Friederike Brockhaus

Von den Flussgrößen zum Bestand: Lassen sich einfache dynamische Systeme intuitiv verständlich darstellen?

Friederike Brockhaus

Von den Flussgrößen zum Bestand:

Lassen sich einfache dynamische Systeme intuitiv verständlich darstellen?

TECHNISCHE UNIVERSITÄT
CHEMNITZ

Universitätsverlag Chemnitz
2018

Impressum

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Angaben sind im Internet über http://www.dnb.de abrufbar.

Das Werk - ausgenommen Zitate, Cover, Logo TU Chemnitz und Bildmaterial im Text - steht unter der Creative-Commons-Lizenz Namensnennung 4.0 International (CC BY 4.0)
https://creativecommons.org/licences/by/4.0/deed.de

Titelgrafik: Friederike Brockhaus
unter Verwendung eines Piktogramms von
Guadalupe Peralta-Ramos
Satz/Layout: Friederike Brockhaus

Technische Universität Chemnitz/Universitätsbibliothek
Universitätsverlag Chemnitz
09107 Chemnitz
https://www.tu-chemnitz.de/ub/univerlag

readbox unipress
in der readbox publishing GmbH
Am Hawerkamp 31
48155 Münster
http://unipress.readbox.net

ISBN 978-3-96100-074-6

http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa2-322531

Von den Flussgrößen zum Bestand:

Lassen sich einfache dynamische Systeme intuitiv verständlich darstellen?

Dissertation zur Erlangung des
akademischen Grades
Doctor rerum naturalium (Dr.rer.nat.)

Vorgelegt an der Fakultät für Human- und
Sozialwissenschaften
der Technischen Universität Chemnitz
von
Friederike Brockhaus

Tag der Einreichung: 29.09.2017
Tag der Verteidigung: 20.12.2017

1. Gutachter: Prof. Dr. Peter Sedlmeier
2. Gutachter: Prof. Dr. Josef F. Krems

Zusammenfassung

Die vorliegende Dissertation handelt von einfachen dynamischen Systemen zu Bestands- und Flussgrößen (BF-Systeme). Diese bestehen aus zwei relevanten Größen: aus dem Bestand und den Flussgrößen, bei denen es sich um Zu- und Abfluss handelt. Ein Beispiel hierfür ist ein Wasserbecken, in das Wasser fließt (Zufluss) und aus dem Wasser herausfließt (Abfluss). Der Wasserstand (Bestand) verändert sich über die Zeit hinweg in Abhängigkeit der beiden Flussgrößen. BF-Systeme sind in unserem Leben omnipräsent: Zum einen stellen sie Subsysteme komplexer dynamischer Systeme dar, wie die Weltwirtschaft oder das Klima der Erde. Zum anderen sind wir mit BF-Systemen für sich genommen in unserem Alltag konfrontiert, beispielsweise mit unserem Bankkonto. In vielen Studien erhielten Versuchspersonen die Flussgrößen für einen bestimmten Zeitraum als Liniendiagramm dargeboten. Ihre Aufgabe war es, zu beurteilen, wie sich der Bestand über die Zeit hinweg veränderte, was den meisten große Probleme bereitete.

Es ist nicht ungewöhnlich, dass es Versuchspersonen schwer fällt beim Urteilen oder Problemlösen zu adäquaten Lösungen zu kommen. Jedoch zeigte sich oft, dass eine Änderung des Darstellungsformats die Leistung der Versuchspersonen stark verbessert.

Beispielsweise ist das der Fall im Bereich der Verarbeitung von Wahrscheinlichkeiten oder beim logischen Schließen. Grundlegend können durch günstige Modifikationen valide Intuitionen angesprochen werden, was zu besserer Leistung in den Urteilen oder beim Problemlösen führt. Ein entsprechender Ansatz wurde im Rahmen der vorliegenden Arbeit für die BF-Systeme gewählt: In vier Studien sollten über Modifikationen der Aufgaben-Darstellung adäquate Intuitionen gefördert werden, die das BF-Verständnis steigern. In den vier Studien dieser Arbeit lag der Schwerpunkt auf der Darstellung der Flussgrößen in einem Diagramm, da dies ein zentraler Bestandteil der Aufgabenbeschreibung von BF-Systemen darstellt.

Die erste Studie untersuchte, inwiefern Änderungen unterschiedlicher Beschreibungsdimensionen der BF-Aufgaben die BF-Leistung beeinflussten. Eine zusätzliche gegebene Abbildung des BF-Systems (beispielsweise die schematische Abbildung eines Wasserbeckens) führte zu besserer Leistung: Die Anteile der korrekten Lösungen nahmen zu. Eine Kombination aus Text und Flussdiagramm verschlechterte die Leistung im Gegensatz zu einer Bedingung nur mit Text. Zwei verschiedene Darstellungsformate der BF-Größe (beispielsweise Wasser in einem Wasserbecken) wurden ebenfalls untersucht, wobei sich diesbezüglich kein Unterschied in der BF-Leistung ergab. Zudem wurden die Fehler

genauer betrachtet; auch in der vorliegenden Studie nahmen viele Personen an, dass der Verlauf des Bestandes dem Verlauf einer Flussgröße entspricht. Es handelt sich dabei um die sogenannte Korrelations-Heuristik. Basierend auf diesen Ergebnissen erfolgten weitere Modifikationen in der zweiten Studie. Zum einen wurden Zu- und Abfluss getrennt im Koordinatensystem des Flussdiagramms abgetragen. In einer anderen Änderung der Aufgabenbeschreibung erhielten die Versuchspersonen den Nettofluss als zusätzliche Größe. Diese Modifikationen verringerten jedoch das BF-Verständnis.

Somit untersuchte die dritte Studie eine etwas stärkere Abwandlung des Flussdiagramms: Die Flussgrößen sahen aus wie Bestandsgrößen für einzelne Zeitpunkte. Um das umzusetzen, wurden die Flussgrößen über Piktogramme dargestellt. In diesen sogenannten symbolischen Bedingungen ergab sich eine starke Verbesserung der BF-Leistung. Jedoch fand keinerlei Transfer-Leistung statt, wenn die Versuchspersonen in einem zweiten Durchgang BF-Aufgaben in klassischem Darstellungsformat, in Form von Liniendiagrammen, bearbeiteten. Die Verbesserung der Leistung resultierte vermutlich aus dem Format in den symbolischen Bedingungen. Weiterhin gingen die Versuchspersonen in diesen Bedingungen schrittweise vor, was ebenfalls das korrekte Lösen der BF-Aufgaben zu unterstützen schien.

Die vierte Studie untersuchte, ob eine ähnlich gute Verbesserung in der BF-Leistung auch auftrat, wenn kein schrittweises Vorgehen notwendig war, um zur Lösung zu gelangen. Das symbolische Darstellungsformat wurde außerdem weiterentwickelt, indem Zufluss und Abfluss auf Piktogrammebene unterschiedlich gestaltet waren: Sie wurden über sogenannte ikonische Zeichen dargestellt. Weiterhin wurde die Differenz von Zu- und Abfluss betont, so dass gleich ersichtlich war, welche Flussgröße um wie viele Einheiten überwog. Tatsächlich schnitten die Versuchspersonen der symbolischen Bedingungen besser ab als die in der Baseline-Bedingung mit Liniendiagrammen. Jedoch war der Unterschied in der Leistung nicht mehr so groß wie in der vorangegangenen Studie. Zudem unterschied sich die Leistung zwischen den verschiedenen symbolischen Bedingungen kaum.

Eine abschließende Diskussion fasst die Ergebnisse der einzelnen Studien zusammen: Es erfolgt ein Überblick über den Einfluss der Modifikationen in der Darstellung von BF-Systemen. Außerdem werden weitere Ergebnisse zu Personenmerkmalen und die Fehleranalysen thematisiert. Abschließend folgen Überlegungen zum Umgang mit BF-Systemen.

Inhaltsverzeichnis

Zusammenfassung.....VII
1 Einleitung.....1
1.1 Systeme zu Bestands- und Flussgrößen.....3
1.1.1 BF-Systeme im Zusammenhang mit komplexen dynamischen Systemen.....4
1.1.2 Beispiele von BF-Systemen.....10
1.1.3 Analyse des Verständnisses bei BF-Aufgaben.....14
1.1.4 Bisherige Versuche zur Erhöhung des BF-Verständnisses.....20
1.1.5 Beschreibungsdimensionen für BF-Systeme.....23
1.1.6 Auffälligkeit bezüglich der BF-Lösungen.....31
1.1.7 Fazit.....33
1.2 Intuition.....34
1.2.1 BF-Systeme und Intuition.....35
1.2.2 Mathematische Intuition.....40
1.2.3 Die externale Repräsentation beim Problemlösen.....44
1.2.4 Diagramme.....48
1.2.5 Fazit.....52
2 Der Einfluss verschiedener Beschreibungsdimensionen auf das BF-Verständnis – Studie 1.....53
2.1 Einleitung.....53
2.1.1 Der Einfluss von Bildern.....54
2.1.1.1 Schematische Abbildung des BF-Szenarios.....55
2.1.1.2 Darstellungsformat der Flussdiagramms.....56
2.1.2 Der Einfluss des genutzten Formats für das BF-Szenario.....57
2.1.1 Fehleranalyse.....60
2.2 Methoden.....61
2.2.1 Teilnehmende.....61
2.2.2 Materialien: die verwendeten Verläufe.....62
2.2.2.1 Bedingung „schematische Abbildung des Szenarios“.....62

2.2.2.2 Bedingungen „Flussdiagramm“ und „BF-Szenario“....64
2.2.3 Abhängige Variable....66
2.2.4 Ablauf....70
2.1.2 Vorgehen bei der Analyse....71
2.3 Ergebnisse....75
2.3.1 Überblick: Anteile korrekter Lösungen und Regressionsanalyse....76
2.3.2 Bedingung "Schematische Abbildung des Szenarios"81
2.3.3 Bedingung "Flussdiagramm"....82
2.3.4 Bedingung "BF-Szenario"....84
2.3.5 Fehleranalyse....85
2.3.6 Zusätzliche Analysen: Positionseffekte in der Bedingung „BF-Szenario“....88
2.4 Diskussion....89
2.4.1 Bedingung "schematische Abbildung des Szenarios" 92
2.4.2 Bedingung "Flussdiagramm"....94
2.4.3 Bedingung "BF-Szenario"....97
2.4.4 Fehleranalyse....100
2.4.5 Fazit....103
3 Modifikation des Flussdiagramms und Einbeziehung des Nettoflusses – Studie 2....105
3.1 Einleitung....105
3.1.1 Modifikation des Flussdiagramms – Teilung des Koordinatensystems....107
3.1.2 Hinzufügen des Nettoflusses....109
3.1.3 Fehleranalyse....112
3.1.4 Kontrollvariablen....113
3.2 Methoden....113
3.2.1 Teilnehmende....113
3.2.2 Materialien: die verwendeten Verläufe....114
3.2.3 Modifikationen, um eine Erleichterung des BF-Verständnisses herbeizuführen....119
3.2.3.1 Die Bedingung „geteiltes Koordinatensystem”. 120

3.2.3.2 Die Bedingungen „Nettofluss-gegeben“ und „Nettofluss-ermitteln“.....121
3.2.4 Abhängige Variable.....122
3.2.4.1 Fehleranalyse.....124
3.2.4.2 Ermittelter Nettofluss und bedingte Wahrscheinlichkeiten.....124
3.2.5 Weitere Variablen.....124
3.2.6 Ablauf.....125
3.2.7 Vorgehen bei der Analyse.....126
3.3 Ergebnisse.....127
3.3.1 Überblick und Regressionsanalyse.....128
3.3.2 Bedingung „geteiltes Koordinatensystem“.....132
3.3.3 Die Bedingungen „NF-gegeben“ und „NF-ermitteln“ 136
3.3.3.1 Korrektheit des ermittelten Nettoflusses.....138
3.3.3.2 Bedingte Wahrscheinlichkeit bei NF-ermitteln..139
3.3.4 Fehleranalyse.....141
3.3.4.1 Weiteres systematisches Fehlermuster.....142
3.3.4.2 Überblick über das Anwenden der Korrelations-Heuristik.....144
3.3.4.3 Für Korrelations-Heuristik genutzte Flussgrößen 145
3.3.4.4 Fazit.....147
3.3.5 Kontrollvariablen.....147
3.3.5.1 Geschlecht.....147
3.3.5.2 Studienfach.....148
3.3.5.3 Teilnahme an vorhergehender Studie.....149
3.3.6 Zusätzliche Analysen.....150
3.3.6.1 Aufgabenschwierigkeit.....150
3.3.6.2 Positionseffekte.....151
3.4 Diskussion.....151
3.4.1 Bedingung geteiltes Koordinatensystem.....152
3.4.1.1 Missinterpretationen bei der neuen Darstellung? 152
3.4.1.2 Probleme im Abgleich der Werte?.....153

3.4.1.3 Manipulation „KS-geteilt“ verschlechterte die BF-Leistung....156
3.4.2 Die Nettofluss-Bedingungen....157
3.4.2.1 Verstanden die Versuchspersonen das Prinzip Nettofluss?....158
3.4.2.2 Manipulationen mit Nettofluss verschlechterten die BF-Leistung....160
3.4.3 Fehleranalyse....161
3.4.4 Einfluss der Kontrollvariablen....162
3.4.5 Zusätzliche Analysen....162
3.4.6 Fazit und Ausblick....163
4 Die Darstellung von Flussgrößen als sukzessive Bestandsgrößen – Studie 3....165
4.1 Einleitung....165
4.2 Methoden....173
4.2.1 Teilnehmende....173
4.2.2 Materialien: die verwendeten Verläufe....174
4.2.3 Modifikationen des Darstellungsformats....178
4.2.3.1 Baseline-Bedingung....179
4.2.3.2 Wasser-Bedingung....180
4.2.3.3 Bus-Bedingung....183
4.2.4 Abhängige Variablen....185
4.2.5 Transfer-Aufgaben....186
4.2.6 Weitere Variablen....190
4.2.7 Ablauf....190
4.3 Ergebnisse....191
4.3.1 Überblick....191
4.3.2 Einfluss des Darstellungsformats....191
4.3.2.1 Regressionsanalyse....195
4.3.2.2 Vergleich von Wasser- und Bus-Bedingung....197
4.3.3 Fehleranalyse....198
4.3.4 Kontrollvariablen Geschlecht und Studienfach....203
4.3.4.1 Geschlecht....203
4.3.4.2 Studienfach....204
4.3.4.3 Interaktionsterme....206

4.3.5 Transfer-Leistung.....208
4.3.6 Zusätzliche Analysen.....219
4.3.6.1 Positionseffekt – bei Durchgang1-Aufgaben.....219
4.3.6.2 Positionseffekt – bei Transfer-Aufgaben.....220
4.4 Diskussion.....220
4.4.1 Das symbolische Darstellungsformat.....220
4.4.1.1 Vergleich Baseline-Bedingung und symbolische Bedingungen.....223
4.4.1.2 Vergleich der beiden symbolischen Bedingungen 223
4.4.2 Fehleranalyse.....225
4.4.3 Kontrollvariablen Geschlecht und Studienfach.....228
4.4.4 Transfer-Leistung.....229
4.4.5 Zusätzliche Analysen.....231
4.4.5.1 Positionseffekte.....231
4.4.5.2 Aufgabenschwierigkeit.....232
4.4.6 Mögliche Wirkfaktoren.....233
4.4.7 Fazit.....239
5 Erweiterung der symbolischen Darstellung – Studie 4.....241
5.1 Einleitung.....241
5.1.1 Hypothesen.....244
5.1.2 Materialien: die verwendeten Verläufe.....250
5.2 Methoden.....252
5.2.1 Teilnehmende.....252
5.2.2 Materialien.....253
5.2.3 Modifikationen des Darstellungsformats.....256
5.2.3.1 Baseline-Bedingung.....257
5.2.3.2 Bedingung „Piktogramm-Menschen, ohne Differenz“ (Pikt-oDiff).....257
5.2.3.3 Bedingung „ikonische Symbole ohne Differenz“ (ikon-oDiff).....259
5.2.3.4 Bedingung „Piktogramm-Menschen, mit Differenz“ (Pikt-mitDiff).....261
5.2.3.5 Bedingung „ikonische Symbole mit Differenz“ (ikon-mitDiff).....262

5.2.4 Abhängige Variablen und Zusatzvariablen.....264
5.2.5 Ablauf.....266
5.3 Ergebnisse.....268
5.3.1 Überblick.....268
5.3.2 Einfluss der Aufgaben-Darstellung.....270
5.3.3 Fehleranalyse.....274
5.3.4 Analysen zum Einfluss von Kontrollvariablen.....284
5.3.4.1 Geschlecht.....284
5.3.4.2 Studienfach.....286
5.3.4.3 Mathematiknote.....286
5.3.4.4 Geschlechterrolle.....287
5.3.4.5 Vorherige Teilnahme.....287
5.3.5 Zusätzliche Analysen.....288
5.3.5.1 Positionseffekte.....288
5.3.5.2 Aufgabenschwierigkeit.....288
5.3.5.3 Fragenkomplex zu den Flussgrößen.....289
5.3.5.4 Abbruchquote.....291
5.4 Diskussion.....292
5.4.1 Einfluss der Aufgaben-Darstellung.....292
5.4.2 Unterschied piktogrammartige und ikonische Darstellung.....292
5.4.3 Betonung des Nettoflusses.....294
5.4.4 Fehleranalyse.....296
5.4.5 Analysen zum Einfluss von Geschlecht, Studienfach und Mathematiknote.....299
5.4.5.1 Geschlecht.....299
5.4.5.2 Studienfach.....300
5.4.5.3 Mathematiknote.....301
5.4.6 Analysen zur Geschlechterrolle.....301
5.4.7 Ablesen der Flussgrößen.....302
5.4.8 Vorherige Teilnahme, Abbruchquote und fehlende Werte.....304
5.4.8.1 Vorherige Teilnahme.....304
5.4.8.2 Abbruchquote.....304
5.4.8.3 Fehlende Werte.....308

5.4.9 Fazit.....309
6 Diskussion der gesamten vorliegenden Arbeit.....311
6.1 Einfluss externaler Repräsentation auf BF-Leistung – eine Zusammenfassung.....311
6.1.1 Studie 1.....312
6.1.2 Studien 2 bis 4.....314
6.2 Fehleranalyse.....317
6.3 Personenmerkmale.....321
6.3.1 Geschlecht.....321
6.3.2 Mathematisches Verständnis.....322
6.3.3 Geschlecht, Geschlechterrolle und mathematisches Verständnis.....323
6.3.4 Fazit.....324
6.4 Intuition.....325
6.5 Implikation für die Kommunikation von BF-Systemen. 329
6.6 Fazit.....331
Referenzen.....333
Anhang.....347
Anhang A.....347
Anhang B.....349
Anhang C.....351
Anhang D.....352
Anhang E.....353
Anhang F.....354
Anhang G.....355
Anhang H.....356

1 Einleitung

Anfang der 1970er Jahre veröffentlichte eine Forschungsgruppe in den USA ein aufsehenerregendes Buch. Der Club of Rome hatte ForscherInnen des Massachusetts Institute of Technology (MIT) beauftragt zu analysieren, welche Gründe und Folgen verschiedene Entwicklungen haben, die auf die Menschheit und ihre Handlungen zurückgehen. Relevant für ihre Analysen waren beispielsweise das Bevölkerungswachstum, die Nahrungsmittelproduktion, der Rohstoffverbrauch, das Industriekapital und die Umweltverschmutzung. Die Forschungsgruppe simulierte ein „Weltmodell“. Das Buch „Die Grenzen des Wachstums“ fasste die daraus resultierenden Ergebnisse zusammen (Meadows, Meadows, Zahn, & Milling, 1972). Darin wurde beispielsweise vorhergesagt: Wenn Industrialisierung und Bevölkerungswachstum, Umweltverschmutzung und weitere damit zusammenhängende Phänomene (wie bspw. Ausbeutung von natürlichen Rohstoffen oder die Nahrungsmittelproduktion) weiterhin so anstiegen wie bis zu diesem Zeitpunkt, könnten die absoluten Wachstumsgrenzen innerhalb des folgenden Jahrhunderts erreicht werden. Die Forschungsgruppe betonte jedoch, dass es möglich erschiene, die beschriebenen Wachstumstrends abzuändern. Diese Änderung

könnte zu einem Zustand von Gleichgewicht führen, aus dem eine langfristige Stabilisierung resultieren könne. Es folgten Jahre später weitere, angepasste Simulationen, die grundlegend die Botschaft enthielten, dass einerseits noch nicht genug Gegenmaßnahmen ergriffen wurden, um einen möglichen „Kollaps“ zu umgehen, dass es andererseits Möglichkeiten gibt, Stabilität herbeizuführen und eine nachhaltige Gesellschaft zu schaffen (Meadows, Meadows, & Randers, 1992).

