reinbek: Rowohlt.

- Dörner, D., Drewes, U. & Reither, F. (1975). Über das Problemlösen in sehr komplexen Realitätsbereichen. In W.H. Tack (Hrsg.), *Bericht über den 29. Kongreß der DGfPs in Salzburg 1974* (Band 1, S. 339-340). Göttingen: Hogrefe.
- Dörner, D., Kreuzig, H. W., Reither, F. & Stäudel, T. (Hrsg.). (1983). *Lohhausen. Vom Umgang mit Unbestimmtheit und Komplexität*. Bern: Huber.
- Dörner, D., Schaub, H. & Strohschneider, S. (1999). Komplexes Problemlösen - Königsweg der Theoretischen Psychologie? *Psychologische Rundschau*, 50, 198-205.
- Dörner, D. & Schölkopf, J. (1991). Controlling complex systems: Or Expertise as „grandmother's know-how.“ In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospect and limits* (pp. 218-239). New York, NY: Cambridge University Press.
- Dörner, D., Stäudel, T., & Strohschneider, S. (1986). *MORO: Programmdokumentation*. Bamberg: Memorandum Nr. 23, Lehrstuhl Psychologie II, Universität Bamberg.
- Dorsch, F., Häcker, H. & Stapf, K. H. (1996). *Dorsch. Psychologisches Wörterbuch* (12. überarbeitete und erweiterte Auflage). Bern: Huber.
- Drosdowski, G., Müller, W., Scholze-Stubenrecht, W. & Wermke, M. (Hrsg.). (1990). *Duden. Das Fremdwörterbuch* (5., neu bearbeitete und erweiterte Auflage). Mannheim: Bibliographisches Institut & F.A. Brockhaus AG.
- Dweck, C. S. & Eliot, E. S. (1983). Achievement motivation. In Mussen, P.H. (Eds.), *Handbook of child psychology* (Vol. 4). New York: Wiley.
- Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments and Computers*, 28, 1-11.
- Funke, J. (1985). Steuerung dynamischer Systeme durch Aufbau und Anwendung subjektiver Kuasalmodelle. *Zeitschrift für Psychologie*, 193, 443-365.
- Funke, J. (1986). Ein Forschungsprogramm zur subjektiven Repräsentation dynamischer Kleinsysteme: Aufbau und Anwendung von Wissen in Abhängigkeit von Person- und Systemmerkmalen. *Berichte aus dem psychologischen Institut der Universität Bonn*, Band 12, Heft 1.
- Funke, J. (1990). Systemmerkmale als Determinanten des Umgangs mit dynamischen Systemen. *Sprache und Kognition*, 9, 143-153.
- Funke, J. (1992). Wissen über dynamische Systeme: Erwerb, Repräsentation und Anwendung. Berlin: Springer-Verlag.
- Funke, J. (1993). Microworlds based on linear equation systems: A new approach to complex

- problem solving and experimental results. In G. Strube & K.-F. Wender (Eds). *The cognitive psychology of knowledge* (pp. 313-330). Amsterdam: Elsevier.
- Funke, J. (1995). Experimental research on complex problem solving. In P. A. Frensch & J. Funke (1995). *Complex problem solving. The european perspektive*. Hillsdale: Lawrence Erlbaum.
- Funke, J. (2001). Dynamic systems as tools for analysing human judgement. *Thinking and reasoning*, 7(1), 69-89.
- Funke, J. (2003). *Problemlösendes Denken*. Stuttgart: Kohlhammer.
- Funke, J., Fahnenbruck, G. & Müller, H. (1986). DYNAMIS. Ein Computerprogramm zur Simulation dynamischer Systeme. *Berichte aus dem psychologischen Institut der Universität Bonn*, Band 12, Heft 3.
- Funke, J. & Müller, H. (1988). Eingreifen und Prognostizieren als Determinanten von Systemidentifikation und Systemsteuerung. *Sprache & Kognition*, 7, 176-186.
- Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155-170.
- Gentner, D. (1989). The mechanisms of analogical reasoning. In S. Vosniadou & A. Ortony (Eds.). (1989). *Similarity and analogical reasoning*. London: Cambridge University Press.
- Gigerenzer, G. (1981). *Messung und Modellbildung in der Psychologie*. München: Reinhardt.
- Gigerenzer, G. (1984). Messung, Modellbildung und die „Kognitive Wende“. In M. Amelang, & H. J. Ahrens (1984). *Brennpunkte der Persönlichkeitsforschung. Band 1*. Göttingen: Hogrefe.
- Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, 98, 254-267.
- Gigerenzer, G., & Selten, R. (2001). *Bounded rationality: The adaptive toolbox*. Oxford: Oxford University Press.
- Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox. In G. Gigerenzer, P. M. Todd, & A. R. Gigerenzer (Eds.), *Simple heuristics that make us smart* (pp. 3-34). Oxford: Oxford University Press.
- Greeno, J. G. & Simon, H. A. (1984). Problem solving and reasoning. In R. C. Atkinson, R. Herrnstein, G. Lindzey & R. D. Luce (Eds.), *Stevens' handbook of experimental psychology*. New York: Wiley.
- Hager, W. & Westermann, R. (1983). Planung und Auswertung von Experimenten. In J.