Was nun im Detail von diesen Simulationen zu halten ist, kann jedeR für sich entscheiden. Was sich jedoch zeigt, ist die hohe Relevanz, derartige Zusammenhänge und Abläufe verständlich darzustellen. Nur dann sind Personen in der Lage, ein realistisches Urteil zu bilden und die in den Berichten zu den Simulationen beschriebenen Sachverhalte danach zu beurteilen, inwiefern sie ihnen zustimmen, und inwieweit sie die Lösungsvorschläge akzeptieren. Denn viele Probleme (das Ausschöpfen natürlicher Ressourcen, Umweltverschmutzung usw.) sind schwer verständlich für die meisten Menschen. Beispielsweise fällt es Personen sehr schwer, exponentielles Wachstum einzuschätzen, wie im Fall von exponentiell wachsender Luftverschmutzung: Der Anstieg wird meist erheblich unterschätzt (z.B. Wagenaar & Timmers, 1979). Somit sollte schwer erkennbar sein, welche Probleme sich beispielsweise im Zusammenhang mit Armut ergeben, wenn die

Bevölkerung der Erde oder einzelner Regionen derart rasant wächst (Meadows et al., 1992).

1.1 Systeme zu Bestands- und Flussgrößen

Die vorliegende Arbeit konzentriert sich auf eine Form von Systemen, die ähnlich wie exponentielles Wachstum schwer zu verstehen ist. Es handelt sich um „Stock and Flow“-Systeme oder im Deutschen Systeme zu Bestands- und Flussgrößen (im Folgenden „BF-Systeme“). In diesen Systemen kann exponentielles Wachstum auftreten, jedoch hat sich gezeigt, dass BF-Systeme auch ohne exponentiell steigende Werte schwer verständlich sind (z.B. Booth Sweeney & Sterman, 2000; Cronin, Gonzalez & Sterman, 2009; Ossimitz, 2002).

Ein BF-System besteht aus einer Bestandsgröße und aus Flussgrößen: dem Zufluss und Abfluss. Ein BF-System ist beispielsweise ein Becken, das Wasser enthält (die Bestandsgröße). Die Flussgrößen sind Wasser, das hineinfließt (der Zufluss) und Wasser, das herausfließt (der Abfluss). Bei vielen der für die Weltsimulation relevanten Größen handelt es sich um Bestandsgrößen, beispielsweise die Bevölkerung, die Umweltverschmutzung (im Sinne von Schadstoffmengen) oder verschiedene Formen von Kapital (Industrie-, Dienstleistungskapital, usw.). Auch

andere, heutzutage aktuelle Themen stellen BF-Systeme dar, wie der CO_2-Gehalt der Atmosphäre oder die Staatsverschuldung. Darauf gehe ich genauer im folgenden Abschnitt ein, in dem ich BF-Systeme kurz in den größeren Gegenstandsbereich „dynamische Systeme“ einordne. In Abschnitt 1.1.2 sind BF-Systeme dann genauer beschrieben. Anschließend zeigen Ergebnisse aus Studien das schlechte Verständnis dieser Systeme.

1.1.1 BF-Systeme im Zusammenhang mit komplexen dynamischen Systemen

Für die Einordnung von BF-Systemen in eine Forschungstradition passt „Systemdynamik“ am besten. In den 1950er Jahren schuf Jay Wright Forrester die „System Dynamics Group“ am MIT [Homepage des MIT] (o.D.). Es ging um die Auseinandersetzung mit komplexen dynamischen Systemen, welche sich beispielsweise durch Eigenschaften wie Zeitverzögerung, nicht-lineares Verhalten von Variablen oder Rückkopplungseffekte auszeichnen. Zudem können Systemteile sich gegenseitig beeinflussen (Gonzalez, Vanyukov & Martin, 2005; Forrester, 1968; 2003; Sterman, 1989, 1994). Menschliche oder natürliche Systeme zeigen oft eine hohe dynamische Komplexität (Sterman, 1994). Beispiele sind die Weltwirtschaft oder das Weltklima. Komponenten davon wurden eingangs beschrieben (z.B. nicht-li-

neares Wachstum der Bevölkerung). Die System Dynamics Group arbeitete daran, komplexe dynamische Systeme zu analysieren und zu simulieren. Ziel war es, dass Personen sie verstehen und fähig sind, sich in ihnen korrekt zu verhalten (Forrester, 1968). Auch „Die Grenzen des Wachstums“ geht auf Forschung aus diesem Fachbereich zurück (Forrester, 1989). Jahre später betonte Sterman (1994), wie wichtig es ist, das schwierige Thema dynamische komplexe Systeme zu verstehen und „System-Denken“ zu entwickeln. Es geht darum, dass Personen ein ganzheitliches Weltbild entwickeln, dass sie dynamische Systeme verstehen und sie gegebenenfalls adäquat beeinflussen können.

Etwa zeitgleich mit dem Erscheinen von „Die Grenzen des Wachstums“ untersuchte in Deutschland Dörner als einer der ersten derartige Systeme (vgl. bspw. Funke, 2003). Hierbei schlüpften die Versuchspersonen unter anderem in die Rolle eines Bürgermeisters (Dörner, 1981) oder waren Entwicklungshelfer (Dörner, 1991). Es handelte sich um Computersimulationen mit vielen miteinander vernetzten Variablen, in denen der Umgang mit komplexen Systemen getestet wurde. Dörners (1999) Meinung nach ist diese Art von Problemlösen eine sehr wichtige Ergänzung zur „klassischen Problemlöseforschung“, bei der der Schwerpunkt auf einzelnen „Denksportaufgaben“ ohne Einbettung in die Realität liegt. Sein Fazit war, dass Menschen schlecht

abschneiden, wenn sie komplexe Systeme verstehen oder steuern wollen, sei es in Computersimulationen oder im wahren Leben. Ungefähr parallel dazu arbeitete Broadbent zu verschiedenen Gedächtnissystemen (Funke, 2003). Er untersuchte eher einfachere dynamische Probleme in experimentellen Settings wie beispielsweise ein „Transport-System“ (Broadbent, 1977). Diese beiden Forschungsstränge stellen die europäischen Wurzeln im Forschungsbereich zu komplexen Problemen dar (Buchner, 1995). Im weiteren Verlauf konzentrierte sich Funke (bspw. 2003) auf komplexe Probleme und den Umgang mit ihnen. Diese Probleme zeichneten sich beispielsweise durch Vernetzung von relevanten Variablen aus, durch Dynamik, auch der Faktor Zeit war mit einbezogen.

In Studien zeigte sich immer wieder, dass komplexe dynamische Systeme nicht durchgängig verstanden werden (Broadbent, 1977; Dörner, 1999; Funke, 2003; Sterman, 1994). Im Gegenteil, ihr Verständnis bereitet eher große Probleme. Sterman (1994) fasste es so zusammen: „[In complex systems,] performance is far from optimal – often far from reasonable – in a wide range of tasks“ (S.32). Nach Dörner (1981) ist das Unvermögen vieler Personen, mit der Komplexität umzugehen, eine Ursache für menschengemachte Katastrophen und Krisen.

Die Frage stellt sich also, woher dieses Unvermögen kommt, das sich in vielen Studien zeigte. Bei vielen AutorInnen lautete die Annahme, dass die Komplexität die kognitiven Fähigkeiten übersteigt und somit die Probleme verursacht (z.B. Brehmer, 1992; Dörner, 1981; Dörner, 2008). Andere AutorInnen betonten, dass schon die einzelnen Bausteine oder Subsysteme von komplexen dynamischen Systemen schwer zu verstehen sind (Booth Sweeney & Sterman, 2000). BF-Systeme sind derartige Bausteine. Beispielsweise kann das oben genannte Wasserbecken ein Kühlbecken darstellen, welches kombiniert mit vielen weiteren Subsystemen das komplexe dynamische System „Kraftwerk“ ergibt. Das Kraftwerk wiederum ist ein Teil der Industrie in einem bestimmten Land. Derartige Systeme sind überall vorhanden und stellen wichtige Bestandteile in unserem Leben dar (z.B. das Unternehmen, in dem wir arbeiten, unsere Familie oder die Umwelt). Booth Sweeney und Sterman wandten sich also den BF-Systemen zu und untersuchten das Verständnis dieser Systeme detailliert und für sich genommen. Tatsächlich fanden sie kein gutes Verständnis für diese Systeme. Wenn Personen die Bausteine oder Subsysteme von komplexen dynamischen Systemen nicht verstehen, so können sie die Systeme an sich ebenfalls nicht verstehen oder auf günstige Art lenken.

Weiterhin sind BF-Systeme nicht nur als Teil von komplexen dynamischen Systemen bedeutsam. Personen sind im Alltag häufig mit „isolierten“ BF-Systemen konfrontiert. Ein Bankkonto ist ein BF-System: das Geld auf dem Konto entspricht dem Bestand, der Lohn beispielsweise entspricht einer Zuflussgröße und Abbuchungen sind Abflussgrößen. Auch der CO_2-Gehalt in der Atmosphäre ist eine Bestandsgröße, die aufgrund von Zufluss (z.B. CO_2-Ausstoß durch den Menschen) und Abfluss (z.B. CO_2-Absorption der Meere) insgesamt zu- oder abnimmt. Diese beiden Beispiele veranschaulichen, dass auch BF-Systeme für sich genommen für die Menschheit wichtig zu verstehen sind. Tatsächlich zeigte sich immer wieder, dass BF-Systeme eben nicht leicht verständlich sind. Beispielsweise werden häufig – absichtlich oder unabsichtlich – Prozesse zu BF-Systemen falsch beschrieben. Ossimitz (2002) erläuterte ein Beispiel aus der Politik. Die Staatsverschuldung ist eine Bestandsgröße. Die Neuverschuldung ist eine weitere Größe, die sich ergibt, wenn die Differenz aus Ausgaben und Einnahmen des Staates berechnet wird. Das Verrechnen von Zu- und Abfluss ergibt den sogenannten Nettofluss. Im Beispiel von Ossimitz wurde die Neuverschuldung, also der Nettofluss halbiert. Vonseiten einiger Politiker wurde das berichtet, als ob sich die *Staatsverschuldung* an sich halbierte, obwohl diese weiterhin zunahm.

Wenn die Medien über Klimagipfel der UNO berichten, so fällt auf, dass sie in erster Linie den CO_2-Ausstoß thematisieren. Teilweise wirkt es, als ob eine Reduktion des CO_2-Ausstoßes die Lösung ist, den CO_2-Gehalt in der Atmosphäre zu verringern. Die CO_2-Absorption stellt die Abflussgröße dar. Sie wird meist vernachlässigt und somit die Tatsache, dass der CO_2-Bestand in der Atmosphäre nur abnimmt, wenn der CO_2-Ausstoß geringer ist als die Absorption. Auch Sterman (2002) betonte, dass viele Personen unterschätzen, wie stark der CO_2-Ausstoß tatsächlich reduziert werden müsste, damit sich der Gehalt in der Atmosphäre überhaupt stabilisiert. Weiterhin berichtete er, dass viele Personen den Abfluss, also die absorbierte Menge CO_2, gar nicht berücksichtigen (Sterman & Booth Sweeney, 2002).

Zusammenfassend lässt sich sagen, es ist sehr wichtig, BF-Systeme zu verstehen. Einerseits begegnen alle Personen ihnen häufig im eigenen Alltag. Andererseits sind sie relevante Größen, wenn es darum geht, bestimmte Maßnahmen seitens der Politik zu verstehen oder zu diskutieren. In einigen Fällen ist es notwendig, BF-Systeme gut zu verstehen, um überhaupt entscheiden zu können, ob man bestimmten Aussagen Glauben schenkt – sei es betreffend Umweltschutz oder politischem Handeln (wie Schuldenabbau). Somit ist es ebenfalls wichtig, gute Möglichkeiten zur Darstellung von BF-Szenarien zur Hand zu haben, so dass diese

Szenarien auch für Laien einfach verständlich sind, beispielsweise, wenn derartige Inhalte in den Medien berichtet werden. Ein Ziel dieser Arbeit ist es, eine Darstellung zu finden, mit der Personen leicht verstehen, wie sich ein Bestand aufgrund von zwei Flussgrößen verändert.

Im folgenden Abschnitt erkläre ich zwei BF-Aufgaben genauer[1]. Weiterhin gehe ich auf die wichtigsten Bestandteile von BF-Aufgaben ein. Das soll erklären, wie die Aufgaben genau aussehen und wo es möglich ist sie zu modifizieren, damit sie leichter zu verstehen sind.

1.1.2 Beispiele von BF-Systemen

Booth Sweeney und Sterman (2000) untersuchten die Leistung in BF-Aufgaben. Sie scheinen die ersten zu sein, die diese Aufgaben systematisch untersuchten. Sie fanden unerwartet schlechte Ergebnisse, was das Lösen der Aufgaben betraf. In ihrer Studie erhielten die Versuchspersonen mehrere BF-Aufgaben, unter anderem die „Badewannen-Aufgabe“ und die „Konto-Aufgabe“. Im ersten Fall handelte es sich beim BF-Szenario um eine

1 In der vorliegenden Arbeit verwende ich den Begriff "Aufgabe" als Übersetzung des englischen Begriffes "task", der meistens in der Literatur im Zusammenhang mit BF-Studien verwendet wurde. Tatsächlich kann das Bearbeiten von BF-Aufgaben als Problemlöseprozess angesehen werden (Schwarz, 2016).

Badewanne, in die Wasser hinein- und aus der Wasser herausfloss. Diese Version mit nur zwei Flussgrößen, also mit einer Zuflussgröße und einer Abflussgröße, ist durchgängig die relevante Version eines BF-Systems für diese Arbeit.

Die Versuchspersonen erhielten zwei Aufgaben. Für jede Aufgabe war jeweils ein Diagramm mit den beiden Flussgrößen gegeben. Die Aufgaben waren einerseits „Rechteckverlauf“ und andererseits der „Dreieckverlauf[2]“ (Abbildung 1.1a und Abbildung 1.2a). Aufgrund des Zu- und Abflusses sollten die Versuchspersonen ermitteln, wie sich der Bestand in der Badewanne über die Zeit hinweg ändert. Abbildung 1.1b und 1.2b zeigen jeweils die korrekte Lösung. Die Versuchspersonen sollten aufgrund der zwei Flussgrößen den Bestand ermitteln und in ein leeres Diagramm ihre Lösung einzeichnen. Hier sind zwei der wichtigsten Bestandteile von BF-Aufgaben enthalten: das Flussdiagramm und das Lösungsdiagramm.

2 Mathematisch gesehen handelt es sich bei den Graphen bzw. Kurven um sogenannte "Rechteckschwingung" und "Dreieckschwingung". Vereinfachend werden diese Graphen als "Rechteckverlauf" und "Dreieckverlauf" bezeichnet.

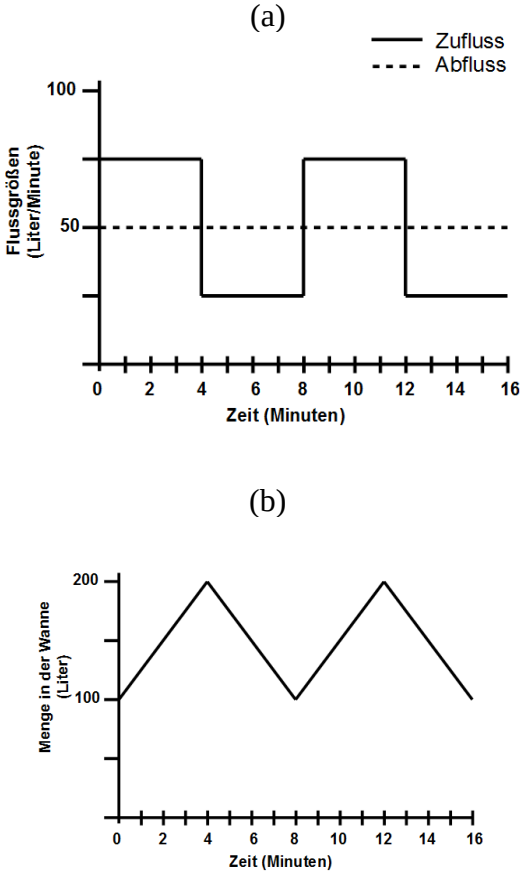


Abbildung 1.1 Flussdiagramm für die Badewannen-Aufgabe mit Rechteckverlauf (a) und die korrespondierende korrekte Lösung (b) (modifiziert nach Booth Sweeney & Sterman, 2000)

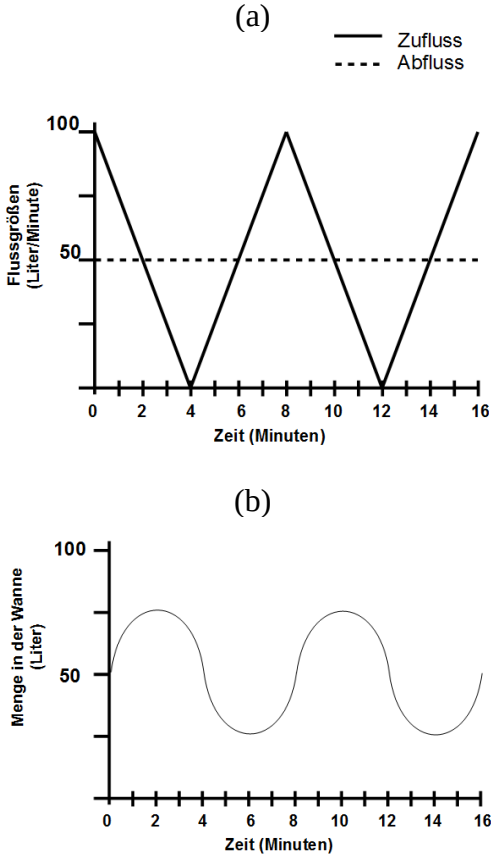


Abbildung 1.2 Flussdiagramm mit Dreieckverlauf (a) und die korrespondierende korrekte Lösung (b) (modifiziert nach Booth Sweeney & Sterman, 2000)

Entsprechend war die „Konto-Aufgabe“ konzipiert. Bei dieser Aufgabe sollten die Versuchspersonen das Guthaben einer Firma (den Bestand) ermitteln. Sie erhielten hierfür Informationen über die Einzahlungen (Zufluss) und Auszahlungen (Abfluss) der Firma, wiederum in Form eines Liniendiagramms. Auch bei diesem Szenario gab es eine Aufgabe mit Rechteck- und eine mit Dreieckverlauf.

1.1.3 Analyse des Verständnisses bei BF-Aufgaben

In vielen Studien wurde untersucht, wie gut BF-Systeme verstanden werden. Es zeigte sich immer wieder, dass die Aufgaben Probleme bereiten. Einige dieser Studien beschreibe ich im weiteren Verlauf.

Booth Sweeney und Sterman (2000) untersuchten alle vier Kombinationen aus BF-Größe (Badewanne vs. Konto) und Flussgrößen-Verlauf (Rechteck- vs. Dreieckverlauf) aus dem vorhergehenden Abschnitt. Um die Lösungsleistung der Versuchspersonen zu beurteilen, nutzten Booth Sweeney und Sterman sieben Kriterien für die Lösung des Rechteckverlaufs und acht Kriterien für die Lösung des Dreieckverlaufs (Anhang A). Die Auswertung der Lösungen erfolgte derart: Booth Sweeney und Sterman prüften bei jedem Verlauf, den eine Versuchsperson für den Bestand an-

gegeben hatte, ob die einzelnen Kriterien erfüllt waren. Dann berechneten sie für jedes einzelne Kriterium den Anteil derer, die das Kriterium korrekt gelöst hatten. Der Mittelwert dieser Anteile ergibt die *durchschnittliche prozentuale Leistung* für eine Aufgabe. Diese lag für den Rechteckverlauf bei 77% und für den Dreieckverlauf bei 48% (in beiden Fällen gemittelt für das Badewannen- und Konto-Szenario). Die gezeigte BF-Leistung war weitaus schlechter als erwartet, da es sich bei den Versuchspersonen um in mathematischen Bereichen formal sehr gut vorgebildet Studierende des MIT handelte. Der Anteil der Personen, die eine Aufgabe komplett korrekt lösten, ist geringer als die oben angegebene durchschnittliche Leistung. Der Anteil lag in dieser Studie für den Dreieckverlauf bei 36% wie Sterman (2002) später berichtete.

Weitere Studien bekräftigten die Ergebnisse von Booth Sweeney und Sterman (2000). Ossimitz (2002) replizierte die Studie mit Badewannen-Szenario und Rechteck- sowie Dreieckverlauf (also ohne das Konto-Szenario). Die BF-Leistung war weitaus schlechter als in der Studie von Booth Sweeney und Sterman (2000): Bei Ossimitz (2002) lag die durchschnittliche Leistung beim Rechteckverlauf bei 42%, während Booth Sweeney und Sterman (2000) für das Badewannen-Szenario 83% gefunden hatten. Für den Dreieckverlauf in Kombination mit Badewannen-

Szenario ergab sich eine durchschnittliche Leistung von 28% statt 48% bei Booth Sweeney und Sterman (2000). Weitere Studien zeigten ebenfalls, dass kein besonders gutes BF-Verständnis vorlag, wenn der Rechteck- oder Dreieckverlauf gegeben war (z.B. Armenia, Onori & Bertini, 2004; Kapmeier, 2004; Kapmeier, Happach & Tilebein, 2014; Schwarz, 2016).

Später wurden Aufgaben mit diskontinuierlichem Flussgrößen-Verlauf untersucht. Das bedeutet, dass im Verlauf keine Regelmäßigkeit oder Muster zu erkennen ist. Diese Aufgaben sollten einfacher zu lösen sein. Sterman (2002) nutzte die Kaufhaus-Aufgabe. Bei dieser Aufgabe stellen Personen, die das Kaufhaus betreten, den Zufluss dar und Personen, die es verlassen, den Abfluss. Beide Flussgrößen weisen einen diskontinuierlichen Verlauf auf (Abbildung 1.3). Zudem beantworteten die Versuchspersonen Fragen statt ihre Lösung zu zeichnen. Zwei Fragen prüften, ob die Versuchspersonen fähig waren, das Flussdiagramm korrekt zu lesen (wann kommen die meisten Personen herein: maximaler Zufluss; wann gehen die meisten heraus: maximaler Abfluss). Die anderen zwei Fragen prüften, ob die Versuchspersonen auf den Bestand schließen konnten (wann waren die meisten Personen im Kaufhaus: maximaler Bestand; wann waren die wenigsten Personen im Kaufhaus: minimaler Bestand). Sterman beschreibt die Aufgabe als einfacher zu lösen als die Aufgaben mit oben be-

schriebenen Verläufen (Rechteck- und Dreieckverlauf). Denn die Fragen der Kaufhaus-Aufgabe sind ohne Berechnungen beantwortbar. Wenn Versuchspersonen verstehen, dass der Bestand zunimmt, solange der Zufluss größer als der Abfluss ist, dann können sie die Fragen nach dem maximalen Bestand korrekt beantworten. Das ist der Fall bis Minute 13, also ist die korrekte Antwort für den maximalen Bestand t = Minute 13. Auch die Frage nach dem minimalen Bestand ist ohne Berechnungen zu lösen. Der Bestand nimmt zu bis Minute 13, danach nimmt er wieder ab, denn ab diesem Zeitpunkt ist der Abfluss größer als der Zufluss. Das heißt, der Bestand ist entweder zu Beginn oder zum Schluss minimal. Um das herauszufinden genügt es, die Fläche zwischen Zufluss und Abfluss vor und nach Minute 13 zu vergleichen. Die Fläche nach Minute 13 ist die größere, somit verlassen nach Minute 13 mehr Personen das Gebäude, als vor Minute 13 in das Gebäude hereinkommen. Das bedeutet, der Zeitpunkt für den minimalen Bestand ist am Ende: Die korrekte Lösung ist Minute 30.

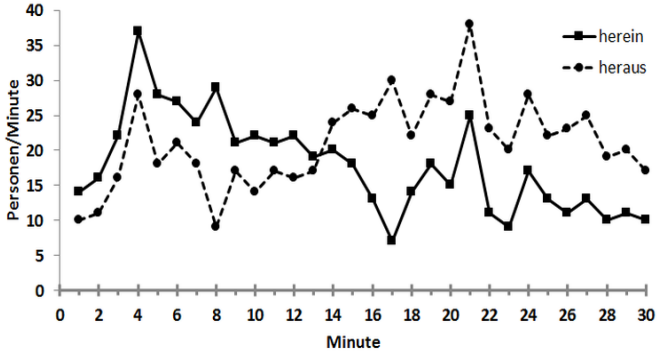


Abbildung 1.3 Flussdiagramm für die Kaufhaus-Aufgabe, bei der Personen ein Kaufhaus betreten („herein“) und es verlassen („heraus“) (modifiziert nach Cronin et al., 2009).
Korrekte Antworten für die Flussgrößen: Zufluss maximal ist bei Minute 4 und Abfluss maximal ist bei Minute 21; für den Bestand: Bestand maximal ist bei Minute 13 und Bestand minimal ist bei Minute 30, wobei eine Minute vor oder nach dem richtigen Zeitpunkt auch als korrekt angesehen wurde.

Die Versuchspersonen schnitten bei den Fragen nach maximalem Zu- und Abfluss gut ab. In beiden Fällen beantworteten 94% die Frage richtig. Jedoch beantworteten nur 42% die Frage nach dem maximalen Bestand korrekt und 30% die Frage nach dem minimalen Bestand. Basierend auf dieser Studie schloss Sterman verschiedene Ursachen für die Probleme beim Lösen von BF-Szenarien aus. Die Probleme sollten nicht daraus resul-

tieren, dass die Versuchspersonen die Diagramme nicht verstehen, dass sie die adäquaten Rechenwege nicht wissen oder dass ihnen Zeit fehlt, die Aufgaben zu lösen. Weiterhin folgerte Sterman, dass für die Versuchspersonen nicht das Problem darin liegt, die Flussdiagramme zu lesen, sondern dass die Versuchspersonen das Konzept der *Akkumulation* nicht verstehen. Um zu akkumulieren, müssen Versuchspersonen zunächst den Zu- und Abfluss verrechnen. Die Differenz von Zu- und Abfluss ergibt den Nettofluss. Dieser Wert wird zum aktuellen Bestand addiert: Dabei handelt es sich um Akkumulation.

In weiteren Studien wurden diskontinuierliche Flussgrößen-Verläufe untersucht. Dabei ergaben sich zumeist keine hohen Anteile für korrekte Antworten (Cronin & Gonzalez, 2007; Cronin et al., 2009; Ossimitz, 2002; Pala & Vennix, 2005). Das eingeschränkte Verständnis von BF-Systemen stützt die zu Beginn erwähnte Annahme, dass komplexe dynamische Systeme schon aufgrund ihrer Bausteine beziehungsweise Subsysteme schwer verständlich sind und nicht ausschließlich aufgrund der Komplexität (vgl. Kapitel 1.1.1). Die einzelnen Bausteine zu verstehen ist demnach wichtig, um mit komplexen dynamischen Systemen adäquat umzugehen. In einigen Studien fanden bereits Versuche statt, das BF-Verständnis zu erhöhen. Das thematisiert der folgende Abschnitt.

1.1.4 Bisherige Versuche zur Erhöhung des BF-Verständnisses

Weitere Studien untersuchten, ob Änderungen von BF-Aufgaben oder des BF-Szenarios zu besserer Leistung führen. Z.B. führten Cronin et al. (2009) mehrere Experimente durch, wobei sie meist die Kaufhaus-Aufgabe nutzten (mit diskontinuierlichem Flussgrößen-Verlauf wie in Abbildung 1.3, aber auch mit anderen BF-Verläufen). Sie replizierten die bereits beschriebene schlechte BF-Leistung (Kapitel 1.1.3). Außerdem untersuchten die AutorInnen, ob begrenzte kognitive Kapazitäten zu den Problemen führten: Vielleicht waren die Versuchspersonen überfordert von den vielen Schritten, die notwendig sind, die Kaufhausaufgabe zu lösen. Zwar reicht bei dieser Aufgabe ein Blick auf die Abbildung und mit ausreichend Vorwissen ist klar, wann der Bestand minimal oder maximal ist. Wenn die Versuchspersonen jedoch rechnen, dann müssen sie herausfinden, wie viele Personen jede Minute herein und heraus gehen, die Differenzen berechnen, diese zum vorhergehenden Bestand addieren und diese Menge im Kopf behalten. Die AutorInnen vereinfachten die Aufgabe, um herauszufinden, ob sich dadurch die BF-Leistung verbesserte. Sie schufen einen Flussgrößen-Verlauf mit weniger Zeitpunkten (12 statt 30 Minuten) und weniger Personen pro Minute (höchstens 15

statt bis zu 40 Personen für eine Flussgröße pro Zeiteinheit). Das führte nicht zu signifikant besseren Lösungen. Das Senken der kognitiven Beanspruchung führt also nicht zu verbesserter BF-Leistung.

Außerdem untersuchten die AutorInnen, ob eine Änderung von Situationsfaktoren die Leistung verbesserte, etwa das Erhöhen der Motivation oder Rückmeldung, wenn die Lösung falsch war. Die AutorInnen hielten es für möglich, dass die Versuchspersonen einer ersten falschen Intuition folgen, obwohl sie eigentlich wissen, wie Fluss- und Bestandsgrößen miteinander zusammenhängen. Somit müssten die Versuchspersonen ihre falsche Lösung leicht korrigieren können, wenn sie informiert werden, dass ihr Ergebnis falsch ist. Außerdem sollten sie mit erhöhter Motivation einfacher zur korrekten Lösung kommen. Jedoch verbesserte keine der beiden Manipulationen die BF-Leistung.

Weiterhin versuchten die AutorInnen latentes BF-Wissen zu aktivieren, wenn es bereits besteht. Das bezeichneten sie als Priming. Die Versuchspersonen erhielten eine BF-Aufgabe als Priming mit sehr simplem Flussgrößen-Verlauf (4 Zeitpunkte und kein Schnittpunkt der Flussgrößen) bevor sie die Kaufhaus-Aufgabe bearbeiteten. Das verbesserte im Vergleich der Gruppen tatsächlich die Lösung: In der Baseline-Bedingung ergaben sich 8% (zum maximalen Bestand) und 16% (zum minimalen Bestand)

korrekte Lösungen, während es in der Priming-Bedingung 27% und 38% waren. Betrachtet man jedoch die Anteile für sich, so ist die BF-Leistung immer noch eher schlecht. Zudem löste fast die Hälfte der Versuchspersonen nicht einmal die sehr einfache Priming-Aufgabe korrekt. Zusätzlich analysierten die AutorInnen die Leistung der Kaufhaus-Aufgabe detaillierter für die Versuchspersonen der Bedingung mit Priming-Aufgabe. Die AutorInnen verglichen die Gruppe der Versuchspersonen, die die Priming-Aufgabe korrekt gelöst hatte mit der Gruppe, die diese Aufgabe nicht korrekt gelöst hatte. Von denjenigen mit korrekter Priming-Aufgabe lösten 53% und 68% die folgende Aufgabe richtig, während es in der anderen Gruppe 0% und 6% waren. Diese Unterschiede waren signifikant ($p = .0004$ und $p = .0001$).

Keine der weiteren vorgenommenen Manipulationen zeigten großen oder überhaupt Erfolg. Auch Trainings für das BF-Verständnis, wie beispielsweise ein Crash-Kurs von 90 Minuten (Kainz & Ossimitz, 2002), waren nicht hilfreich. Da alle bisher beschriebenen Modifikationen zu keiner Verbesserung in der BF-Leistung geführt haben, stellt sich die Frage, was stattdessen helfen könnte. Vor Überlegungen zu dieser Frage folgt ein Abschnitt zu einzelnen Beschreibungsdimensionen von BF-Systemen.

1.1.5 Beschreibungsdimensionen für BF-Systeme

Im Folgenden gehe ich auf verschiedene Bestandteile und mögliche Eigenschaften von BF-Systemen ein, wobei ich diese unter dem Begriff „Beschreibungsdimensionen“ zusammenfasse. Änderungen in den Aufgabenstellungen können systematisch den einzelnen Dimensionen zugeordnet werden, und die folgenden Begriffe nutze ich durchgängig für die vorliegende Arbeit. Bei den Dimensionen handelt es sich um den Flussgrößen-Verlauf (kontinuierlich versus diskontinuierlich; oder Rechteckverlauf) und das Antwortformat (Lösungsgraph zeichnen oder Fragen nach dem Bestand beantworten). Weitere Beschreibungsdimensionen sind das Darstellungsformart der Flussgrößen (z.B. Werte der Flussgrößen in Diagramm oder in Tabelle), das verwendete Szenario (z.B. Wasserbecken oder Kaufhaus), damit eng zusammenhängend die BF-Größe (z.B. Wasser oder Personen) und eine Abbildung des BF-Szenarios (schematische Abbildung, z.B. ein gezeichnetes Wasserbecken). Zu einigen dieser Beschreibungsdimensionen wurden bereits Studien durchgeführt, deren Ergebnisse ebenfalls erwähnt sind.

Flussgrößen-Verlauf

Hier bezeichne ich als Flussgrößen-Verlauf wie genau die Flussgrößen aussehen in Bezug auf ihren Verlauf über die Zeit hinweg. In der vorliegenden Arbeit gelten folgende Einteilungen und Begriffe, die ich so für die komplette Arbeit nutze: Eine grobe Einteilung ist diskontinuierlicher versus kontinuierlicher Verlauf. Die Kaufhaus-Aufgabe (z.B. Cronin et al., 2009, siehe Abbildung 1.3) ist ein Beispiel für einen diskontinuierlichen Flussgrößen-Verlauf. Die Graphen (bzw. Kurven) für Zu- und Abfluss enthalten unregelmäßige Sprünge, während die Graphen für Rechteck- oder Dreieckverlauf (Booth Sweeney & Sterman, 2000, siehe Abbildung 1.1 und 1.2) ein regelmäßiges Muster aufweisen. Diese zweite Form wird als kontinuierlicher Flussgrößen-Verlauf bezeichnet. Eine weitere Einteilung für die kontinuierlichen Flussgrößen-Verläufe erfolgt für die häufig genutzten Funktionen, beispielsweise Rechteck- und Dreieckverlauf. Bei den diskontinuierlichen Aufgaben können einzelne Merkmale für die Einteilung herangezogen werden, wie mit und ohne Schnittpunkt der Flussgrößen.

Direkte und systematische Vergleiche von „kontinuierlichem“ und „diskontinuierlichem“ Verlauf fanden meines Wissens nicht statt in dem Zeitraum bevor die hier beschriebenen Studien

durchgeführt wurden. In späteren Studien bearbeiteten Versuchspersonen mehrere Aufgaben in derselben Bedingung. Dort liegen zumindest einzelne Vergleichswerte für die verschiedenen Verläufe vor (Schwarz, 2016). In einer dieser Studien zeigte der Dreieckverlauf eine ähnliche Lösungsgüte wie ein diskontinuierlicher Verlauf (angelehnt an den Verlauf einer Studie von Sterman, 2002 bzw. Cronin et al., 2009) mit etwa 30%, während der Rechteckverlauf in dieser Studie etwa 60% erbrachte. In einer weiteren Studie schnitt der Rechteckverlauf leicht schlechter ab als ein diskontinuierlicher Verlauf, der sehr ähnlich war wie in der vorher genannten Studie (etwa 20% versus etwa 30%). Booth Sweeney und Sterman (2000) verglichen Rechteck- und Dreieckverlauf miteinander, wobei sich der Dreieckverlauf als schwieriger erwies (siehe Kapitel 1.1.3). Ossimitz (2002) untersuchte verschiedene Aufgaben in einer Stichprobe (bspw. in Bezug auf BF-Größe oder Flussgrößen-Verlauf). Die Lösungsraten für den Dreieckverlauf waren wiederum schlechter als die für Rechteckverlauf: Für Rechteckverlauf ergab sich als durchschnittliche Leistung 42% und für den Dreieckverlauf 28%. Der Anteil der KorrektlöserInnen bei einem diskontinuierlichen Flussgrößen-Verlauf lag bei 24% für die Frage nach dem maximalen Bestand. Das ist in etwa derselbe Wert wie beim Rechteckverlauf.

Auch in der vorliegenden Arbeit werden Rechteck- und Dreieckverlauf häufig genutzt. Die Ergebnisse, was die Schwierigkeit der Flussgrößen-Verläufe betrifft, sind nicht konsistent. Jedoch scheint die Aufgabe mit Dreieckverlauf tendenziell am schwierigsten zu lösen zu sein. Ein diskontinuierliches Muster, wie von Cronin et al. (2009) genutzt, scheint teilweise zu ähnlichen korrekten Lösungsraten zu führen wie der Rechteckverlauf.

Antwortformat

Es existieren in der Regel zwei Versionen für das Antwortformat: Häufig zeichneten die Versuchspersonen entweder den Verlauf des Bestands in ein vorgegebenes Diagramm. Das wird in der vorliegenden Arbeit als Lösungsgraph bezeichnet. Alternativ beantworteten die Versuchspersonen Fragen zum Bestand. Meist ergab sich das Antwortformat aus dem Flussgrößen-Verlauf: In den meisten Studien mit diskontinuierlichen Flussgrößen-Verläufen beantworteten die Versuchspersonen gegebene Fragen. Bei kontinuierlichen Flussgrößen-Verläufen zeichneten sie ihre Lösung in ein gegebenes Diagramm. Meines Wissens nach gab es vor Beginn der vorliegenden Arbeit keinen direkten Vergleich von Antwortformaten: die Versuchspersonen zeichneten entweder ihre Lösung ein oder sie beantworteten Fragen, taten aber nie bei-

des. In späteren Studien erfolgte dies, darauf gehe ich in Abschnitt 4.4.1 ein.

Darstellungsformat der Flussgrößen

Für die Flussgröße können verschiedene Darstellungsformate gefunden werden: zum Beispiel werden die Daten in einem Diagramm oder in Tabellenform dargestellt. In den meisten Fällen werden Liniendiagramme verwendet, im weiteren Verlauf werden diese als „klassische Diagrammdarstellung“ für BF-Aufgaben bezeichnet.

Cronin et al. (2009) untersuchten, ob die Art der Darstellung die BF-Leistung beeinflusst. Sie vermuteten, dass die Versuchspersonen die relevanten Informationen nicht erkennen, die notwendig für das Lösen der Aufgabe sind. Um herauszufinden, ob das Liniendiagramm zu den Problemen führt, zeigten sie die Daten auch als Text und in einer Tabelle. Außerdem untersuchten sie, ob das Problem in der speziellen Darstellung mit Linien liegt. Sie verwendeten als Vergleichsbedingung ein Säulendiagramm. Sie vermuteten, dass das Liniendiagramm ungünstig ist, weil Linien sich verändernde Raten zwischen den aufeinanderfolgenden Zeitpunkten suggerieren. Ein Liniendiagramm könnte also auf Einheiten hinweisen, die sich in jedem Moment kontinuierlich ändern (wie Wasser), während tatsächlich die Einheiten der Kauf-

haus-Aufgabe diskreter Natur sind (Personen). Somit nutzen sie ein Säulendiagramm, in dem häufig Gesamtbeträge über kleine Zeiteinheiten (hier Minuten) zusammengefasst und abgetragen werden.