- Bredenkamp & H. Feger (Hrsg.), *Enzyklopädie der Psychologie, Serie: Forschungsmethoden der Psychologie, Band 5: Hypothesenprüfung* (S. 24-238). Göttingen: Hogrefe.
- Halpern, D. F., Hanson, C. & Riefer, D. (1990). Analogies as an aid to understanding and memory. *Journal of Educational Psychology*, 82(2), 298-305.
- Hauptmanns, P. (1999). Grenzen und Chancen von quantitativen Befragungen mit Hilfe des Internet. In B. Batinic, A. Werner, L. Gräf und W. Bandilla (Hrsg.), *Online Research – Methoden, Anwendungen und Ergebnisse* (S. 21-38). Göttingen: Hogrefe.
- Havlicek, L. L. & Peterson, N. L. (1977). Effect of the violation of assumptions upon significance levels of the Pearson r. *Psychological Bulletin*, 84, 373-377. [zitiert nach Bortz, 1993]
- Hebb, D. O. (1955). Drives and C.N.S. (Conceptual Nervous System). *Psychological Review*, 62, 243-254.
- Hesse, F. W. (1979). *Zur Verbesserung menschlichen Problemlöseverhaltens durch den Einfluß unterschiedlicher Trainingsprogramme*. Aachen: Dissertation. RWTH. [zitiert nach Dörner, 1992]
- Holyoak, K. J. (1984). Mental models in problem solving. In J. R. Anderson & S. M. Kosslyn (Eds.). (1984). *Tutorials in learning and memory: Essays in honour of Gordon Bower* (pp. 193-218). San Francisco, CA: W. H. Freeman. [zitiert nach Robertson, 2001]
- Holyoak, K. J. (1985). *The pragmatics of analogical transfer*. In G. H. Bower (Ed.). (1985). *The psychology of learning and motivation* (Vol. 19). New York: Academic Press. [zitiert nach Robertson, 2001]
- Holyoak, K. J. & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory and Cognition*, 15, 332-340). [zitiert nach Robertson, 2001]
- Holyoak, K. J. & Thagard, P. (1989a). Analogical mapping by constraint satisfaction. *Cognitive Science*, 13(3), 295-355. [zitiert nach Robertson, 2001]
- Holyoak, K. J. & Thagard, P. (1989b). A computational model of analogical problem solving. In S. Vosniadou & A. Ortony (Eds.). (1989). *Similarity and analogical reasoning*. London: Cambridge University Press. [zitiert nach Robertson, 2001]
- Holzkamp, K. (1972). *Kritische Psychologie. Vorbereitende Analysen*. Frankfurt a. M.: Fischer.
- Hunt, E. B. (1962). *Concept learning: An information processing problem*. New York:

Wiley.

- Hussy, W. (1984). *Denkpsychologie. Ein Lehrbuch. Band 1*. Mainz: Kohlhammer.
- Hussy, W. (1986). *Denkpsychologie. Ein Lehrbuch. Band 2*. Mainz: Kohlhammer.
- Janetzko, D. (1999). *Statistische Anwendungen im Internet*. München: Addison-Wesley-Longman.
- Jäger, A. O., Süß, H.-M. & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test. BIS-Test; Form 4; mit einem separat verwendbaren Kurztest der allgemeinen Intelligenz und der Verarbeitungskapazität*. Göttingen: Hogrefe.
- Johnson, D. M. (1972). *Systematic introduction to the psychology of thinking*. New York: Harper & Row.
- Kersting, M. (1999). Diagnostik und Personalauswahl mit computergestützten Problemlöseszenarien? Zur Kriteriumsvalidität von Problemlöseszenarien und Intelligenztests. Göttingen: Hogrefe.
- Kelly, G. A. (1986). *Die Psychologie der persönlichen Konstrukte*. Paderborn: Junfermann-Verlag.
- Klahr, D. & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12, 1-48.
- Knoblich, G. (2002). Problemlösen und logisches Schließen. In J. Müsseler & W. Prinz (Hrsg.) (2002). *Allgemeine Psychologie*. Heidelberg: Spektrum Akademischer Verlag.
- Krantz, J. H. & Dalal, R. (2000). Validity of web-based psychological Research. In M. H. Birnbaum (Hrsg.). *Psychological Experiments on the Internet* (S. 35-60). San Diego: Academic Press.
- Larkin, J. H., McDermott, J., Simon, D.P. & Simen, H. A. (1980a). Expert and novice performance in solving physics problems. *Science*, 208, 1334-1342. [zitiert nach Nickerson, 1988]
- Larkin, J. H., McDermott, J., Simon, D.P. & Simen, H. A. (1980b). Modes of competence in solving physics problems. *Cognitive science*, 4, 317-345.
- Mayer, R. E. (1997). *Thinking, problem solving, cognition*, (2nd ed.). New York: W. H. Freeman and Company.
- Mietzel, G. (2001). *Pädagogische Psychologie des Lernens und Lehrens* (6., korrigierte Auflage). Göttingen: Hogrefe.
- Müller, H. (1993). *Komplexes Problemlösen: Reliabilität und Wissen*. Bonn: Holos.
- Müller, H., Funke, J. & Rasche, B. (1988). *Wechselseitige Abhängigkeiten: Zum Einfluss*