Cronin et al. testeten die nicht-graphischen versus graphischen Bedingungen, das heißt Text und Tabelle versus Linien- und Säulendiagramm. Die graphischen Bedingungen schnitten etwas besser ab. Es folgen die Werte für jeweils 1. maximaler und 2. minimaler Bestand: Für die graphische Bedingung ergaben sich 61% und 48%, während sich für die nicht graphische Bedingung 51% und 44% ergaben (p = .137 und p = .537). Außerdem verglichen Cronin et al. innerhalb der graphischen Bedingung beide Diagrammarten: Das Säulendiagramm schnitt leicht besser ab als das Liniendiagramm (Säulendiagramm: 69% und 55%; Liniendiagramm: 52% und 41%; mit p = .071 und p = .157).

BF-Szenario und BF-Größe

Worum geht es bei dem gegebenen BF-Szenario? Geht es um ein Kaufhaus oder ein Konto? Damit hängt zusammen, welche BF-Größen vorliegen. Das können Personen sein wie in der Kaufhaus-Aufgabe (Cronin et al., 2009) oder Geld wie bei der „Konto-Aufgabe“ (Booth Sweeney & Sterman, 2000, siehe Abschnitt 1.1.2). Hierzu erfolgten bereits Studien. Booth Sweeney

und Sterman (2000) verglichen die Leistung für das BF-Szenario mit Badewanne und die Leistung beim Konto-Szenario. Wasser als BF-Größe führte beim Rechteckverlauf zu signifikant besserer Gesamtleistung (83% für Wasser vs. 69% für Geld), während die BF-Größe beim Dreieckverlauf zu keinem signifikantem Unterschied führte. In einer weiteren Studie verglichen Cronin et al. (2009) verschiedene BF-Größen. Sie wollten herausfinden, ob die Probleme daher kommen, dass Versuchspersonen ihr latentes BF-Verständnis nicht aktivieren. Sie nahmen an, eine anderes BF-Szenario könnte dieses aktivieren. Beispielsweise könnte latentes Wissen vorhanden sein, jedoch nur gebietsspezifisch. Sie verglichen die BF-Größe Personen (klassische Kaufhaus-Aufgabe) mit den Größen Wasser (Badewannen-Aufgabe) und Geschwindigkeit. Im letzten Fall handelte es sich um eine „Auto-Aufgabe“. Bei dieser sind Zu- und Abfluss die Geschwindigkeit zweier Autos, die in dieselbe Richtung fahren und der Bestand ist ihr Abstand. Das stellte ihrer Meinung nach eine Aufgabe dar, mit deren Kontext die Versuchspersonen besser vertraut sein sollten als mit Personen, die in ein Kaufhaus hineingehen und es verlassen. Auch das Badewannen-Szenario sollte den Versuchspersonen vertrauter sein und somit zu besserer BF-Leistung führen. Für die Fragen nach dem maximalen und minimalen Bestand beim Vergleich von Kaufhaus-Aufgabe und Auto-Aufgabe schnitt die BF-

Größe Auto tendenziell besser ab: Bei der Kaufhaus-Aufgabe ergab sich 22% für die Frage nach maximalen Bestand und 17% für minimalen Bestand. Bei der Auto-Aufgabe ergaben sich 38% und 31%. Jedoch zeigte sich keine signifikante Verbesserung in der Leistung ($p = .457$ und $p = .429$). Beim Vergleich der Badewannen- und Kaufhaus-Aufgabe führte erstere zu tendenziell besserer Leistung bei der Frage nach dem minimalen Bestand (Kaufhaus 17%, Badewanne 31% mit $p = .413$), bei der Frage nach dem maximalen Bestand ergab sich kein Unterschied (Kaufhaus 22%, Badewanne 23% mit $p = 1.000$). Die AutorInnen folgerten, dass das Problem nicht daher kommt, dass BF-Verständnis gebietsspezifisch ist, sondern dass es nicht existiert. Wobei anzumerken ist, dass ihre Stichprobe mit 47 Versuchspersonen recht klein war und der Anteil der Korrektlösenden beim Auto-Szenario tatsächlich fast doppelt so groß war wie der beim Kaufhaus-Kontext. Somit könnte durch eine zu geringe Teststärke ein vorhandener Effekt zu einem nicht signifikanten Ergebnis geführt haben. Andererseits handelte es sich auch bei der Auto-Aufgabe um Anteile von 38% und 31%, was den häufig vorliegenden Wertebereich der Lösungsraten derartiger Aufgaben nicht auffällig übersteigt (vgl. oben genannte Werte).

Schematische Abbildung des BF-Szenarios

In einigen Studien ist eine Abbildung des gegebenen BF-Systems beigefügt, beispielsweise eine schematische Abbildung einer Badewanne (Booth Sweeney & Sterman, 2000). In anderen Studien ist keine BF-Abbildung gegeben oder es wurde zumindest nicht erwähnt (Ossimiz, 2002; Pala &Vennix, 2005; Phuah, 2010; Brunstein, Gonzalez & Kanter, 2010; Cronin & Gonzalez, 2007). In einer Studie mit einer Reihe von BF-Aufgaben von Kainz und Ossimitz (2002) sind alle Aufgaben ohne zusätzliche Abbildung des Szenarios gegeben. Eine Ausnahme stellt die Badewannen-Aufgabe dar, die eine Abbildung mit der Badewanne wie in der Studie von Booth Sweeney und Sterman (2000) enthielt. Bisher wurde jedoch nicht systematisch variiert, ob eine Abbildung des BF-Szenarios hilfreich für die BF-Leistung ist.

1.1.6 Auffälligkeit bezüglich der BF-Lösungen

Wenn die falschen Lösungen der BF-Aufgaben näher betrachtet werden, so zeigt sich ein systematisches Fehlermuster. Die Versuchspersonen nehmen an, die Bestandsgröße ändert sich entsprechend einer Flussgröße, sei es Zu-, Ab-, oder Nettofluss (also die Differenz der beiden vorher genannten Flussgrößen). Booth Sweeney und Sterman (2000) vermuteten, die Versuchs-

personen orientieren sich am Verlauf einer der Flussgrößen und übertragen das auf die Bestandsgröße. Cronin et al. (2009) bezeichneten diese Vorgehensweise als Korrelations-Heuristik und untersuchten sie genauer. Zunächst untersuchten die AutorInnen in drei Studien, welche Zeitpunkte die Versuchspersonen fälschlicherweise für den maximalen und minimalen Bestand angaben. Die häufigsten falschen Angaben für den Zeitpunkt des maximalen Bestandes waren der des maximalen Nettoflusses (33% bis 54% der falschen Antworten) und entsprechend für den minimalen Bestand der Zeitpunkt des geringsten Nettoflusses (18% bis 43%). Eine dieser Studien war die obengenannte Studie mit verschiedenen Darstellungsformaten (graphische Bedingung: Werte der Flussgrößen in Säulen- oder Liniendiagramm gegeben vs. nicht-graphische Bedingung: Werte in Tabelle oder Text). Je nach Darstellungsformat variierte die Häufigkeit, die Korrelations-Heuristik zu nutzen. Versuchspersonen verwendeten sie bei gezeigtem Liniendiagramm häufiger als beim Säulendiagramm. In den Bedingungen mit graphischem Darstellungsformat verwendeten die Versuchspersonen sie häufiger als in den nicht-graphischen Bedingungen. Für eine weitere Studie entwickelten die AutorInnen acht verschiedene Flussgrößen-Verläufe. Beispielsweise verliefen Zu- und Abfluss parallel oder der Zufluss zeigte die Form eine Glockenkurve, während der Abfluss konstant bei ei-

nem Wert blieb. Sie analysierten, wie oft die Versuchspersonen die Korrelations-Heuristik orientiert an Zu- oder Nettofluss anwandten. In einem Fall waren es 33% der falschen Lösungen, ansonsten schwankten die Anteile von 56% bis 89%. Die Korrelations-Heuristik ist somit ein häufig auftretender systematischer Fehler und wird in den Fehleranalysen der vorliegenden Arbeit berücksichtigt.

1.1.7 Fazit

BF-Systeme sind sehr schwierig zu verstehen, die Lösungsraten sind – abgesehen von vereinzelten Fällen – niedrig bis extrem niedrig. Verschiedene Modifikationen führten zu keiner oder nur geringer Verbesserung der Leistung. Weder erhöhte Motivation noch Rückmeldung war hilfreich, während die Aktivierung von latentem BF-Wissen zu einer leichten Verbesserung in der Leistung führte. Die Ergebnisse zum Einfluss verwendeter BF-Szenarien sind nicht eindeutig: In manchen Fällen beeinflusste das verwendete Szenario die Leistungsgüte, in anderen Fällen nicht. Das Darstellungsformat führte zu tendenziellen Unterschieden: Mit einem Säulendiagramm lösten die Versuchspersonen BF-Aufgaben etwas häufiger korrekt als mit einem Liniendiagramm, und insgesamt ergab sich ein leicht höherer Anteil an korrekten Lösungen in den graphischen Bedingungen.

Es wird immer wieder betont, dass fundamentale Fehler gemacht werden, insbesondere, dass das Prinzip der Akkumulation nicht verstanden wird. Ossimitz (2002) vermutete, das Problem besteht darin, dass die Versuchspersonen die Flussgrößen für Bestandsgrößen halten. Das scheint zum häufigen Nutzen der Korrelations-Heuristik zu passen: die Versuchspersonen nehmen anscheinend an, dass sich der Bestand wie eine Flussgröße ändert. Bisher existieren keine Theorien, die direkt zum BF-Verständnis entwickelt wurden (vgl. auch Schwarz, 2016). Somit sollen theoretische Aspekte aus anderen Bereichen herangezogen werden, um passende Modifikationen für eine Verbesserung der BF-Leistung zu finden. Im Folgenden werden entsprechende Erkenntnisse aus dem Bereich Intuition und Problemlösen beziehungsweise mathematische Intuitionen beschrieben.

1.2 Intuition

In den bisher genannten Studien zu BF-Systemen zeigten sich keine bis höchstens minimale Verbesserungen in der BF-Leistung durch die vorgenommenen Modifikationen. Die vorliegende Arbeit untersucht, ob es möglich ist, die Leistung durch intuitives Verständnis zu verbessern. Bei den BF-Aufgaben geht es darum, von zwei Flussgrößen auf den Bestand zu schließen. Diese Aufgaben können also zum Bereich mathematischer Aufgaben

gezählt werden. Andererseits können sie als passend zum Themenbereich Problemlösen angesehen werden. Die folgenden Abschnitte erläutern wie in diesen beiden Bereichen Änderungen der Aufgaben- oder Problemstellung hilfreich waren. Zunächst wird geklärt, was in vorliegender Arbeit unter Intuition verstanden wird, insbesondere Intuitionen beim Problemlösen und mathematischer Intuition.

1.2.1 BF-Systeme und Intuition

Booth Sweeney und Sterman (2000) schrieben zunächst, dass sie annahmen, Intuitionen zu BF-Systemen sollten bei Menschen vorhanden sein. Denn diese Aufgaben sind meist schon mit sehr simplen Berechnungen zu lösen (vgl. Abschnitt 1.1.3). Außerdem sollte das Konzept der Akkumulation uns allen aus unseren alltäglichen Erfahrungen klar sein: Wir können beispielsweise eine Badewanne mit Wasser füllen und unser Konto verwalten. Aufgrund ihrer gefundenen Ergebnisse änderten Booth Sweeney und Sterman ihre Meinung, dass Personen häufig über korrekte Intuitionen zu BF-Systemen verfügen. Booth Sweeney und Sterman kamen zu dem Schluss, dass bei einem großen Anteil der Versuchspersonen die Erfahrung im mathematischen Bereich nicht dazu führt, dass sie Intuitionen für das Prinzip der Akkumulation entwickeln und das trotz hoher Vorbildung seitens der Ver-

suchspersonen. Ihr Fazit aus der Studie war somit, diese Intuitionen existieren nicht und sollten entwickelt werden. Auch Cronin et al. (2009) interpretierten die Ergebnisse ihrer Studien dahingehend, dass bei vielen Personen kein intuitives BF-Verständnis existiert.

An dieser Stelle – vor dem Durchführen der eigenen Studien – schien es mir verfrüht, davon auszugehen, dass keine Intuitionen für BF-Systeme existieren. Auch wenn schon einige Modifikationen der Aufgabendarstellung vorgenommen wurden, gibt es noch weitere mögliche Veränderungen, die noch nicht näher untersucht wurden und hilfreich sein könnten, Intuitionen zu triggern. Weiterhin könnten in Studien (z.B. von Cronin et al., 2009) Effekte übersehen worden sein, da nur auf Signifikanz geprüft wurde und die Stichproben teilweise recht klein waren (natürlich könnten die teilweise auftretenden tendenziellen Unterschiede auch auf Zufall zurückgehen). Zunächst beschreibe ich Intuition, wie sie in der vorliegenden Arbeit gesehen wird, im weiteren Verlauf gehe ich wieder auf Intuition im Zusammenhang mit BF-Aufgaben ein. Da der Begriff Intuition in der Literatur teilweise sehr unterschiedlich verwendet wird, kann hier kein umfassender Überblick gegeben werden. Einen Überblick geben beispielsweise Hogarth (2001) und Plessner, Betsch und Betsch (2008).

Intuitives und analytisches Vorgehen wird oft einander gegenübergestellt (z.B. Hammond, 1988). Er nimmt an, dass analytische Prozesse eher auf Symbolebene ablaufen, während bildliches Material eher zu intuitiven Prozessen führt. Hogarth (2001) versucht Intuition zu fassen und zu beschreiben, wie man sie „schulen“ kann. Auch Hogarth unterteilt Prozesse der Informationsverarbeitung grob in zwei Teile: das „tacit system“ und das „deliberate system“, welche er eher als Kontinuum und weniger als zwei strikt getrennte Systeme ansieht (S.191). Hogarth beschreibt beispielsweise im Zusammenhang mit Urteilen bei Wahrscheinlichkeiten, dass wir oft nicht wissen, welcher Prozess (analytisch oder intuitiv) welchen Anteil am Lösungsprozess hat. Aber es ist erkenntlich, dass natürliche Formate das Lösen erleichtern (S.255, vgl. auch Abschnitt 1.2.3). Nach Hogarth gehört Intuition zu den informationsverarbeitenden Prozessen, die im unbewussten oder impliziten Bereich ablaufen (im „tacit system“)[3]. Sie umfasst kognitive implizite Prozesse, die z.B. beim Problemlösen oder bei Entscheidungen ablaufen. In Abgrenzung zu anderen AutorInnen (z.B. Dijksterhuis & Nordgren, 2006) können für Hogarth (2001) Emotionen mit intuitiven Prozessen

3 Er betont, nicht bei allen dort ablaufenden Prozesse handelt es sich um Intuition. Für ihn gehört nur ein Teil der automatisch ablaufenden Prozesse zu Intuition. Er grenzt beispielsweise alles was unter „instinktiv“ fällt, von intuitiven Prozessen ab.

in Zusammenhang stehen, jedoch sie sind nicht notwendig für intuitives Vorgehen. Er schreibt zu Intuition oder intuitiven Reaktionen: „They are reached with little effort, and typically without conscious awareness. They involve little or no conscious deliberation“ (S.14). Bei Plessner et al. (2008) findet sich als Definition:

> Intuition is a process of thinking. The input to this process is mostly provided by knowledge stored in long-term memory that has been primarily acquired via associative learning. The input is processed automatically and without conscious awareness. The output of the process is feeling that can serve as a basis for judgments and decisions. (S.24)

Wobei „feeling“ hier sehr weit gefasst ist: ein Gefühl etwas zu mögen oder dass etwas riskant ist. Auch nicht-emotionale Gefühle wie „feeling of knowing“ sind möglich. Diese Art von Gefühlen sollen beispielsweise Entscheidungen oder Urteilen zugrunde liegen, die auf Intuition zurückgehen. Auch nach Plessner et al. ist das Vorgehen eine Hybridform, bei der der Schwerpunkt stärker bei analytisch oder intuitiv liegt. In den gerade beschriebenen Sichtweisen ist Intuition erlernbar. Hogarth (2001) beschreibt für diesen Prozess „freundliche“ und „schlechte“ Lernumgebungen (S.216 ff.). Wenn die Lernumgebung freundlich ist, so sollten gute Intuitionen erlernt werden, die beispielsweise

zu korrekten Urteilen oder Handeln führen. Wichtig ist unter anderem, dass Verbindungen zwischen Handeln und dessen Folgen klar sind, so dass assoziatives Lernen möglich ist.

Entsprechend dieser Sichtweisen wird für die vorliegende Arbeit der Begriff Intuition genutzt. Es handelt sich also um einen Teil von informationsverarbeitenden Prozessen und ist der impliziten Kognition zuzuordnen. Für die vorliegende Arbeit ist nicht von Bedeutung, ob ein Gefühl (sei es emotional oder nicht-emotional) notwendig ist, um von Intuition zu sprechen. Der Schwerpunkt liegt darauf, dass Wissen aus dem Langzeit-Gedächtnis eine Rolle spielt, und dass dieses über die passende Darstellung für das Problemlösen „aktiviert“ werden kann. In dieser Arbeit wird teilweise von "analytisch versus intuitiv" geschrieben, was nicht bedeuten soll, dass das Verhalten klar auf einer Seite liegt. In Anlehnung an hier genannte Autoren (Plessner et al., 2008; Hogarth, 2001) ist eher eine Tendenz gemeint und im Folgenden vereinfacht beschrieben, zu welchem Verhalten die Tendenz neigt.

Im Bereich BF-Aufgaben wurde also nach Durchführen einiger Studien angenommen, dass keine Intuitionen für BF-Systeme existieren. Im Folgenden Abschnitt gehe ich auf den Bereich „mathematische Intuition“ ein, in dem eine entsprechende Sichtweise in Bezug auf mathematische Aufgaben besteht.

1.2.2 Mathematische Intuition

Es gibt die Ansicht, dass sich Intuition in primäre versus sekundäre Intuitionen unterteilen lässt (Ben-Zeev & Star, 2001). Primäre Intuition wird als angeboren angesehen. Im Bezug auf mathematische Intuitionen handelt es sich beispielsweise um „Zähl-Prinzipien“ bei Vorschulkindern. Ein derartiges Prinzip ist, dass beim Zählen einer bestimmten Menge von Objekten jedes Objekt berücksichtigt wird, jedoch nur genau ein Mal (Gelman & Meck, 1983). Sekundäre Intuitionen sind nach Ben-Zeev und Star (2001) Intuitionen, die über langes, formales Training entstehen. In der weiterführenden Mathematik sollen sekundäre Intuitionen entwickelt werden, die AutorInnen sprechen hierbei von symbolischen Intuitionen im Bezug auf den Bereich symbolischer Mathematik. Das Konzept sekundärer Intuitionen passt gut zu Hogarths Begriff von Intuition (2001), da seiner Meinung nach Intuitionen erlernbar sind. Das kann sich bei ihm ebenfalls auf Problemlöse-Prozesse, auch im mathematischen Bereich, beziehen.

Eine Textaufgabe nach Fischbein (1994) veranschaulicht, was mit mathematischer Intuition gemeint ist. Die Anordnung in einer Multiplikationsaufgabe kann das Lösen erleichtern oder erschweren. Es handelte sich um folgende Textaufgabe: „Aus 1 kg Getreide gewinnt man 0,75 kg Mehl. Wieviel Mehl gewinnt man

aus 15 kg Getreide?“. Wurde das Problem nun umgedreht beschrieben: „Von einer bestimmten Chemikalie benötigt man 1 kg um 15 kg Seife herzustellen. Wie viel Seife kann man aus 0,75 kg der Chemikalie herstellen?“ fiel den SchülerInnen das Berechnen schwerer, obwohl es sich grundlegend um dieselbe Berechnung handelte. Jedoch widersprach das Ergebnis der Intuition, dass sich eine Zahl vergrößert, wenn sie mit einer anderen multipliziert wird. Das stimmt, solange es sich um natürliche Zahlen handelt, mit denen zunächst die Multiplikation erlernt wird. Bei einer Zahl kleiner 1, wie im vorliegenden Fall, gilt das nicht mehr.