von Eigendynamik und Nebenwirkungen auf die Bearbeitung dynamischer Systeme. Berichte aus dem Psychologischen Institut der Universität Bonn, Band 13, Heft 2. [zitiert nach Müller, 1993].

- Newell, A. & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Nickerson, R. S. (1998). *The Teaching of Thinking and Problem Solving*. In Sternberg, R. J. (1998). *Thinking and Problem Solving* (2nd ed). London: Academic Press
- Pervin, L.A. (2000). *Persönlichkeitstheorien*.(4. Aufl.). München: Reinhardt (UTB).
- Plötzner, R. & Spada, H. (1995). Wissenserwerb: Funktionsprinzipien von Lernprozessen. In D. Dörner & E. Van der Meer (Hrsg.). (1995). *Das Gedächtnis. Probleme – Trends – Perspektiven* (S. 381-412). Göttingen: Hogrefe.
- Preußler, W. (1996). Zur Rolle expliziten und impliziten Wissens bei der Steuerung dynamischer Systeme. *Zeitschrift für Experimentelle Psychologie*, 43, 399-434.
- Preußler, W. (1998). Strukturwissen als Voraussetzung für die Steuerung komplexer dynamischer Systeme. *Zeitschrift für Experimentelle Psychologie*, 45, 3, 218-240. Göttingen: Hogrefe.
- Putz-Osterloh, W. (1993a). Strategies for knowledge acquisition and transfer of knowledge in dynamic tasks. In G. Strube & K. F. Wender (Eds). *The cognitive psychology of knowledge* (pp. 331-350). Amsterdam: Elsevier.
- Putz-Osterloh, W. (1993b). Unterschiede im Erwerb und in der Reichweite des Wissens bei der Steuerung eines dynamischen Systems. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 40, 386-410.
- Putz-Osterloh, W. & Lüer, G. (1981). Über die Vorhersagbarkeit komplexer Problemlöseleistungen durch Ergebnisse in einem Intelligenztest. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 28, 309-334.
- Reips, U.-D. (1997a). Das psychologische Experimentieren im Internet. In B. Batinic (Hrsg.). *Internet für Psychologen* (S. 245-265). Göttingen: Hogrefe.
- Reips, U.-D. (1999). Theorie und Techniken des Web-Experimentierens. In B. Batinic, A. Werner, L. Gräf und W. Bandilla (Hrsg.), *Online Research. Methoden, Anwendungen und Ergebnisse* (S. 277-295). Göttingen: Hogrefe.
- Reips, U.-D. (2000). The web experiment method: Advantages, disadvantages, and solutions. In M. H. Birnbaum (Hrsg.). *Psychological Experiments on the Internet* (S. 89-117). San Diego: Academic Press.

- Reither, F. (1979). *Über die Selbstreflexion beim Problemlösen*. Gießen: Universität – Dissertation am FB 06 Psychologie. [zitiert nach Dörner, 1992]
- Robertson, S. I. (2001). *Problem solving*. Hove: Psychology Press.
- Rollett, W. (2002). *Strategieinsatz und Informationsnutzung bei der Exploration und Steuerung komplexer dynamischer Systeme*. In E. van der Meer, H. Hagendorf, R. Beyer, F. Krüger, A. Nuthmann & S. Schulz (Eds.), *43. Kongress der Deutschen Gesellschaft für Psychologie (DGPs)* (pp. 443). Lengerich: Pabst Science Publishers.
- Rollett, W. (in Vorb.). *Strategieinsatz, Informationsnutzung und Motivation beim Explorieren und Steuern komplexer dynamischer Systeme*. Dissertation am Institut für Psychologie, Humanwissenschaftliche Fakultät der Universität Potsdam, Potsdam.
- Sanderson, P. M. (1989). Verbalizable knowledge and skilled task performance: Association, dissociation, and mental models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 729-747.
- Schank, R. C., Berman, T. R. & Macpherson, K. A. (1999). Learning by Doing., In C. M. Reigeluth (Ed.), *Instructional-Design Theories and Models* (pp. 160-181). Mahwah, NJ: Lawrence Erlbaum.
- Sedlmeier, P. & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, *105*, 309-316.
- Simon, H. A. (1972). Theories of bounded rationality. In C. B. McGuire & R. Radner (Eds.). *Decision and Organization* (pp. 161–176). Amsterdam: North-Holland.
- Simon, H. A. (1992). *Economics, bounded rationality and the cognitive revolution*. Aldershot: Elgar.
- Simon, H. A. (2001). *The sciences of the artificial*, (3rd ed.). Cambridge, Mass: MIT Press.
- Spring, M. (2001). *Emotionen und Kontrollüberzeugungen beim komplexen Problemlösen. Eine experimentelle Untersuchung anhand des computersimulierten Problemlöseszenarios FSYS 2.0*. Unveröffentlichte Diplomarbeit, Ruprecht-Karls-Universität Heidelberg.
- Sternberg, R. J. (1997). *Thinking Styles*. New York: Cambridge University Press.
- Süß, H.-M. (1996). *Intelligenz, Wissen und Problemlösen. Kognitive Voraussetzungen für erfolgreiches Handeln bei computersimulierten Problemen*. Göttingen: Hogrefe.
- Süß, H.-M. (1999). Intelligenz und komplexes Problemlösen. *Psychologische Rundschau*, *50*, 220-228.
- Simon, H. A. & Lea, G. (1974). Problem solving and rule induction: A unified view. In L.