In Bezug auf sekundäre Intuitionen wird betont, dass in der Problemstruktur oft Hinweisreize vorliegen, die zu einem bestimmten Vorgehen beim Lösen führen. Problematisch ist es, wenn irrelevante Merkmale eines Problems (z.B. ein Wort) zu einer bestimmten Lösungsstrategie führen. Dazu kann es kommen, wenn dieses Merkmal oft in einem Problem vorkam, welches mit dieser Strategie zu lösen war. Wenn dieses Merkmal in einer anderen Art von Problem vorhanden ist, kann das zu der falschen Lösungsstrategie führen. Ein anekdotisches Beispiel aus englischsprachigem Mathematikunterricht (Schoenfeld, 1991, S.323) veranschaulichte dies. Die SchülerInnen sollten Schlüsselworte in Textaufgaben beachten. Wenn sie das Wort „left“ sehen (z.B. „(…) how many are *left*“, sollten sie die Werte der Aufgabe sub-

trahieren. Es entstehen Probleme, wenn nun folgende Aufgabe zu lösen ist: „John had 7 apples. Mary left John 4 apples“. Der erlernte Lösungsprozess würde dann zu 7 – 4 führen statt zu 7 + 4. Hierbei handelte es sich um ein überspitztes anekdotisches Beispiel. Jedoch scheinen tatsächlich ähnliche Effekte zu existieren, die in Studien untersucht wurden. Ben-Zeev (1996) nennt diese „irreführende Korrelationen“. Hierbei entsteht in Augen von SchülerInnen ein Zusammenhang zwischen einem tatsächlich irrelevanten Merkmal in der Aufgabe und dem Lösungsprozess. Zusammenfassend lässt sich sagen, dass Intuitionen im Bereich mathematischer Probleme eine Rolle spielen können: Probleme sind aufgrund vorhandener Intuitionen einfacher oder schwieriger zu lösen.

Zwar liegt der Schwerpunkt bei mathematischer Intuition auf Intuitionen, die im formalen Setting erlernt sind, beispielsweise im Mathematikunterricht. Wenn nun Intuitionen allgemein als erlernbar angesehen werden, so ist es möglich bei mathematischen Problemen die Problemstruktur derart zu gestalten, dass sie passende Intuitionen triggert. Intuitionen können zu einer korrekten Lösung führen, unabhängig davon wie das genaue Vorgehen beschaffen ist. Es könnten auch korrekte analytische Prozesse folgen. Die Intuition hat die Person dazu gebracht, diese anzuwenden.

Eine relevante Frage ist: Existiert in Personen BF-Verständnis, so dass Intuition getriggert werden kann? Cronin et al. (2009) gingen wohl davon aus, als sie über eine Änderung des BF-Szenarios latentes, gebietsspezifisches BF-Wissen triggern wollten (Abschnitt 1.1.5). Hierbei wandelten sie die BF-Größen ab in Abstand (Bestand) und Geschwindigkeit (Flussgrößen). Um auf diese Art latentes BF-Wissen zu aktivieren, sollte dieses bereits vorliegen. Die Versuchspersonen sollten also über das Wissen verfügen, wie die Veränderung der beiden Flussgrößen (die Geschwindigkeit von vorderem und hinterem Auto) mit dem Bestand (Abstand der Autos) zusammenhängt. Die Versuchspersonen schnitten hierbei kaum besser ab, als beim Bearbeiten anderer BF-Größen (vgl. auch Kap.1.1.5). Grundsätzlich ist Hogarth (2001) der Meinung, es muss eine „freundliche Lernumgebung“ vorliegen, damit Personen Intuitionen erlernen. Die Frage ist, ob das für die Auto-Aufgabe der Fall ist. Es muss beispielsweise für die Fahrenden oder beobachtenden Personen klar sein, wie schnell die zwei Autos fahren, und wie sich in Abhängigkeit davon der Abstand der Autos verändert – ist das in Realsituationen tatsächlich gut zu erkennen, so dass Personen diesbezüglich gute Intuitionen erlangen? Unter Umständen ist es nicht der günstigste Fall, um sich intuitives BF-Verständnis anzueignen. Dann

kann aus diesem Experiment nicht eindeutig geschlossen werden, dass keine Intuitionen für BF-Systeme vorhanden sind.

Für die vorliegende Arbeit gilt die Annahme, dass Intuitionen für BF-Systeme existieren können, denn es handelt sich um sehr alltägliche Konzepte, mit denen Personen im Alltag immer wieder konfrontiert sind. Möglicherweise existieren aus dem Alltag kommend Intuitionen, die in den bisherigen Studien zu BF-Szenarien jedoch nicht oder selten getriggert wurden und stattdessen kam es zu falschem Vorgehen im Sinne eines falschen Lösungsprozesses. In der vorliegenden Arbeit soll also das korrekte Lösen über eine adäquate Darstellung der BF-Aufgaben erreicht werden. Oft hat sich gezeigt, dass die externale Repräsentation eines Problems beeinflusst, wie gut das Problem gelöst wird.

1.2.3 Die externale Repräsentation beim Problemlösen

Sedlmeier und Hilton (2011) geben einen Überblick, inwiefern die externale Repräsentation eines Problems beeinflusst, wie gut es lösbar ist. Die Passung zwischen Dargestelltem und dem Darstellungsformat ist demnach für korrektes Urteilen oder Entscheiden relevant. Für BF-Studien wurde ein derartiger Ansatz bereits verfolgt: Cronin et al. (2009) variierten die externale Repräsentation des Szenarios (Wasser, Geschwindigkeit oder Perso-

nen). Jedoch führte diese Änderung zu keiner signifikanten Verbesserung.

In Bereichen mit anderen Problemen führten derartige Änderungen jedoch zu Verbesserungen in der Leistung, zum Beispiel beim logischen Schließen. Bei der klassischen Wason selection task (z.B. Wason & Shapiro, 1971) gibt es Karten, auf deren einer Seite Buchstaben und auf deren anderer Seite Zahlen stehen. Es existiert beispielsweise die Regel: „Jede Karte mit D auf der einen Seite hat eine 3 auf der anderen Seite“. Nur 30-40% der Versuchspersonen wählen die passenden Karten, um diese Regel korrekt zu überprüfen. Wenn der Inhalt zu einem „thematischen Inhalt“ umgewandelt wird, so verbessert sich die Leistung der Versuchspersonen. Beispielsweise sollte die Regel geprüft werden: „Wenn eine Person Alkohol trinkt, so muss sie über 20 Jahre (sprich volljährig) sein“. Hier lösten über 70% das Problem (Cosmides, 1989).

Auch bei Einsichtsproblemen kann die externale Repräsentation hilfreich sein, um schneller zum „Aha-Effekt“ zu kommen. Das zeigte sich beispielsweise beim Schachbrett-Problem (Kaplan und Simon, 1990). Versuchspersonen erhielten ein Schachbrett, bei welchem an den diagonal gegenüberliegenden Ecken jeweils ein Feld derselben Farbe nicht mehr vorhanden war. Die Versuchspersonen sollten versuchen, die restlichen Felder mit

Dominosteinen abzudecken oder logisch zu beweisen, dass das nicht möglich war. Tatsächlich ist zweiteres der Fall war: Es ist nur dann möglich, das gesamte Brett abzudecken, wenn gleich viele der verschiedenartigen Felder vorhanden sind. Denn es können immer nur Paare verschiedener Farben von den Dominosteinen abgedeckt werden. Versuchspersonen fanden sehr viel schneller die Begründung, wenn sie kein Schachbrett vorliegen hatten, sondern ein Brett auf dem für die hellen Felder „Butter“ und für die schwarzen Felder „Brot“ notiert war. Auf diese Art war schneller ersichtlich, dass es nicht aufgehen kann, wenn zwei Felder der gleichen Sorte, wie zwei Mal „Butter“, fehlt.

Im Bereich der Wahrscheinlichkeiten gibt es weitere Hinweise für den Einfluss der Problemrepräsentation in Bezug auf das genutzte Format. Sind Wahrscheinlichkeiten in der klassischen Weise als Prozentzahlen dargestellt, so ist der Umgang damit für die meisten Personen sehr schwierig (z.B. Kahneman & Tversky, 1972; Bar-Hillel, 1973). Hier besteht die Möglichkeit das Darstellungsformat zu ändern: 80% wird beispielsweise als „acht von zehn Personen“ ausgedrückt. Wenn statt Prozentzahlen oder Wahrscheinlichkeiten diese „natürlichen Häufigkeiten“ genutzt werden, so verbessert sich die Leistung im Umgang mit Wahrscheinlichkeiten. Es folgt ein Beispiel aus dem Bereich des Bayesianischen Folgerns, das im Allgemeinen sehr schwer zu

verstehen ist. In einer Studie wurde ein Problem (Mammographieproblem) einerseits mit Prozentzahlen gegeben, andererseits mit natürlichen Häufigkeiten. Die Rate der KorrektlöserInnen stieg hierbei von 16% auf 46% (Gigerenzer & Hoffrage, 1995). Über verschiedene Studien hinweg gab es eine Zunahme der korrekten Leistung von etwa 10-20% zu ca. 40-60% (Gigerenzer, Hertwig, Hoffrage & Sedlmeier, 2008). Warum hilft eine derartige Änderung der externalen Repräsentation? Eine Annahme ist, dass die Änderung in bestimmten Fällen „gute“ oder „valide“ Intuitionen (Begriff vgl. Sedlmeier, 2007) herbeiführt. Das bedeutet, dass in diesem Fall die Intuitionen zu korrektem Verständnis beziehungsweise Handeln führen. Die Erklärung hierfür sehen viele darin, dass im Laufe der Evolution der Mensch immer wieder mit natürlichen Häufigkeiten in Berührung kam und lernte damit umzugehen (z.B. Cosmides & Toobey, 1996; Sedlmeier & Gigerenzer, 2001).

Wie oben beschrieben (Abschnitt 1.1.5), wurde die externale Repräsentation der BF-Systeme bereits variiert: Bei Booth Sweeney und Sterman (2000) ergaben sich leichte Unterschiede in Folge der genutzten BF-Szenarien: Beim Rechteckverlauf schnitt die Wasserbecken-Aufgabe besser ab als die mit dem Konto. Bei Cronin et al. (2009) zeigten sich keine signifikanten Unterschiede in Abhängigkeit der BF-Größen Wasser, Geschwindigkeit oder

Personen. Möglicherweise helfen andere Modifikationen der Darstellung, valide Intuitionen herbeizuführen. Flussdiagramme sind wichtig, um die BF-Aufgaben zu bearbeiten. Die vorliegende Arbeit untersucht Modifikationen in den Flussdiagrammen genauer. Erkenntnisse hierzu beschreibt der nächste Abschnitt.

1.2.4 Diagramme

Nach Erkenntnissen von Cronin et al. (2009) schnitten die graphischen Bedingungen mit Säulen- oder Liniendiagramm tendenziell besser ab als die nicht-graphischen in Text- oder Tabellenform. Innerhalb der graphischen Bedingungen ergab sich kein nennenswerter Unterschied: Das Säulendiagramm schnitt nur tendenziell besser ab als das Liniendiagramm (detailliert in Abschnitt 1.1.5). Ein anderes Ergebnis berichteten Ossimitz, Kotzent-Pietschnig, Kreisler und Zoltan (2001). Dort lösten die Versuchspersonen eine diskontinuierliche Aufgabe in Tabelle besser als eine diskontinuierliche Aufgabe als Liniendiagramm. Es handelte sich jedoch um verschiedene Flussgrößen-Verläufe, weshalb dieser Vergleich nur als möglicher Hinweis zu werten ist. Zusammenfassend lässt sich sagen: Es ist unklar, ob graphische oder nicht-graphische Darstellungen günstiger sind. Insgesamt sind die meisten BF-Aufgaben als Diagramme, dann wieder

am häufigsten als Liniendiagramme dargestellt. Zudem zeigen sich durchgängig keine guten Lösungsraten.

MacDonald-Ross (1977) schreibt zu Fehlern bei Diagramm-Darstellungen: „... the most frequent and crucial experimental mistakes – [is] asking subjects to perform tasks inappropriate to the format or purpose of communication“ (S.363). Wenn nun also das klassische Flussdiagramm – gezeigt als Liniendiagramm – das falsche Format ist, um zur korrekten Lösung für den Bestand zu kommen, so besteht die Möglichkeit, dass eine andere Diagrammdarstellung günstiger ist, um BF-Aufgaben zu lösen. Die Modifikation von Cronin et al. (2009) könnte nicht ausreichend gewesen sein, um zu einer bedeutsamen Verbesserung der BF-Lösungen zu führen. Möglicherweise wirkt eine andere Art der Darstellung besser. Die Diagrammdarstellung bietet mehr Möglichkeiten, Diagramme zu verändern, als Liniendiagramme durch Säulendiagramme zu ersetzen. Beispielsweise ist es auch möglich, ein Balkendiagramm (mit vertikaler Zeitachse) zu verwenden oder andere graphische Darstellungen der Flussgrößen zu entwickeln.

Niegemann (2008, S.207, ff.) gibt eine Einordnung von verschiedenen Arten von Bildern. Darunter fallen beispielsweise realistische Bilder, die eine Abbildung von Objekten aus der realen Welt zeigen. Bei ihnen reichen Detailgenauigkeit und Realitäts-

grad von Fotos bis zu Piktogrammen. Ausschlaggebend ist, dass die Struktur von dargestelltem Objekt und dessen Repräsentation stark übereinstimmen. Eine andere Art von Bildern sind die logischen Bilder, zu denen die Diagramme gehören. Sie zeigen abstrakte Sachverhalte, wobei sie keine Ähnlichkeit mit dem aufweisen, was sie darstellen. Über ihre visuellen und räumlichen Eigenschaften transportieren sie Informationen. Nach Pinker (1990) mögen wir es, quantitative Informationen in graphischem Format zu zeigen oder zu lesen. Tatsächlich könnte alles auch nicht-graphisch dargestellt werden wie beispielsweise in Tabellen. Wir scheinen bestimmte Informationen jedoch graphisch besser verarbeiten zu können. Pinker erwähnt, dass Graphiken – in dieser Art – noch nicht sehr lange existieren. Seiner Meinung nach sind sie so günstig für die Kommunikation, weil sie generelle kognitive und Wahrnehmungs-Mechanismen effektiv nutzen. Es existiert also keine spezielle geistige Fähigkeit für das Graph-Lesen oder -Verstehen, die evolutionär entstanden ist, sondern andere Mechanismen werden hierfür genutzt.

Beispiele für Graphiken sind Diagramme wie das Liniendiagramm, Kreisdiagramm, Säulendiagramm, Balkendiagramm und vieles mehr. Für die vorliegende Arbeit werden das Säulen- und Balkendiagramm voneinander abgegrenzt (z.B. Zelazny, 1989). Das Säulendiagramm besitzt eine horizontale Zeit- oder Katego-

rienachse, während beim Balkendiagramm diese Achse vertikal ist. Verschiedene AutorInnen (z.B. Kosslyn, 1994; MacDonald-Ross, 1977; Shah & Hoeffner, 2002; Zelazny, 1989) betonen immer wieder, dass es nicht einfach die günstigste Darstellung für bestimmte Daten gibt, sondern dass die Hauptaussage aus den Daten beeinflusst, was eine günstige Darstellung ist, und somit welche Art von Diagramm gewählt werden sollte. Beispiele zeigen, wie eine einfache Änderung der Darstellung zu verschiedenen Interpretationen führt. Shah, Mayer und Hegarty (1999) beschreiben, wie die Modifikationen in Diagrammen zu sehr unterschiedlichen Interpretationen führen können. Beispielsweise waren die Daten einmal als Säulendiagramm gezeigt und im anderen Fall aufgeteilt in zwei Liniendiagrammen. Das heißt, die Art wie Informationen in Abbildungen gezeigt werden beeinflusst welche Größen die Diagrammlesenden vergleichen. In ihren Experimenten ergaben sich teilweise große Effektstärken für die Modifikationen der Diagramme.

Für die vorliegende Arbeit wird angenommen, dass das BF-Verständnis davon abhängt, wie die Diagramme der Flussgrößen gestaltet sind.

1.2.5 Fazit

BF-Systeme sind sehr schwierig zu verstehen – wie ist es möglich, das BF-Verständnis zu verbessern? Eine passende externale Repräsentation soll zu validen Intuitionen (sofern vorhanden) führen, wobei Intuition als erlernbar angesehen wird. Zudem stellt sie einen Teil von impliziten Prozessen dar, die beim Problemlösen oder beim Urteilen eine Rolle spielen. Bisher scheinen BF-Aufgaben oft nicht valide Intuitionen oder fehlerhaftes analytisches Vorgehen zu aktivieren. Eine Änderung der externalen Repräsentation von BF-Aufgaben soll dem abhelfen. Zwar wurden derartige Modifikationen schon erfolglos vorgenommen (Cronin et al., 2009), aber es gibt weitere Möglichkeiten, die externale Repräsentation zu ändern. Der Schwerpunkt der vorliegenden Arbeit liegt auf den klassischen Flussdiagrammen: Hier sollen die Flussgrößen klar ablesbar sein und zusätzlich sollen die Diagramm-Lesenden korrekt auf den Bestand schließen. Es handelt sich also nicht um eine reine Ableseaufgabe, sondern zusätzlich soll ein gegebenes Problem gelöst werden: Versuchspersonen sollen aufgrund gegebener Flussgrößen auf den Bestand schließen. Die Lösung soll durch das passende Format für die Flussgrößen erleichtert werden, so dass intuitiv der korrekte Prozess für das Lösen des Problems gefunden und angewandt wird.

2 Der Einfluss verschiedener Beschreibungsdimensionen auf das BF-Verständnis – Studie 1

Wie bereits beschrieben, zeigten viele Studien, dass BF-Systeme schwierig zu verstehen sind (Kapitel 1.1.3). Ein möglicher Ansatzpunkt ist es, die Darstellung der BF-Aufgaben zu ändern. Diese Kapitel untersucht verschiedene Modifikationen der Aufgabendarstellung.

2.1 Einleitung

In Kapitel 1.1.5 wurden verschiedene Beschreibungsdimensionen von BF-Aufgaben erläutert. Insbesondere die Dimensionen Darstellung von Flussgrößen und Abbildung des BF-Szenarios bieten sich an, um zu prüfen, ob beziehungsweise inwiefern eine bildliche Darstellung hilfreich ist. Des Weiteren können BF-Größen variiert werden. Bei BF-Studien ergaben sich diesbezüglich verschiedene Ergebnisse: Manchmal hatte die BF-Größe Einfluss und manchmal nicht (z.B. Booth Sweeney & Sterman, 2000; Cronin et al., 2009). Jedoch sind weitere Modifikationen möglich als die dort untersuchten. Daher soll die erste Studie prü-

fen, ob andere Modifikationen auf diesen Ebenen das BF-Verständnis beeinflussen.

2.1.1 Der Einfluss von Bildern

Domagk (2008, S.221, f.) fasst verschiedene Funktionen von Bildern zusammen, die einen Text oder Lernmaterial begleiten. Neben beispielsweise motivierenden oder dekorativen Funktionen betont sie die kognitiven Funktionen. Diese beinhalten auch die Aufmerksamkeitsfunktionen: Die Aufmerksamkeit wird auf relevante Informationen gelenkt. Außerdem können Bilder Aspekte zeigen, die schwierig in Worte zu fassen sind, und sie konkretisieren verbal gegebene Information (Konkretisierungsfunktion). Bilder zu Text oder Lernmaterial hinzuzufügen ist hilfreich (Levie & Lentz, 1982; Levin, Angelin, & Carnitiy, 1987). Mayer (2009) spricht vom Multimedia Prinzip: Für das Lernen ist es vorteilhaft, sowohl Text als auch Bild zu geben. Wenn die Bilder in Beziehung mit dem Text stehen, dann verbessern sie in der Regel die Leistung seitens der Betrachtenden. Auch nach Mayer wirken Bilder auf intuitiver Ebene, und es ist somit vorteilhaft, sie zu nutzen.

Wie in Kapitel 1.2.4 beschrieben gibt es verschiedene Arten von Bildern: Eine schematische Abbildung fällt unter die realistischen Bilder. Ein Diagramm ist ein logisches Bild, das abstrakte

Informationen illustriert (Domagk, 2008, S.209). Die vorliegende Studie untersucht, ob diese beiden Formen von bildlicher Darstellung hilfreich sind.