- W. Gregg (Ed.), *Knowledge and cognition* (pp. 105-127). Hillsdale, NJ: Erlbaum.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12, 257-285.
- Sweller, J., Mawer, R.F., & Ward, M. R. (1983). Development of expertise in mathematical problem solving. *Journal of Experimental Psychology: General*, 112, 639-661. [zitiert nach Nickerson, 1988]
- Vollmeyer, R., Burns, B. D. & Holyoak, K. J. (1996). The impact of goal specificity on strategy use and the acquisition of problem structure. *Cognitive Science*, 20, 75-100.
- Vollmeyer, R. & Funke, J. (1999). Personen- und Aufgabenmerkmale beim komplexen Problemlösen. *Psychologische Rundschau*, 50, 213-219.
- Wagener, D., & Conrad, W. (1997). *Testmanual zu FSYS 2.0*. Universität Mannheim, Mannheim.
- Westermann, R. (1987). Wissenschaftstheoretische Grundlagen der experimentellen Psychologie. In G. Lüer (Hrsg.), *Allgemeine Experimentelle Psychologie. Eine Einführung in die methodischen Grundlagen mit praktischen Übungen für das Experimentelle Praktikum* (S. 5-42). Stuttgart: Fischer.
- Yerkes, R. & Dodson, J. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurological Psychology*, 18, S. 459-482.

Anhang

Anhang A

Die DYNAMIS – Steuerdatei für das System „WITS“

```

What Is The System (WITS.prm)
2 2 1 7 7 4 0 (nx ny gr mx zt dg fb)
  0  0  100  200  Startwerte
  0  0   0   0  Fehlerwerte
  0  0   20  500  Zielwerte

  1  0   Axx --    „Wirkung“ von X- auf X-Variablen
  0  1   Axx

  0 -1.5   Ayx --    Wirkung von X- auf Y-Variablen
  1 -2.0   Ayx

  1 0.25   Ayy --    Wirkung von Y- auf Y-Variablen
  0 1.1    Ayy

-999999 -999999 -999999 -999999 min
999999 999999 999999 999999 max
      0      0      0      0 add. Konstanten

de
I
II
A
B
en
I
II
A
B

```

Anhang B

Die Werbung in Emails, Newsgroups und Email-Verteilern

Mail an Erstsemester-Mail-Verteiler

Subject: Dein Beitrag zur Wissenschaft - ein Online-Experiment zum "Komplexen Problemlösen"

From: Carsten Schulz

Date: Tue, 22 Oct 2002 15:58:01 +0200

Hallo,

erst zwei Wochen hier und schon der ultimative Info-Overkill?? Das geht nicht nur Dir so...

Am Anfang sind tausend Dinge zu bewältigen: Wohnung suchen, Stundenplan machen, Leute kennen lernen. Wie solche Aufgaben gelöst werden können und welche Prozesse dabei in unseren Köpfen ablaufen - das untersucht die "Komplexe-Problemlöse-Forschung".

Der Info-Overkill ist eines von vielen Problemen, die Menschen mit komplexen Umgebungen haben können. Wir haben nun ein Online-Experiment gestartet, bei dem genauer untersucht werden soll, unter welchen Bedingungen der Umgang leichter oder schwerer fällt.

Das Tolle ist: Jeder kann mitmachen!

Der Link, der dort hinführt lautet:

<http://www.dynamis.uni-hd.de>

Es gibt viele Gründe, nicht draufzuklicken:

- Bei Sonnenschein ist es draußen schöner als vorm Computer (aber schau mal raus...)
- das Experiment ist nicht ganz leicht (aber anonym)
- 30-60 Minuten kann's schon dauern...
- Und: Jeder darf nur einmal teilnehmen!!