2.1.1.1 Schematische Abbildung des BF-Szenarios

Wenn eine BF-Aufgabe zu einem Szenario Wasserbecken gegeben ist, kann dieses Wasserbecken abgebildet und der Aufgabe beigefügt werden: Das stellt eine schematische Abbildung des Szenarios dar. Booth Sweeney und Sterman (2000) nutzten eine schematische Abbildung für eine Badewanne (Badewannen-Aufgabe) und eine Art Kassensystem für die Konto-Aufgabe. Sie untersuchten jedoch nicht systematisch, ob das Hinzufügen oder Weglassen einer derartigen Abbildung das BF-Verständnis erleichtert. In der vorliegenden Studie lagen die schematischen Abbildungen als abstrakte realistische Bilder vor, wobei sich ihr Design im mittleren Feld von realistisch befand. Das bedeutet, dass ein realistisches Bild einerseits nicht zu viele Details enthält, andererseits nicht zu simpel in der Darstellung ist. Es hatte sich als günstig für das Verständnis erwiesen, Bilder derart zu gestalten (Dwyer, 1975). Wie oben beschrieben sollten die zusätzlich gegebenen Abbildungen das Verständnis eines Sachverhaltes erleichtern. Somit ergab sich als erste Hypothese:

Hypothese 1.1: Die BF-Leistung verbessert sich, wenn eine schematische Abbildung des BF-Szenarios die gegebene Aufgabe ergänzt.

2.1.1.2 Darstellungsformat der Flussdiagramms

Eine weitere Beschreibungsdimension für BF-Systeme ist das Darstellungsformat für die Flussgrößen. In den meisten der Studien stellten die AutorInnen die Flussgrößen in einem Liniendiagramm dar: Das ist hier gemeint mit „klassische Flussgrößen-Darstellung“. Ist die klassische Darstellung tatsächlich hilfreich? Um das zu untersuchen, gab es eine graphische und eine nicht-graphische Version der Aufgabe; also eine Aufgaben-Beschreibung mit und eine ohne Flussgrößen-Diagramm.

Dies untersuchten Cronin et al. (2009) schon für die diskontinuierliche Kaufhaus-Aufgabe: Die AutorInnen verglichen Bedingungen mit Tabelle oder Text (nicht graphisch) mit einem Liniendiagramm oder einem Säulendiagramm (graphisch). Die Versuchspersonen schnitten in den graphischen Bedingungen besser ab. In der vorliegenden Arbeit wurden kontinuierliche Verläufe genutzt. Das erlaubte ein etwas anderes Vorgehen: Beim Rechteck- und Dreieckverlauf war es möglich, Zeiträume zusammenzufassen und nicht jeden Zeitpunkt einzeln zu beschreiben. Beispielsweise ist der Abfluss über die gesamte Zeit hinweg kon-

stant, und der Zulauf beim Dreieckverlauf nimmt in einem Zeitraum von 4 Minuten kontinuierlich zu, dann kontinuierlich ab, und dieses Muster wiederholt sich. Das erleichterte es, einen fließenden Text zu formulieren. Also existierte für jeden Flussgrößen-Verlauf ein Text. Verglichen wurde eine Bedingung mit den Flussgrößen in Form von *Text und Diagramm* mit einer Bedingung mit *nur Text*. Der Text ist in beiden Bedingungen identisch. Es bestand die Annahme:

Hypothese 1.2: Das Hinzufügen des klassischen Flussdiagramms verbessert die BF- Leistung.

2.1.2 Der Einfluss des genutzten Formats für das BF-Szenario

In der vorliegenden Arbeit wird das BF-Szenario beziehungsweise die darin genutzte BF-Größe auch als eine Beschreibungsdimension angesehen. Die BF-Größe kann eine kontinuierliche Größe sein wie Wasser oder eine diskrete Größe wie Personen. Booth Sweeney und Sterman (2000) haben schon eine diskrete Größe mit einer kontinuierlichen vergleichen. Im ersten Fall nutzten sie Geldeinheiten im Konto-Szenario, im anderen nutzten sie Wasser als BF-Größe. Hierbei schnitten Versuchspersonen beim Rechteckverlauf signifikant besser ab, wenn sie die

Aufgabe mit Wasser bearbeiteten. Die eigene nachträgliche Berechnung der Effektstärke ergab ein *r* von .24. Beim Dreieckverlauf ergab sich kein Unterschied. Cronin et al. (2009) verglichen die Lösungsgüte verschiedener BF-Szenarien: von Kaufhaus-, Badewannen- und Auto-Aufgabe. In ihrer Studie wurden zwei Szenarien als aus dem Alltag kommend und somit als vertrauter angesehen: die Badewannen-Aufgabe und die Auto-Aufgabe mit Geschwindigkeit der Autos als Flussgrößen und Abstand der Autos als Bestand. Zwar schnitten bei Cronin et al. die vertrauten Bedingungen kaum besser ab. Jedoch scheint es mir problematisch, daraus zu folgern, dass die Ausprägung der BF-Größe an sich keinen Einfluss hat. Das BF-Szenario mit Auto-Aufgabe wirkt sehr abstrakt, woraus resultieren könnte, dass diese Art der Änderung der BF-Größe nicht in starkem Ausmaß hilfreich war. Das heißt ebenfalls, dass möglicherweise eine andere Veränderung zu einer Verbesserung des BF-Verständnisses führt. Außerdem sind die wenigen Ergebnisse, die es bis dato zu dieser Fragestellung gab, eher widersprüchlich. Es zeigte sich vereinzelt unterschiedlich gute BF-Leistung je nach verwendeter BF-Größe. Somit soll in der aktuellen Studie ein Vergleich verschiedener BF-Größen vorgenommen werden. Die diesbezügliche Modifikation folgt jedoch etwas anderen Überlegungen.

Der Schwerpunkt in der Modifikation der BF-Größe liegt in der aktuellen Studie auf dem Vergleich kontinuierlich und diskret, wobei beide Größen konkret sein sollen. Mit konkret ist hier greifbar und nicht abstrakt wie Geschwindigkeit gemeint. Diese Modifikation ist stark orientiert an einer Änderung der Problem-Repräsentation bei Studien zu Wahrscheinlichkeiten (vgl. Kapitel 1.2.3). In diesen Studien bearbeitete ein Teil der Versuchspersonen Aufgaben mit Häufigkeiten, sprich konkreten Größen. Diese Versuchspersonen schnitten besser ab als die Personen, die mit Prozentzahlen arbeiteten. Das Format der Aufgaben führte also zu unterschiedlich guter Leistung.

Die vorliegende Studie berücksichtigte diese Ergebnisse insofern, dass die Eigenschaft von BF-Größe (diskret oder kontinuierlich) als möglicher Wirkfaktor angesehen wurde. Es sollte sich bei den verglichenen Größen um möglichst „neutrale“ Größen handeln. Murmeln stellten die neutrale und konkrete Größe dar, für die kontinuierliche Größe wurde wiederum Wasser genutzt. Das barg zusätzlich den Vorteil, dass die schematischen Abbildungen der BF-Szenarien ähnlich gestaltbar waren: In beiden Fällen handelte es sich um ein Becken mit Ab- und Zufluss. Die Ähnlichkeit der schematischen Abbildungen könnte die Wahrscheinlichkeit für den Einfluss aufgabenspezifischer Variablen senken. Diese könnten bei stärkeren Unterschieden der Abbildun-

gen zusätzlich auf die Ergebnisse wirken. Die Annahme zum BF-Szenario lautete:

Hypothese 1.3: Die BF-Leistung verbessert sich, wenn die BF-Größe im Szenario mit einer diskreten Größe dargestellt ist.

Es ergaben sich also zwei Arten schematischer Abbildungen: für ein Wasserbecken und für ein Murmelbecken. In das jeweilige Becken fließt Wasser oder rollen Murmeln hinein und heraus.

2.1.1 Fehleranalyse

Weiterhin interessierte das Fehlerverhalten der Versuchspersonen. Oft nutzten Versuchspersonen die Korrelations-Heuristik und kamen dadurch zu einer falschen Lösung (Booth Sweeney & Sterman, 2000; Cronin & Gonzalez, 2007; Cronin et al. 2009). Auch in der vorliegenden Studie wurde untersucht, ob die Versuchspersonen die Korrelations-Heuristik anwendeten, und ob die Gestaltung der BF-Aufgaben beeinflusste, wie oft sie sie anwendeten und an welcher Größe sich die Korrelations-Heuristik orientierte. Die anderen falschen Lösungen wurden dahingehend überprüft, ob weitere systematische Fehler ersichtlich waren. Als zusätzliche Frage ergibt sich:

Fragestellung 1.1: Treten systematische Fehlermuster auf?

2.2 Methoden

2.2.1 Teilnehmende

Levin et al. (1987) untersuchten über eine Meta-Analyse, wie gut Bilder einen Text unterstützen. Sie fanden eine Effektgröße von etwa .7 (Abstandsmaß). Dieser Wert wurde für die Poweranalyse genutzt, durchgeführt mit dem Programm G-Power (Version 3.1; Faul, Erdfelder, Lang, & Buchner, 2007. Auch für alle weiteren Studien wurde dieses Programm genutzt). Es wurde ein einseitiges α von 5% und eine Power von 85% festgelegt. Das ergab eine erforderliche Gruppengröße von 31 Personen für die Einzelvergleiche der einzelnen Bedingungen. Damit sich eine gleichmäßige Verteilung der einzelnen Bedingungen ergab, also in den Bedingungen mit und ohne Diagramm beziehungsweise mit oder ohne schematische Abbildung, wurde diese Zahl auf 32 pro Gruppe erhöht. Somit lagen für alle Kombination dieser beiden Variablen gleich viele Versuchspersonen vor. 64 Studierende nahmen teil, das mittlere Alter war 22 (SD = 3.8; Spannweite: 19 bis 43), 58 Personen (91%) waren weiblich. Die Mehrheit der Personen studierte Psychologie (75%), 22% studierten Erziehungswissenschaften, 2 Personen kamen aus anderen Studienrichtungen. Alle erhielten Versuchspersonenstunden für die Teilnahme.

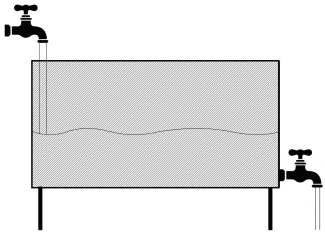
2.2.2 Materialien: die verwendeten Verläufe

Es handelte sich um ein 2x2x2-Design: Between variiert wurden die Abbildung des Szenarios (vorhanden oder nicht), ein Flussdiagramm (vorhanden oder nicht) und within variiert wurde die BF-Größe (Murmelszenario vs. Wasserszenario). Jede Person bearbeitete in Form eines Papier-und-Bleistift-Tests zwei Aufgaben: eine mit Rechteckverlauf und eine mit Dreieckverlauf; ein Beispiel hierzu befindet sich in Anhang B. Zu jeder Aufgabe erhielt die Versuchsperson einen ausführlichen Text, mit dem allein es möglich war, die Aufgabe zu lösen. Die abhängige Variable war die Lösungsgüte, sie ergab sich aus den gezeichneten Lösungsgraphen. Für die verschiedenen Bedingungen existierten die Abbildungen, die im Folgenden beschrieben werden.

2.2.2.1 Bedingung „schematische Abbildung des Szenarios“

Die Versuchspersonen erhielten zu der gegebenen BF-Aufgabe entweder eine Abbildung des BF-Szenarios, oder sie erhielten diese nicht. Die Bedingung schematische Abbildung war also between variiert. Abbildung 2.1 zeigt die BF-Szenarien: Ein Wasser- (a) und ein Murmelbecken (b).

(a)



(b)

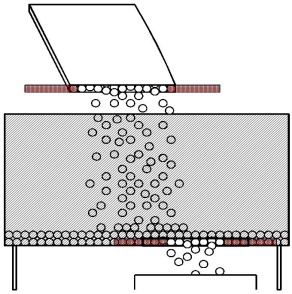


Abbildung 2.1 die schematischen Abbildungen vom Wasserbecken (a) und vom Murmelbecken (b).

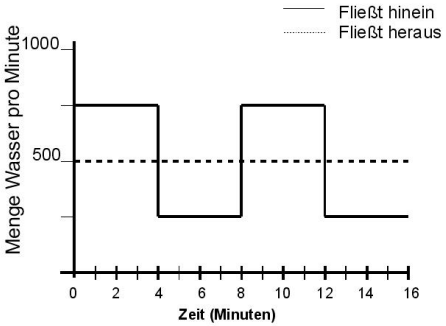
Das Murmelbecken sollte dem Wasserbecken möglichst ähnlich sehen, wobei der Zu- und Abfluss unterschiedlich dargestellt war. Das Problem war die Regelung der Menge für die beiden Flussgrößen. In diesem Fall entschied ich mich für eine Dar-

stellung, in der der Murmelzu- und -abfluss jeweils über eine Art Schleusenfunktion geregelt ist.

2.2.2.2 Bedingungen „Flussdiagramm“ und „BF-Szenario“

Für jede Aufgabe existierte ein Liniendiagramm, in dem Zu- und Abfluss abgetragen sind. In Abbildung 2.2 ist der Dreieck- und Rechteckverlauf zu sehen, ein Mal für das Wasser-Szenario (a) und ein Mal für das Murmel-Szenario (b). Die Werte, die Booth Sweeney und Sterman (2000) nutzten, sind mit 10 multipliziert. Eine Versuchsperson erhielt für ihre Aufgaben entweder nur den Text oder den Text und das entsprechende Flussdiagramm, diese Bedingung war ebenfalls between variiert.

(a)



(b)

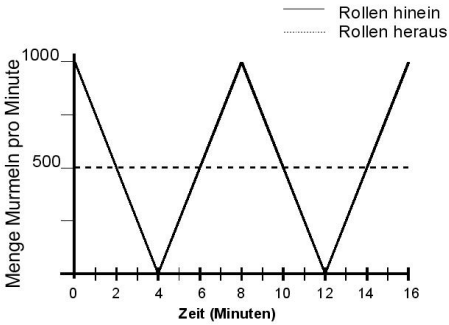


Abbildung 2.2 Beispiel für zwei Flussdiagramme einer Bedingung. BF-Größe ist für den Rechteckverlauf Wasser (a) und für den Dreieckverlauf Murmeln (b).

Jede Person erhielt beide BF-Szenarien, Murmel- und Wasserbecken, wobei ein Szenario mit Dreieckverlauf und das andere Szenario mit Rechteckverlauf kombiniert war.

2.2.3 Abhängige Variable

Abbildung 2.3 zeigt ein Lösungsdiagramm für das Wasser-Szenario. In allen Bedingungen war das Lösungsdiagramm entsprechend gestaltet. Mit einer Ausnahme: Für das BF-Szenario Murmelbecken ergab sich eine andere Beschriftung (Anzahl der Murmeln im Becken).

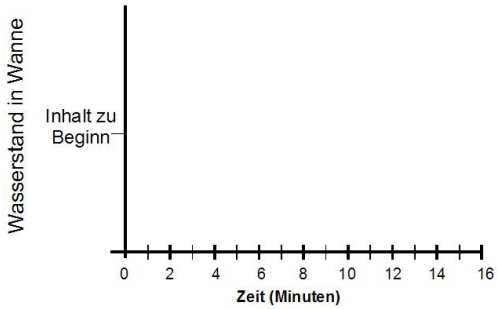


Abbildung 2.3 Lösungsdiagramm für das Wasser-Szenario. Das Lösungsdiagramm für das Murmel-Szenario sah gleich aus, mit Ausnahme der Beschriftung an der X-Achse. Dort stand: "Anzahl der Murmeln in Wanne".

Die Lösungsgüte wurde über sieben Kriterien beurteilt (Tabelle 2.1). Diese waren von Booth Sweeney und Sterman (2000) übernommen und wurden – wie bei ihnen – als intervallskaliert angesehen. Das gilt ebenfalls für die Lösungs-Kriterien der folgenden Studien. Für den Dreieckverlauf wurden von den ursprünglich 8 Kriterien, die Booth Sweeney und Sterman nutzten, das 6. und 7. Kriterium zusammengefasst zu einem. Somit lagen für beide Verläufe 7 Kriterien vor. Das Ziel war eine einfachere Vergleichbarkeit der beiden Aufgaben. Auch andere AutorInnen

änderten die Kriterien für ihre Studien ab (Lungwitz, Sedlmeier, & Schwarz 2018; Ossimitz et al., 2001; Schwarz, 2016).

Tabelle 2.1

Die Kriterien für die BF-Lösungsgüte von Rechteck- und Dreieckverlauf angelehnt und leicht modifiziert nach den Kriterien von Booth Sweeney und Sterman (2000).

Rechteckverlauf	***Dreieckverlauf***
Wenn Zufluss über Abfluss liegt, nimmt der Bestand zu [1]	
Wenn Abfluss über Zufluss liegt, nimmt der Bestand ab [2]	
Die Höchst- und Tiefstpunkte des Bestandes treten auf, wenn der Nettofluss gleich 0 ist:	
(t= 4, 8, 12, 16) [3]	(t = 2,6,10,14) [3]
Der Graph für den Bestand sollte keine diskontinuierlichen Sprünge enthalten (ist kontinuierlich) [4]	
In jedem Zeitabschnitt ist der Nettofluss konstant, der Bestand muss linear fallen oder steigen [5]	In jedem Zeitabschnitt ist Steigung des Bestandes gleich der Nettorate also a) Wenn Nettofluss positiv und fallend, nimmt Bestand mit abnehmender Geschwindigkeit/Rate zu (0<t<= 2; 8 < t =< 10) b) Wenn Nettofluss negativ und fallend, nimmt Bestand mit zunehmender Rate ab (2<t<=4; 10<t=<12)

Rechteckverlauf	***Dreieckverlauf***
	c) Wenn Nettofluss negativ und zunehmend, nimmt Bestand mit abnehmender Rate ab (4< t <= 6; 12 < t <=14) d) Wenn Nettofluss positiv und zunehmend, nimmt Bestand mit zunehmender Rate zu (6 < t =< 8; 14<= 16) [5]
Steigung von Bestand während jedes Zeitabschnittes ist Nettorate (d.h. +/- 25 Einheiten pro Minute) [6]	Steigung von Bestand wenn Nettorate am höchsten, ist 50 Einheiten pro Zeiteinheit (min) (bei t = 0, 8, 16) und Steigung von Bestand wenn Nettorate am niedrigsten, ist -50 Einheiten pro Zeiteinheit (min) (bei t = 4, 12) [6]
Die Menge, die während jedes Zeitabschnittes zum Bestand dazu kommt bzw. weggeht, ist die Fläche, die von der Nettorate eingeschlossen wird (z.B. 25 Einheiten pro Zeiteinheit * 4 Zeiteinheiten = 100 Einheiten), also ist der Höchststand des Bestandes bei 200 Einheiten erreicht und fällt bis zu einem Minimum von 100 Einheiten [7]	Die Menge, die während jedes Zeitabschnittes von 2 Zeiteinheiten zum Bestand dazu kommt bzw. weggeht, ist gleich der Fläche, die von der Nettorate eingeschlossen wird (d.h. ein Dreieck mit Fläche (1/2)*50 Einheiten pro Minute * 2 Minuten = +/- 50 Einheiten), also ist der Höchststand vom Bestand bei 150 Einheiten erreicht und fällt bis zu einem Minimum von 50 Einheiten [7]

Die gezeichneten Lösungen wurden jeweils von drei Personen beurteilt. Bei Differenzen in der Auswertung, ob ein Kriterium erfüllt war oder nicht, wurde nach Mehrheitsprinzip entschieden. Auf diese Art wurden in den folgenden Studien alle Auswertungen von gezeichneten Lösungen vorgenommen.