Aber es gibt noch viel mehr (und vor allem viel bessere Gründe) doch draufzuklicken:

- Du bekommst eine VPN-Stunde (und davon braucht man ja bekanntlich jede Menge...),
- Du kannst was Tolles gewinnen,
- Du leistest einen Beitrag zur Problemlöse-Forschung,
- Du kannst Dir den Termin frei aussuchen (einfach daheim oder im CIP-Pool mitmachen),
- Du darfst diese Mail auch an Freunde und Bekannte weitergeben (was sogar erwünscht ist),
- und wenn Dein nicht zu stillender Wissensdurst danach verlangt, dann werden Dir die Forschungsergebnisse per Mail zugesandt.

Und jetzt: Viel Spaß!

Newsgroups: de.sci.psychologie**Subject:** Online-Experiment / Komplexes Problemlösen**From:** Carsten Schulz**Date:** Fri, 11 Oct 2002 10:58:44 +0200**Newsgroups:** de.sci.psychologie

Hallo,

unter www.dynamis.uni-hd.de

läuft derzeit meine Untersuchung zum Umgang mit komplexen Szenarien und dem Wissenserwerb in eben solchen.

Vielleicht bin ich hier nicht ganz richtig, dennoch suche ich ein wenig ernsthaftere Teilnehmer als die, die bisher teilgenommen haben.

Da der Versuch über 60 min. dauert, nützen mir 'Reingucker' wenig.

Auch zur Theorie des Forschungsgebietes, also dem Komplexen Problemlösen, finden sich Hinweise unter diesem Link.

Ich danke jedem, der mir hilft für meine Diplomarbeit die nötigen vollständigen Datensätze zu erhalten.

Viele Gruesse,
Carsten Schulz

Newsgroups: de.alt.umfragen**Date:** Wed, 16 Oct 2002 18:22:13 +0200**From:** Carsten Schulz**Newsgroups:** de.alt.umfragen

Hi,

im Rahmen meiner Diplomarbeit im Bereich der Kongnitionsfoschung untersuche ich den Umgang mit komplexen, dynamischen Systemen.

Unter: <http://www.dynamis.uni-hd.de>

läuft diese psychologische Fragestellung noch bis Weihnachten und wahrscheinlich weit darüber hinaus.

Die Teilnahmedauer variiert zwischen 30-60 min. und zu gewinnen gibt es kleine Büchergutscheine.

Allen interessierten Teilnehmenden werden Ergebnisse dieser Untersuchung anfang nächsten Jahres zugestellt.

Vielen Dank allen Teilnehmenden,

Carsten Schulz

Allgemeine Werbung über private Emails

Subject: Suche Versuchspersonen für meine Diplomarbeit

From: Carsten Schulz

Date: Thu, 31 Oct 2002 08:38:45 +0100

To: dynamis@uni-hd.de

Liebe Freunde,

tatsächlich arbeite ich derzeit mit Eifer an meiner Diplomarbeit. Und wie das so ist bei experimentellen, psychologischen Untersuchungen, brauche auch ich viele Versuchspersonen.

Da ich mein Experiment in Java programmiert habe, bin ich in der 'glücklichen' Lage, über das Internet Versuchspersonen zu erheben. Es tut mir Leid, wenn einige diese Mail mehrfach bekommen, dennoch ist es die beste Möglichkeit für mich viele Teilnehmer zu gewinnen.

Was muß man wissen?

- die Dauer des Experimentes: es dauert etwa 45-60 min.
Daher vor Beginn Zeit nehmen und bitte EINEN ernsthaften Versuch wagen.
- Vollständigkeit: nur volle Datensätze kommen bei mir an.
Bitte brecht deshalb nicht ab, nur weil ihr glaubt nicht so gut zu sein. Ich kann mit den Daten schon etwas anfangen.
- die Software: es läuft mit allen Browsern, die Java (ab Version 1.1.8) eingeschaltet haben und mit dem Microsoft Internet Explorer (ab Version 5).
- Es bleibt alles anonym.
- Bitte halte Dich möglichst genau an die Anweisungen.
- Jeder sollte nur einmal mitmachen, den ersten Versuch also bitte durchziehen. Auch wenn er zu misslingen scheint. Ich kann mit den Daten etwas anfangen.

--> Es gibt auch eine Kleinigkeit zu gewinnen:

<http://www.dynamis.uni-hd.de>

Es ist nicht leicht, aber auch kein IQ Test oder so. Mir geht es um den Wissenserwerb in komplexen Systemen unter verschiedenen Bedingungen (werden in den Anweisungen genannt). Leider kann ich vor dem Versuch nicht viel mehr dazu sagen, bin aber gerne bereit hinterher jedem Auskunft zu erteilen.

Und nun noch viel Spaß,

herzliche Grüße aus der Uni-Heidelberg,
Carsten

PS: Ich nehme auch mir unbekanntes VP, vor allem nicht-Studenten sowie Mütter, Väter, Freunde etc.
Also sendet die Mail ruhig weiter und nehmt den Link auf in jede Homepage auf. Danke.