2.2.4 Ablauf

Aus dem vorliegenden Design ergaben sich acht verschiedene Versionen. Ein Beispiel für eine komplette Version des Aufgabenmaterials befindet sich in Anhang B. Die beiden Aufgaben wurden in unterschiedlicher Reihenfolge gezeigt, was zu 16 verschiedenen Versionen des Aufgabenmaterials führte (Anhang C). Die Versuchspersonen wurden diesen per Blockrandomisierung zugeordnet. Ein Ablauf für eine Versuchsperson sah beispielsweise so aus: Eine Versuchsperson wurde zufällig vom Versuchsleiter einer der 16 Bedingungen zugeordnet. Sie befand sich in der Bedingung mit schematischer Abbildung – ohne Flussdiagramm – mit Dreieckverlauf in Form des Wasser-Szenarios als erste Aufgabe. Die Versuchsperson erhielt eine allgemeine Instruktion, was den Ablauf des Experimentes betraf. Als nächstes bearbeitete sie ihre erste Aufgabe: ein Blatt mit ausführlichem Text zu den die Flussgrößen für das Wasser-Szenario und mit der schematischen Abbildung des Szenarios. In das Antwortdiagramm zeich-

nete die Versuchsperson ihren Lösungsgraphen ein. Dann meldete sie sich beim Versuchsleiter, der die bearbeiteten Unterlagen an sich nahm und ihr die zweite Aufgabe aushändigte (Aufgabe mit Zufluss in Form von Rechteckverlauf als Murmel-Szenario, wieder mit schematischer Abbildung und ohne Flussdiagramm). Wiederum bearbeitete sie die Aufgabe und meldete sich beim Versuchsleiter, wenn sie den Lösungsgraphen eingezeichnet hatte. Sie füllte noch ein Blatt mit demographischen Daten aus (Alter, Geschlecht, Studienfach) und mit weiteren Fragen einerseits zu den Aufgaben (welche Aufgabe war einfacher und warum), andererseits zu ihrem Vorgehen beim Lösen. Von den offenen Fragen werden Ergebnisse berichtet, wenn sie für die vorliegende Arbeit relevant sind. Komplette qualitative Analysen würden jedoch den Rahmen der Arbeit sprengen. Entsprechend wird bei den folgenden Studien vorgegangen.

2.1.2 Vorgehen bei der Analyse

Abbildungen wie Histogramme und Boxplots zeigen die Lösungsgüte in Abhängigkeit der Bedingungen für die explorative Datenanalyse. In einigen Fällen berichte ich Prozentangaben. Zum einen erfolgt die Angabe, wie viele Personen die Aufgabe *komplett korrekt* gelöst haben. Dabei handelt es sich um den Anteil der Personen, die alle sieben Kriterien korrekt erfüllten. Zum

anderen berichte ich den *mittleren prozentualen Anteil*. Hierfür berechnete ich den Mittelwert der BF-Leistung über alle Personen hinweg. Der maximal erreichbare Wert entsprach 100%. In der aktuellen Studie bearbeiteten die Versuchspersonen zwei Aufgaben, die beide maximal mit 7 Punkten bewertet wurden, somit entsprachen 14 Punkte 100%. Innerhalb dieser Skala wurde der Mittelwert für die BF-Leistung in einen Prozentwert umgerechnet. Diese beiden Prozentangaben berichte ich auch in den folgenden Studien. Inferenzielle Analysen erfolgen mit Regressionsanalysen. Diese werden mit SPSS (IBM, Version 24) durchgeführt. Zu den gerichteten Hypothesen werden einseitige Tests durchgeführt, die entsprechenden p-Werte werden somit halbiert. Die Partialkorrelationen (*pr*) sind entsprechend einem Maß für Effektstärken berichtet, denn diese entsprechen am ehesten der Effektgröße *r* (Sedlmeier & Renkewitz, 2013). Diese Werte sind auch in den folgenden Studien für die Regressionsanalysen angegeben.

Beispiel für die Auswertung über die sieben Kriterien

Es folgen Beispiele für verschiedene Lösungsgraphen, die bei Rechteckverlauf auftraten, und welche Punkte die Versuchspersonen für einen derartigen Lösungsgraphen erhielten (Abbil-

dung 2.4 bis Abbildung 2.6). Zunächst ein falscher Lösungsgraph, bei dem keines der Kriterien korrekt erfüllt ist (Abbildung 2.4).

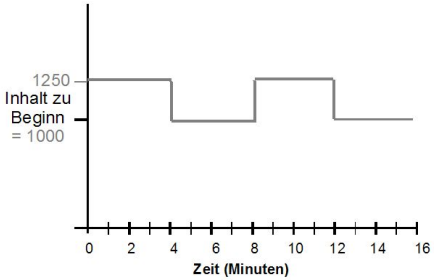


Abbildung 2.4 Dieser fehlerhafte Lösungsgraph ergibt 0 Punkte; Anwendung der Korrelations-Heuristik, orientiert an Netto-/Zufluss.

Die folgende Lösung (Abbildung 2.5) ist nicht komplett falsch, für sie gibt es Punkte, da einzelne der Kriterien erfüllt sind (Tabelle 2.1, dort sind Kriterien 1, 2, 3, 4, 5 korrekt; 6 und 7 nicht korrekt). Es ergeben sich für diese Lösung also 5 korrekte Kriterien. Abbildung 2.6 zeigt die korrekte Lösung.

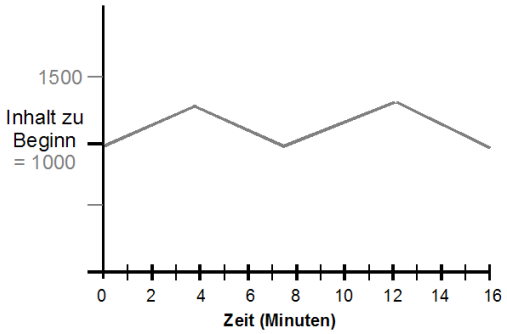


Abbildung 2.5 Fehlerhafter Lösungsgraph, der 5 Punkte ergibt.

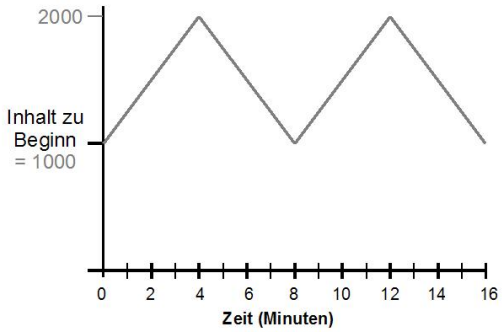


Abbildung 2.6 Die Lösung ist komplett korrekt, es ergeben sich 7 Punkte.

2.3 Ergebnisse

Zunächst beschreibe ich ein erstaunliches Ergebnis betreffend des Dreieckverlaufes. Es folgen Ergebnisse für den Überblick, anschließend gehe ich auf die einzelnen Bedingungen ein. Daraufhin folgt die Fehleranalyse und einige weitere Auffälligkeiten in den Daten.

Die erste Auswertung zeigte, dass nur vier Versuchspersonen die Aufgaben mit Dreieckverlauf komplett korrekt lösten. 57 Versuchspersonen (fast 90%) hatten kein einziges oder nur ein Kriterium erfüllt. Es lag also ein sehr starker Bodeneffekt für diesen Flussgrößenverlauf vor (siehe Abbildung 2.7). Somit werden die meisten Auswertungen nur für die Aufgabe „Rechteckverlauf“ durchgeführt.

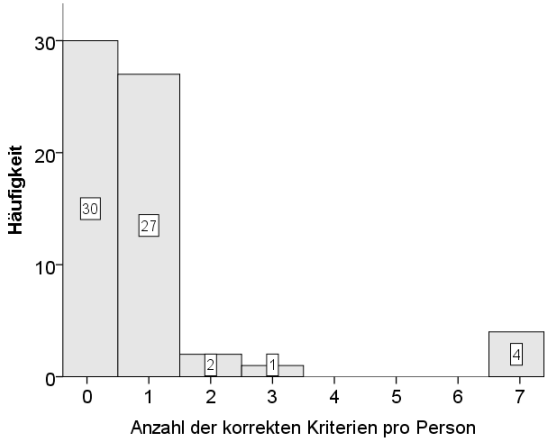


Abbildung 2.7 Die Säulen zeigen, wie viele Versuchspersonen die jeweilige Anzahl an korrekten Kriterien in ihrem Lösungsgraphen zum Dreieckverlauf erreichten.

2.3.1 Überblick: Anteile korrekter Lösungen und Regressionsanalyse

Als Mittelwert für die BF-Leistung (Gesamtsumme) ergab sich 4.92 (*SD* = 3.67). Das entspricht einem mittleren prozentualen Anteil von 35%. Für den Rechteckverlauf ergab sich eine mittlere prozentuale Leistung von 56% (Mittelwert: 3.95, *SD* = 3.00) und für den Dreieckverlauf eine mittlere prozentuale

Leistung von 14% (Mittelwert: .97, *SD* = 1.69). Der Anteil komplett korrekter Lösungen in der Stichprobe ergab für Rechteckverlauf 41% und für Dreieckverlauf 6%. Abbildung 2.8 zeigt, wie viele Versuchspersonen welche Werte für die Gesamtsumme erhielten. Hierfür wurden alle korrekt gelösten Kriterien für jede Person aufsummiert. Dadurch reichte die Gesamtsumme von 0 Punkten (kein einziges Kriterium bei einer der zwei Aufgaben korrekt) bis maximal 14 Punkte (beide Aufgaben wurden korrekt gelöst, somit sind alle Kriterien erfüllt). Abbildung 2.8 zeigt, dass 10 Versuchspersonen kein einziges Kriterium erfüllten und nur 2 Versuchspersonen beide Aufgaben komplett korrekt lösten.

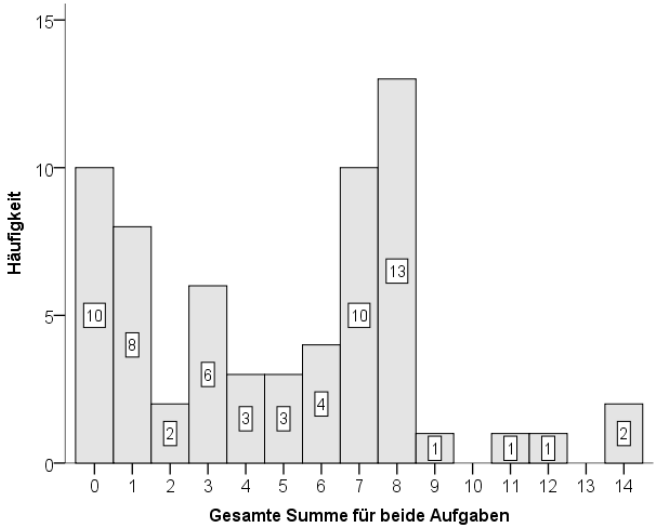


Abbildung 2.8 Anzahl der Versuchspersonen, die die jeweilige Gesamtsumme für beide BF-Verläufe erreichten (0-14).

Abbildung 2.9 zeigt Boxplots für die Bedingungen und für beide Flussverläufe. Wiederum ist die sehr schlechte Leistung beim Dreieckverlauf ersichtlich. Die Bedingung ohne Diagramm und mit schematischer Abbildung schnitt am besten ab, während die Bedingung mit Diagramm und ohne schematische Abbildung zur schlechtesten Leistung führte (bei Rechteckverlauf).

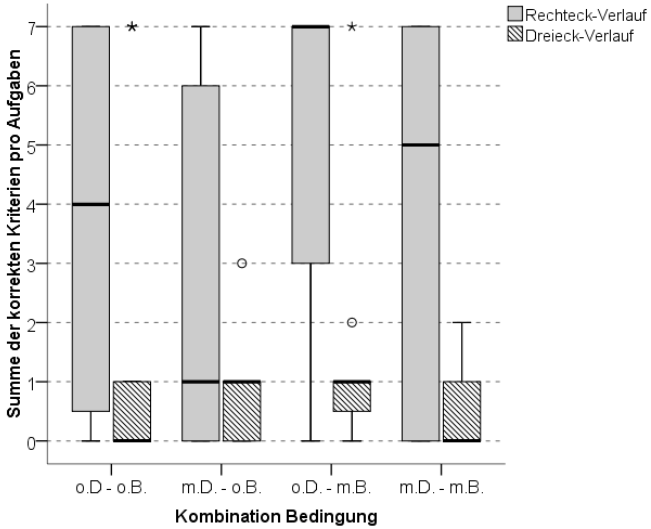


Abbildung 2.9 Boxplots für die Summe der Kriterien jeweils für beide Flussgrößen-Verläufe.
Anmerkung: m.D: Flussdiagramm vorhanden; o.D: Flussdiagramm nicht vorhanden; entsprechend m.B: mit schematischer Abbildung; o.B: ohne schematische Abbildung

Wegen der fehlenden Varianz in den Lösungen der Aufgabe mit Dreieckverlauf vergleiche ich im Folgenden die verschiedenen Bedingungen ausschließlich für die Aufgabe mit Rechteckverlauf. Es folgt das vollständige Regressionsmodell als Überblick

über die Regressionsanalyse mit den Werten für die einzelnen Prädiktoren (Tabelle 2.2). Hier und in den folgenden Studien enthält das „vollständige Modell“ alle Prädiktoren der Analyse in Abgrenzung zu vereinzelten Auswertungen, die nicht alle Prädiktoren enthalten. Auch wenn in vorliegender Untersuchung die Bedingungen mit Flussdiagrammen entgegen *Hypothese 1.2* die Leistung verschlechterten, war die Variable für eine detaillierte Auswertung weiterhin in der Regressionsanalyse enthalten.

Tabelle 2.2

Regressionsanalyse mit *pr*: Partialkorrelation; die abhängige Variable ist die Summe korrekter Kriterien für die Aufgabe mit Rechteckverlauf.

Prädiktor	β	*t*	*p*	*pr*
Abbildung	.24	1.992	.03	.24
Diagramm	-.20	-1.726	.05	-.21
BF-Größe/Szenario	.05	.398	.35	.05

Anmerkung: $R^2 = .10$ mit $F(3, 60) = 2.213$ mit $p = .10$ für das vollständige Regressionsmodell. Dummy-Kodierungen: mit Abbildung 1, ohne Abbildung 0; mit Diagramm 1, ohne Diagramm 0; Murmel 0, Wasser 1. Die *p*-Werte sind halbiert.

2.3.2 Bedingung "Schematische Abbildung des Szenarios"

Zunächst folgen die Ergebnisse zu *Hypothese 1.1* betreffend der schematischen Abbildung des Szenarios. Der mittlere prozentuale Anteil war mit schematischer Abbildung 67% und ohne 46%. In der Bedingung ohne Abbildung lösten 9 Personen (28%) die Aufgabe komplett korrekt. In der Bedingung mit schematischer Abbildung des Szenarios waren es 17 Personen (53%). Abbildung 2.10 zeigt Boxplots für den Vergleich der Gruppen mit und ohne schematische Abbildung.

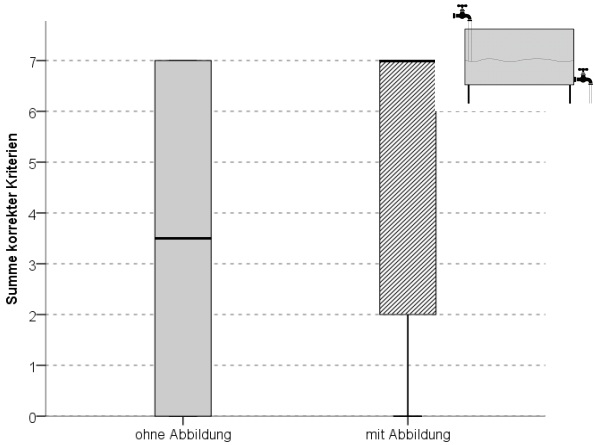


Abbildung 2.10 Boxplots für die Leistung der Gruppen mit vs. ohne schematischer Abbildung.

In der Regressionsanalyse ergab sich $\beta = .24$ und $pr = .24$ mit $p = .03$. Die Lösungsgüte verbesserte sich signifikant mit schematischer Abbildung des Szenarios.

2.3.3 Bedingung "Flussdiagramm"

Nach *Hypothese 1.2* sollte es hilfreich sein, wenn das Flussdiagramm zusätzlich zu dem ausführlichen Text gegeben ist. Der mittlere prozentuale Anteil lag ohne Diagramm bei 65% und mit

Diagramm bei 48%. Komplett korrekt gelöst haben die Aufgabe in der Bedingung ohne Flussdiagramm 15 Personen (47%) und mit Flussdiagramm 11 Personen (34%). Boxplots zeigen die Leistung in den beiden Gruppen (Abbildung 2.11).

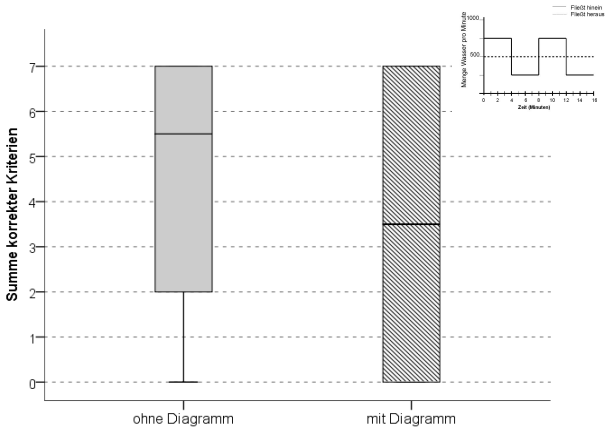


Abbildung 2.11 Boxplots für die Leistung der Gruppen mit versus ohne Diagramm.

Es ergab sich $\beta = -.20$, $pr = -.21$ und $p = .05$[4]. Das klassische Flussdiagramm verschlechterte die Lösungsgüte deutlich, wenn es zusätzlich zum Text vorliegt.

2.3.4 Bedingung "BF-Szenario"

Nach *Hypothese 1.3* sollte die BF-Leistung für eine Aufgabe mit diskreter BF-Größe besser ausfallen. Wenn das BF-Szenario ein Wasserbecken darstellte, so ergab sich ein mittlerer prozentualer Anteil für die Leistung von 54%, beim Murmelbecken lag er bei 58%. Beim Wasser-Szenario zeichnete 14 Versuchspersonen (44%) den Lösungsgraphen komplett korrekt und beim Murmel-Szenario waren es 12 Versuchspersonen (38%) (in beiden Fällen nur für Rechteckverlauf). Abbildung 2.12 zeigt die Boxplots für den Vergleich der beiden Gruppen.

4 Auch im Fall dass sich Leistung entgegen der angenommenen Richtung ändert, wird einseitig getestet und der entsprechende *p*-Wert berichtet.

Abbildung 2.12 Boxplots für die Leistung der Gruppen mit Murmeln versus Wasser als BF-Größe.

Es ergab sich $\beta = .05$ und $pr = .05$ mit $p = .35$. Die hier vorgenommene Änderung des Darbietungsformats schien zu keiner Veränderung in der Lösungsgüte zu führen.

2.3.5 Fehleranalyse

Nach *Fragestellung 1.1* war von Interesse, ob systematische Fehler auftraten. Ein aus der Literatur berichteter, systematischer Fehler ist die Verwendung der Korrelations-Heuristik. Somit

wurde zunächst untersucht, ob sich die Versuchspersonen an der Korrelations-Heuristik orientierten. Die Analyse der falschen Lösungen zeigte, dass das für einen Teil der Versuchspersonen der Fall war. Abbildung 2.13 zeigt für beide Flussgrößen-Verläufe die Anzahl der falschen Lösungen (schwarz umrandete Blöcke) und zusätzlich die Personen, die die Korrelations-Heuristik anwendeten (grau gefüllte Blöcke). Das ist jeweils unterteilt nach der Zugehörigkeit für die Bedingungen mit und ohne schematische Abbildung beziehungsweise Flussdiagramm. Die Prozentangaben in der Abbildung beziehen sich auf den Anteil (Korrelations-Heuristik verwendet) an den falschen Lösungen.