Mail an Drittsemester-Mail-Verteiler

Date: Tue, 26 Nov 2002 16:06:09 +0000

Subject:: Dein Beitrag zur Wissenschaft - ein Online-Experiment zum "Komplexen Problemlösen"

Body:

Hallo,

ich schreibe diese Mail, in der Hoffnung dadurch Versuchspersonen für ein Online-Experiment in Deinem Semester zu gewinnen. Es stellt sich nämlich heraus, dass das gar nicht so leicht ist, wie es scheint. Und wie schwer es ist, brauche ich Dir wohl nicht zu schildern, jetzt nach dem die Versuchspersonen von Dir und anderen aus dem 3. Semester erhoben werden.

Inhaltlich geht es um Komplexes Problemlösen, ja, genau ein Teilbereich der Denkpsychologie und damit der Allgemeinen. Dementsprechend läuft das Experiment in der Abteilung von Prof. Funke.

So nun folgt noch die Werbebotschaft, die die anderen Semester auch bekommen haben und der Link, wo es losgeht.

Ich danke jedem, der sich die Zeit nimmt. Das Experiment läuft noch bis kurz vor Weihnachten.

Gruß, Carsten Schulz

Wir haben ein Online-Experiment gestartet, bei dem genauer untersucht werden soll, unter welchen Bedingungen der Umgang mit komplexen Systemen leichter oder schwerer fällt.

Das Tolle ist: Jeder kann mitmachen!

Der Link, der dort hinführt lautet:

<http://www.dynamis.uni-hd.de>

Es gibt viele Gründe, nicht draufzuklicken:

- Bei Sonnenschein ist es draußen schöner als vorm Computer (aber schau mal raus...)
- das Experiment ist nicht ganz leicht (aber anonym)
- 30-60 Minuten kann's schon dauern...

- Und: Jeder darf nur einmal teilnehmen!!

Aber es gibt noch viel mehr (und vor allem viel bessere Gründe) doch draufzuklicken:

- Du bekommst eine VPN-Stunde (und davon braucht man ja bekanntlich jede Menge...),
- Du kannst was Tolles gewinnen,
- Du leistest einen Beitrag zur Problemlöse-Forschung,
- Du kannst Dir den Termin frei aussuchen (einfach daheim oder im CIP-Pool mitmachen),
- Du darfst diese Mail auch an Freunde und Bekannte weitergeben (was sogar erwünscht ist),
- und wenn Dein nicht zu stillender Wissensdurst danach verlangt, dann werden Dir die Forschungsergebnisse per Mail zugesandt.

Und jetzt: Viel Spaß!

Anhang C

Instruktionstexte und Erklärungen zum DYNAMIS-System „WITS“

Die erste Zeile gibt jeweils an, wie die Textdatei **heißt** und wo sie im DYNAMIS-Programm zu finden ist. Die zweite unterstrichene Zeile gibt an, welche Gruppe den Text zu lesen bekam.

/de/instructions/group0/text0.txt

Text_1 für alle vier Gruppen (KG1, KG2, EG1, EG2)

Hallo und herzlich willkommen zur Simulationssteuerung WITS.

Zwei mal Technisches vorneweg:

- Sollte der Text nicht ganz zu sehen sein, benutzen Sie bitte den Scrollbalken rechts im Fenster oder passen Sie das Fenster ihrem Monitor an. -
- Sie können dieses Fenster gern schließen und diesen Text noch mal anzeigen lassen, wenn Sie einen Blick auf das System geworfen haben, dadurch wird die Erklärung verständlicher. -

Es wird Ihre Aufgabe sein, ein Ihnen völlig fremdes System zu steuern. Das System wird dargestellt durch vier Variablen, zwei zum direkten Eingreifen, zwei zum Beobachten.

Sobald Sie dieses Informationsfenster schließen, können Sie bei zwei der Variablen Zahlen eingeben. Durch Ihre Eingabe setzen Sie jeweils etwas hoch oder herunter. Klicken Sie dann auf 'weiter'. Es erscheinen neue Werte, die die Veränderung des Systems in den oberen Variablen (A,B) anzeigen.

Die Veränderungen treten immer ein, nachdem Sie den 'weiter'-Button gedrückt haben. Damit ist jeweils ein Takt zu ende. Neben positiven Zahlen können auch negative Zahlen und die Null eingegeben werden. Keine Eingabe entspricht auch der Null. Ihre Eingabe (bei I,II) wirkt sich übrigens nur im direkt folgenden Takt aus. Ihnen werden nun drei Durchgänge mit sieben solcher Takte zum Testen gegeben. Im vierten Durchgang sollen Sie dann das System steuern können. Genaue Ziele werden Ihnen dann genannt.

Bei jedem neuen Durchgang handelt es sich um ein exakt identisches System, das auf ihre Eingaben reagiert. Ein Durchgang ist natürlich unabhängig vom vorhergehenden.

Wichtig !!!

Ihre Aufgabe ist es in drei Durchgängen möglichst viel Wissen über das System zu sammeln, da Sie im vierten Durchgang nach bestimmten Anweisungen das System steuern sollen.

Am Ende eines Durchgangs folgt jeweils ein Wissenstest über die Zusammenhänge des Systems, bevor ein neuer Durchgang beginnt.

Tip:

Es ist nicht möglich, das System im ersten Takt bereits zu durchschauen. Sie können sich Notizen machen, nötig sollte es jedoch nicht sein.

Achten Sie auch auf die Wirkungen der Variablen a und b untereinander.

Die Variablen I,II wirken sich nicht aufeinander aus.

/de/instructions/group0/text1.txt

Text_2 für die unspezifische Kontrollgruppe (KG1) und Text_3 und Text_4 für alle außer EG2

Dieser Probedurchgang ist nun zu Ende.

Es folgt ein weiterer Durchgang zum Üben.

/de/instructions/group1/text1.txt

Text_2 für die taktische Kontrollgruppe (KG2)

Es folgen die Probedurchgänge zum Üben und Erkunden.

Hier noch einige Tips und Tricks, die Ihnen helfen können.

- 1.Hypothesen: Bilden Sie sich immer wieder Hypothesen, was passieren wird, wenn diese oder jene Variable von Ihnen verändert wird.
- 2.Einzeleingriffe: Verändern Sie nicht immer alle Variablen gleichzeitig, sondern lassen Sie auch mal eine Variable leer.
- 3.Nulleingriffe: Um Dynamiken zu erkennen sollten Sie auch mal nichts verändern.
- 4.Starke Eingriffe: Um Beziehungen eindeutig zu erkennen nützen hohe Eingabewerte.
- 5.Negative Ausgabevariablen: Steuern Sie die Ausgabevariablen nacheinander auf Null oder in den negativen bzw. positiven Bereich.

/de/instructions/group3/text1.txt

Text_2 für die Experimentalgruppen (EG1, EG2)

Damit Sie nicht völlig ohne Variablenbezeichnungen arbeiten müssen, stellen Sie sich bitte das folgende System als 'kleine Mensa' vor, auf die Sie Einfluß nehmen können. Dabei spielen vier Variablen eine Rolle:

(alphabetisch sortiert)

Besucher,
Preis,
Werbung und
Zufriedenheit.

Ordnen Sie, während Ihrer Tests, die Variablenbezeichnungen einigermaßen sinnvoll dem System zu. Diese Variablenbezeichnungen sollen eine Hilfe für Sie sein, schneller das System zu erkennen und später besser steuern zu können. Setzen Sie sie ein, wie ein Hilfswerkzeug, das Sie ständig verändern können.

Die Zuordnungen bleiben natürlich immer noch sehr abstrakt, da eine Mensa sicherlich aus mehr Variablen besteht, dennoch können Sie für sich daraus Informationen gewinnen.

Testen Sie Ihr System und verändern Sie die Bezeichnungen, bis es für Sie ein einigermaßen sinnvolles System ergibt, was Ihnen beim Steuern helfen kann und Zusammenhänge erkennen läßt.

/de/instructions/group3/text2.txt

Text_3 für die EG2

Auch im nächsten Durchgang handelt es sich um das gleiche System, jedoch soll Ihre Aufgabe jetzt sein, sich eine andere Analogie mit diesem System vorzustellen. Bitte ordnen Sie die folgenden Variablenbezeichnungen eines 'kleinen Ökosystems' zu:

(alphabetisch sortiert)

kleine Käfer,
große Käfer,
Licht und
Wärme.

Dabei gibt es kein Richtig oder Falsch, sondern nur eine Vorstellungshilfe für Sie. Zusammenhänge des Systems bleiben genau wie in Durchgang eins.

/de/instructions/group3/text3.txt

Text_3 für die EG2

Auch im nächsten Durchgang handelt es sich noch mal um das gleiche System, jedoch soll Ihre Aufgabe jetzt sein, sich eine andere Analogie mit diesem System vorzustellen. Bitte ordnen Sie die folgenden Variablenbezeichnungen eines 'kleinen Abfallsystems' zu:

(alphabetisch sortiert)

Anzahl der Entleerungen,
Gestank aus den Tonnen,
Füllstand der Tonnen und
Menge des Abfalls.

/de/instructions/twintest.txt

Text für die Wissensdiagnostik

Wissenstest über die Struktur des bearbeiteten Systems

- Dieses Fenster besteht aus drei Teilen, die Sie über die Button der obigen Leiste erreichen können. Diese Leiste ist in allen Teilen zu sehen:

'Erklärung' zeigt diese Erklärung an, also wie der Test zu bedienen ist.

'Testfenster' zeigt das Testfenster an, in dem Sie Ihr Wissen unter Beweis stellen sollen.

'System' zeigt nur noch die Buttonleiste, so daß Sie (auch bei kleinem Monitor) den vorhergehenden Durchgang vollständig sehen können.

In diesem Test sollen Sie ihr Wissen über das System darlegen. Dazu dient das Fenster, das erscheint, wenn Sie den Button 'Testfenster' drücken.

Zunächst ein Beispiel:

Angenommen Sie haben einen negativen Zusammenhang zwischen Variable II und B festgestellt, d.h. wenn Sie bei II eine positive Zahl (z.B. 50) eingegeben haben, sank der Wert bei B (z.B. um 200).

Dann müssten Sie in der untersten Zeile (= Variable II ; Zusammenhang zu...B) unter Richtung ein '-' auswählen, da ja ein negativer und kein positiver Zusammenhang besteht. Danach könnten Sie sogar noch die Stärke dieses Zusammenhangs in das Textfeld unter Stärke eingeben, so Sie diese kennen. Im Beispiel wäre das '4' bzw. '4.0'.

Schritt für Schritt:

1. Wenn Sie zwischen den Variablen (aus Spalte 'Variable' und 'Zusammenhang_zu') einen Zusammenhang erkannt haben oder sich sicher sind, daß es keinen Zusammenhang gibt, dann wählen Sie in der Spalte 'Richtung' etwas aus dem betreffenden Auswahlfeld aus:

'+' für einen positiven Zusammenhang der Variablen (ist die eine positiv, steigt die andere bzw. ist die eine negativ, sinkt die andere),

'-' für einen negativen Zusammenhang der Variablen (ist die eine positiv, sinkt die andere bzw. ist die eine negativ, sinkt die andere),

'+' oder '-' wenn Sie wissen es gibt einen Zusammenhang, aber nicht welche Richtung dieser hat,

'keinen' wenn Sie sicher sind, daß diese Variablen keinen direkten Zusammenhang haben und

'- ? -' wenn Sie nicht wissen, ob ein Zusammenhang besteht.

2. Haben Sie eine Richtung durch '+' oder '-' angegeben, können Sie die Stärke der Verbindung angeben, so Sie eine Vorstellung davon haben. Geben Sie dazu eine positive Zahl in das nebenstehende Textfeld ein. Für Dezimalzahlen benutzen Sie bitte den Punkt '.' und bitte nicht das Komma ','. Eine Angabe von 5 beispielsweise würde bedeuten, dass die Variable unter 'Zusammenhang' um das 5fache von der Variable in dieser Zeile steigt (+) oder fällt (-).

Dieser Wissenstest wird Ihnen nach jedem Durchgang der Simulation gestellt.

Viel Erfolg

/de/instructions/goal.txt

Zieltext (Text_5) für alle Gruppen (KG1, KG2, EG1, EG2)

Es folgt der entscheidende Durchgang. Sie haben dabei gleich zwei Aufgaben, zum einen sollen Sie versuchen die beiden nachfolgenden Regeln so gut es geht einzuhalten, zum anderen sollen Sie die unten angegebenen Ziele für A und B möglichst schnell und genau erreichen.

Regel 1: A soll beständig kleiner werden oder zumindest gleich bleiben.

Regel 2: B soll niemals kleiner werden und möglichst konstant bleiben.

Bitte erreichen Sie die nachfolgenden Ziele möglichst früh und beachten dabei die obigen Regeln.

/de/instructions/concludingquest.txt

Abschlußtext für alle Gruppen

Abschließender Kommentar

Bitte geben Sie Ihre email-Adresse an, wenn Sie am Gewinnspiel teilnehmen und über Ergebnisse der Untersuchung informiert werden möchten.

Darunter besteht die Möglichkeit eines abschließenden Kommentars zum Experiment, z.B. über mögliche Schwierigkeiten, Verbesserungsvorschläge etc.

Möchten Sie keine Informationen erhalten oder keinen Kommentar abgeben, drücken Sie einfach 'OK' und lassen Sie die Felder leer.

Diese Angaben haben nichts mit dem laufenden Experiment zu tun und werden anonym behandelt.

Erklärung

Hiermit erkläre ich, dass ich diese Arbeit selbständig und ohne Benutzung von anderen als den angegebenen Quellen und Hilfsmitteln angefertigt habe. Alle den Quellen wörtlich und inhaltlich entnommenen Stellen wurden als solche kenntlich gemacht.

Heidelberg, 11. August 2003

Carsten A. Schulz

Freiwillige Erklärung

Ich stimme ausdrücklich zu, dass meine durch Prof. Dr. Joachim Funke und Prof. Dr. Karlheinz Sonntag betreute Diplomarbeit mit dem Titel:

Strategien zum komplexen Problemlösens.

Eine experimentelle Analyse der Rolle von Analogien beim Wissenserwerb in komplexen, dynamischen Systemen.

nach Beendigung der Diplomprüfung wissenschaftlichen Zwecken zugänglich gemacht und in der Institutsbibliothek aufgestellt wird (Veröffentlichungen nach § 6 Abs. 1 UrhG), sowie hieraus im Rahmen des § 51 UrhG zitiert werden kann.

Sämtliche Verwertungsrechte nach § 15 UrhG verbleiben beim Verfasser der Diplomarbeit.

Heidelberg, 11. August 2003

Carsten A. Schulz

Carsten A. Schulz

Punkerstr. 8

69126 Heidelberg

Der.Schulz@web.de