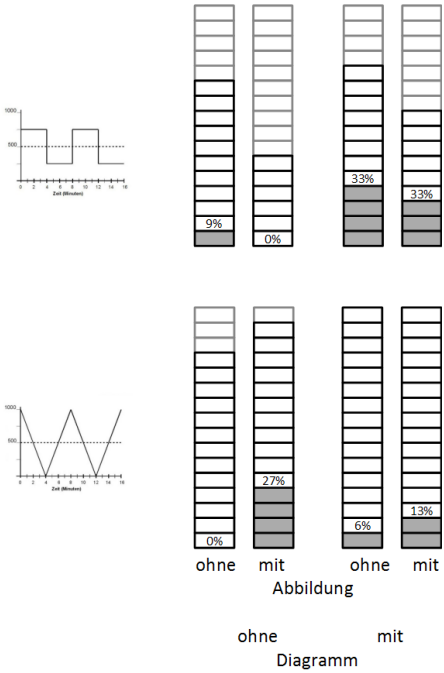


Abbildung 2.13 Anzahl der FalschlöserInnen (schwarz umrandete Rechtecke) und daran der Anteil derer, die Korrelations-Heuristik anwandten (grau gefüllte Rechtecke) – obere Reihe für Rechteckverlauf und untere für Dreieckverlauf.

Für Aufgaben mit Rechteckverlauf wurde die Korrelations-Heuristik häufiger verwendet (21%) als beim Dreieckverlauf (11%). In der Bedingung mit schematischer Abbildung verwendeten die Versuchspersonen die Korrelations-Heuristik häufiger (25% vs. 6% ohne schematische Abbildung). In der Bedingung mit Flussdiagramm kam sie häufiger vor (20% vs. 11% ohne Diagrsamm). Jedoch handelte es sich teilweise um sehr geringe Personenzahlen, weshalb diese Auswertung nur erste Hinweise geben kann. 12 Versuchspersonen nutzten Zufluss oder Nettofluss als Orientierungsgröße für die Korrelations-Heuristik, und drei nutzten den Abfluss. Weitere, systematische und fehlerhafte Lösungen, die oft auftraten wurden nicht entdeckt.

2.3.6 Zusätzliche Analysen: Positionseffekte in der Bedingung „BF-Szenario“

Unter Einbezug der Position des jeweiligen BF-Szenarios zeigte sich eine Besonderheit in den Ergebnissen (Abbildung 2.14): Je nach Position lösten die Versuchspersonen die Aufgabe in den verschiedenen BF-Szenarien unterschiedlich gut (wiederum nur auf Rechteckverlauf bezogen). Es ergab sich ein Positionseffekt: Das Wasser-Szenario wurde an 1. Stelle besser gelöst, das Murmel-Szenario besser an 2. Stelle. Wurden Position und die Interaktion von Position und BF-Größe mit in das vollständi-

ge Modell aufgenommen, so ergab sich für die Interaktion $\beta = .25$ mit $p = .02$ und $pr = .26$.

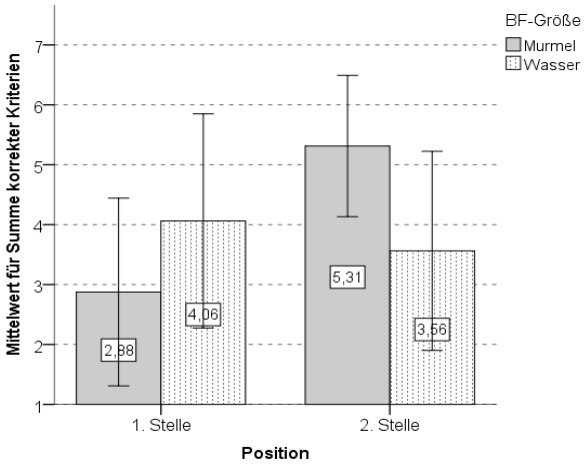


Abbildung 2.14 Mittelwerte und Konfidenzintervalle (95%) für Gesamtsumme in Abhängigkeit von BF-Größe und Position.

2.4 Diskussion

Bevor einzelne Aspekte der Studie näher betrachtet werden, gehe ich auf das erstaunlich schlechte Abschneiden der Versuchs-

personen ein. Das betrifft auf der einen Seite die Bedingung mit Diagramm und auf der anderen Seite die Aufgaben mit Dreieckverlauf. Die Lösungsgüte bei Rechteckverlauf liegt im mittleren bis unteren Bereich, wenn sie mit Werten aus vorhergehenden Studien verglichen wird. In den meisten Studien wurden gemittelte Prozente für alle Kriterien angegeben. Für die aktuelle Studie ergab sich ein gemittelter Wert von 56% gegenüber Werten aus vorhergegangenen Studien von 23% bis 83%, wobei Werte um 80% Ausnahmen darstellten (Armenia, Onori, & Bertini, 2004; Booth Sweeney & Sterman, 2000; Kapmeier et al., 2014; Ossimitz, 2002; Schwarz, 2016). Für den Rechteckverlauf ergab sich ein Anteil an komplett korrekten Lösungen von 41%. Die Aufgabe mit Dreieckverlauf lösten nur vier Personen komplett korrekt. In der Studie von Booth Sweeney und Sterman (2000) waren es 36% gewesen (Sterman, 2002). Die BF-Leistung in vorliegender Studie ist für den Dreieckverlauf viel schlechter als in vorhergehenden Studien. Dasselbe zeigte sich, wenn die gemittelten Prozentwerte für die Kriterien betrachtet werden. Diese Werte schwankten in bisherigen Studien von 28% bis 57% (Armenia et al., 2004; Booth Sweeney & Sterman, 2000; Kapmeier et al., 2014; Lungwitz et al., 2018; Ossimitz, 2002; Schwarz, 2016). In der vorliegenden Studie ergab sich 14% – woher kommt ein derart schlechtes Ergebnis?

Nach bisherigen BF-Studien schnitten Frauen schlechter ab als Männer (Booth Sweeney & Sterman, 2000; Kasperidus, Langfelder, & Biber; Kainz & Ossimitz, 2002; Schwarz, 2016; Sterman, 2010). Ossimitz (2002) gab in seiner Untersuchung die gemittelten Prozentwerte für Männer und Frauen getrennt an, wobei sich für Frauen 16% ergaben. Dieser Wert lag in einem vergleichbaren Bereich wie in der vorliegenden Studie. Ebenfalls zeigte sich vereinzelt ein Einfluss von mathematischer Vorbildung auf die BF-Leistung. Mit mathematischer Vorbildung ist gemeint, dass die Personen Fächer studierten, in denen ein hoher Anteil an mathematischen Inhalten vorlag. Ein Indiz hierfür ist, dass MIT-Studierende aus der Untersuchung von Booth Sweeney und Sterman (2000) bessere BF-Leistung zeigten als Studierende aus einer Studie von Ossimitz (2002). In der zweiten Studie kamen die Teilnehmenden von einer österreichischen Universität, die verschiedene Fächer mit eher geringeren Anteil an mathematischen Inhalten studierten. Außerdem zeigte sich in Experimenten, in denen ein direkter Vergleich möglich war, dass Studierende von naturwissenschaftlichen Fächern besser abschnitten als Studierende aus den Sozialwissenschaften (Booth Sweeney & Sterman, 2000; Kapmeier, 2004). In anderen Untersuchungen ergab sich jedoch kein Zusammenhang von Vorbildung und BF-Leistung (z.B. Sterman, 2010).

Es ist möglich, dass die Stichprobe zufällig vermehrt aus Personen mit schlechtem BF-Verständnis bestand. Tatsächlich sprach schon die Zusammensetzung der Versuchspersonen für eine Stichprobe, in der eher niedriges BF-Verständnis vorlag. Die aktuelle Stichprobe wies einen extrem hohen Frauenanteil auf (91%). Alle Versuchspersonen kamen aus nicht naturwissenschaftlichen Fächern, somit könnte geringere mathematische Vorbildung vorgelegen haben.

2.4.1 Bedingung "schematische Abbildung des Szenarios"

Die Daten unterstützten *Hypothese 1.1*: Eine schematische Abbildung des Szenarios verbesserte die Lösungsgüte. Tatsächlich zeigten Studien in anderen Bereichen, dass nicht in allen Fällen das Hinzufügen von Abbildungen von Vorteil war. Wenn eine Abbildung zu rein dekorativen Zwecken eingefügt war, verbesserte sich die Lernleistung nicht (Weidenmann, 2006, S.449). Das vorliegende Ergebnis passt zu Erkenntnissen aus dem Bereich Multimedia Learning: Zusätzliche Abbildungen erleichterten häufig das Verständnis/ Lernen, beispielsweise über kognitive Funktionen (vgl. Kapitel 2.1.1). In der vorliegenden Studie ergab sich als Effektstärke eine Partialkorrelation von .24. Die Partialkorrelation wurde gewählt, da sie am ehesten vergleichbar ist mit der

Einfachkorrelation aus Studien, in denen zwei Gruppen verglichen wurden (Sedlmeier & Renkewitz, 2013). Bei Schwarz (2016) ergaben sich vereinzelt Effektstärken, die etwa einem $r = .1$ entsprachen im Sinne einer Verbesserung gegenüber der Baseline-Bedingung. In den meisten bisherigen Studien sind keine derartigen Effektstärken angegeben, daher ist der Vergleich innerhalb des Forschungsgebietes schwierig. Nach Cohens Konventionen (1988) handelt es sich bei einer Korrelation von .24 um einen mittelgroßen Effekt.

Die Versuchspersonen in der vorliegenden Studie schienen aufgrund ihrer Zusammensetzung (weiblich und mathematisch eher wenig vorgebildet) besonders große Schwierigkeiten zu haben, die BF-Aufgaben zu lösen (vgl. vorhergehenden Abschnitt). Das könnte für den Rechteckverlauf bedeuten, dass das Hinzufügen einer schematischen Abbildung sehr gut wirkte. Denn in diesem Fall lag ein Anteil von 53% Korrektlösenden im hohen Bereich. Wie oben geschrieben ergab sich in der Studie von Booth Sweeney und Sterman (2000) ein höherer Anteil. Dort handelte es sich jedoch um eine Stichprobe mit größerem Anteil an Männern und mathematisch gut gebildeten Personen. In der Studie von Ossimitz lag ein höherer Männeranteil vor, und dennoch schnitten die Versuchspersonen dort schlechter ab als in der vorliegenden Studie die Gruppe mit schematischer Abbildung, wäh-

rend die Leistung für die Gruppe ohne schematische Abbildung etwa im gleichen Leistungsbereich lag.

Eine zusätzliche schematische Abbildung des Szenarios scheint also hilfreich zu sein. Bei Modifikationen aus vorangegangenen Studien zu BF-Aufgabenstellungen ergaben sich keine Hinweise, was konkret und in relativ allgemeiner Form zu ändern ist, um das Verständnis zu erhöhen. Mit einer zusätzlichen schematischen Abbildung liegt ein derartiger erster Hinweis vor.

2.4.2 Bedingung "Flussdiagramm"

Die Daten widersprachen klar der Annahme, dass das klassische Flussdiagramm hilft, die BF-Aufgaben besser zu lösen (*Hypothese 1.2*). Im Gegenteil: Die Versuchspersonen in der Bedingung mit Text und Graphik schnitten schlechter ab als in der Bedingung nur mit Text. Eine mögliche Ursache könnte das Kontiguitätsprinzip darstellen. Demnach sollen Text und Bild räumlich nah beieinander präsentiert werden (Domagk, 2008, S.230 f.), da sonst die kognitive Beanspruchung zu hoch sein könnte. In der vorliegenden Studie war die Anordnung von Text und Abbildungen immer gleich gehalten: Zuerst kam der Text, daraufhin folgten die Abbildungen. In den Bedingungen mit beiden Abbildungen kam zuerst die schematische Abbildung, dann das Flussdiagramm. Da nicht alles auf ein Aufgabenblatt passte, war in ei-

nigen Fällen das Flussdiagramm auf einem zweiten Blatt gegeben, was nach dem Kontiguitätsprinzip ungünstig ist. Andererseits bestand für die Versuchspersonen die Möglichkeit, die Blätter nebeneinander zu legen, so dass sie Text und Abbildungen gleichzeitig betrachten konnten. Das sollte zumindest einer zu hohen kognitive Beanspruchung abhelfen.

Weiterhin erstaunlich ist die Tatsache, dass Cronin et al. (2009) ein anderes Ergebnis fanden. In ihrer Studie schnitten die Versuchspersonen in den graphischen Bedingungen besser ab als die in den nicht-graphischen. Existiert ein grundlegender Unterschied zwischen den beiden Studien, der diese Diskrepanz erklärt? In der vorliegenden Studie war in der graphischen Bedingung der Text gegeben, während bei Cronin et al. (2009) in der graphischen Bedingung ausschließlich ein Diagramm vorlag. Nach dem Multimedia Prinzip (Mayer, 2009) sollte das Geben von Text und Diagramm von Vorteil sein. Jedoch kann die Kombination aus Text und Bild manchmal zu schlechterer Leistung führen, beispielsweise wenn ExpertInnen redundante Information aus Bild und Text erhalten. Das wird „general redundancy Effekt“ oder „expertise reversal Effekt“ genannt. Wenn jedoch Personen wenig Vorwissen haben, kann die Kombination aus Text und Bild das Verstehen erleichtern (Schnotz, 2005, S.63; Kalyuga, 2005, S.327; Kalyuga, Chandler, & Sweller, 2000). Wie zu

Beginn der Diskussion beschrieben, sollte es sich bei der vorliegenden Stichprobe nicht um eine Gruppe mit hoher BF-Expertise handeln. Somit sollte der Unterschied nicht aufgrund des general redundancy Effektes auftreten.

Ein weiterer Unterschied zu dem Vorgehen von Cronin et al. (2009) lag im genutzten Flussgrößen-Verlauf. Cronin et al. gaben eine Aufgabe mit diskontinuierlichen Verlauf, weshalb bei den nicht-graphischen Bedingungen alle Zeitpunkte einzeln aufgelistet waren. In der vorliegenden Studie fasste der Text Zeiträume zusammen. Beispielsweise ist bei der Aufgabe mit Rechteckverlauf der Zufluss über ein Intervall von vier Minuten konstant, der Abfluss sogar über den kompletten Zeitraum hinweg. Das ergibt eine andere Textstruktur und könnte eine Art Chunking darstellen, was einen Vorteil in der nicht-graphischen Bedingung darstellen könnte. Das könnte ein Hinweis sein, warum die Versuchspersonen der nicht-graphischen Bedingungen mit kontinuierlichem Flussgrößen-Verlauf besser abschnitten als die Versuchspersonen mit diskontinuierlichem Verlauf wie in der Studie von Cronin et al.

Zudem fand Schwarz (2016) ähnliche Ergebnisse wie die hier vorliegenden. Er verglich BF-Aufgaben, bei denen die Flussgrößen nur in Textform, als Text und Diagramm und nur in Diagrammform gegeben waren. Die Leistung verschlechterte sich in

der graphischen Bedingung zwar nicht derart auffällig wie in der vorliegenden Studie. Ein klassisches Flussdiagramm führte jedoch ebenfalls nicht zu besseren Ergebnissen in der Lösungsgüte. Auch Fischer, Degen und Funke (2015) berichteten, dass eine Flussgrößen-Darstellung als Text vorteilhaft ist.

Was nun genau die Ursache für den Unterschied zu den anderen Studien war – ob Zufall, die Zusammensetzung der Stichprobe oder ein weiterer nicht bekannter Grund – ist nicht klar zu sagen. Jedoch ist die Befundlage bis zu diesem Zeitpunkt recht eindeutig: Selbst wenn sich das Ergebnis mit Diagramm nicht verschlechterte, so zeigten die gerade beschriebenen Studien, dass ein klassisches Flussdiagramm die BF-Leistung nicht verbesserte. Es scheint somit sinnvoll, das klassische Flussdiagramm genauer zu untersuchen. Denn je nach Gestaltung eines Diagramms können unterschiedliche Interpretationen gefördert werden (vgl. Kapitel 1.2.4).

2.4.3 Bedingung "BF-Szenario"

Die Daten stützten zunächst nicht die Annahme, dass sich die BF-Leistung erhöht, wenn eine diskrete BF-Größe vorliegt (*Hypothese 1.3*). Grundlegend ergab sich kein Unterschied, je nachdem welches Format für die BF-Größe genutzt wurde. Das widersprach einigen Ergebnissen aus dem BF-Forschungsbe-

reich. Brunstein et al. (2010) fanden unterschiedlich gute BF-Leistung in verschiedenen BF-Szenarien. Der Flussgrößen-Verlauf war immer identisch, das Szenario war aus dem medizinischen Bereich, beispielsweise der Glucose-Gehalt des Bluts in Abhängigkeit von Glucagon und Insulin, oder eher allgemein gehalten beispielsweise das Kaufhaus-Szenario. Medizin-Studierende und Studierende anderer Fächer lösten den diskontinuierlichen Flussgrößen-Verlauf mit sechs verschiedenen BF-Szenarien. Die Annahme der AutorInnen war, dass die Medizin-Studierenden in BF-Szenarien aus dem Bereich Medizin besser abschneiden als die Studierenden anderer Fächer. Diese Annahme bestätigte sich nicht. Andererseits variierte die Lösungsgüte teilweise sehr stark für die verschiedenen BF-Szenarien. Die höchsten Anteile an korrekten Lösungen ergaben sich für die Kaufhaus-Aufgabe und eine Aufgabe zum Gewicht, das sich in Abhängigkeit von konsumierter und verbrauchter Energie veränderte. Ähnliches fanden Kapmeier et al. (2014). In ihrer Studie war eine Aufgabe (Dreieckverlauf) immer mit dem Konto-Szenario gegeben. In drei verschiedenen Gruppen schwankte die Lösung von 35% bis 41%. Die andere Aufgabe (Rechteckverlauf) war mit unterschiedlichen BF-Szenarien gegeben. Am besten schnitt das Szenario ab, das ein Bewerbungsportal darstellte (mit Bewerbungen als Flussgrößen). Es folgte das klassische Badewannen-Szenario und am

schlechtesten schnitt das Auto-Szenario ab (mit Geschwindigkeit als Bestand und Beschleunigung und Verzögerung als Flussgrößen). In Zahlen ergab sich: 72%, 47% und 29%. In dieser Studie zeigte die genutzte BF-Größe ebenfalls Einfluss auf die Leistung.

In der Studie von Booth Sweeney und Sterman (2000) zeigten sich unterschiedliche Ergebnisse. Beim Rechteckverlauf unterschied sich die Leistung je nach Szenario (Badewanne schnitt besser ab als Konto). Beim Dreieckverlauf unterschied sie sich nicht (detaillerter vgl. Kapitel 1.1.3). Das zweitere entsprach den Ergebnissen von Cronin et al. (2009), bei denen sich die Leistung nicht aufgrund einer Änderung der BF-Größe verbesserte (z.B. bei Auto-Aufgabe vs. Kaufhaus-Aufgabe). Es ist also unklar, inwiefern BF-Szenario und BF-Größe auf die Leistung wirken. Weitere Studien wären sinnvoll, da bei diesen starken Schwankungen zumindest ein Einfluss von BF-Szenario beziehungsweise der BF-Größe nicht auszuschließen ist. Diese können unter Umständen aufzeigen, in welchen Fällen und wie diese Beschreibungsdimensionen wirken.

In der vorliegenden Studie ergab sich eine Auffälligkeit bei der Variable BF-Größe, wenn zusätzlich die Position berücksichtigt wurde. Die Versuchspersonen lösten die Murmel-Aufgabe besser an 2. Stelle, dann sogar besser als die Wasser-Aufgabe (egal an welcher Position). Eventuell war es ungünstig, die Mur-

mel-Aufgabe an erster Stelle zu bearbeiten, da ein „Murmelbecken“ nicht gängig ist, und die Versuchspersonen es besser verstanden, wenn sie zuvor die Wasser-Aufgabe bearbeitet hatten. In diesem Fall könnte man tatsächlich von einem Positionseffekt sprechen. Zudem scheint dann die BF-Größe Murmeln besser verständlich als Wasser, wenn die Versuchspersonen das Prinzip des Szenarios verstanden hatten. Weitere Studien müssten klären, ob auf dieser Ebene tatsächlich Einfluss von der BF-Größe auf das BF-Verständnis vorliegt.

2.4.4 Fehleranalyse

Das systematische Fehlermuster Korrelations-Heuristik trat auch in der vorliegenden Studie auf. Somit war *Fragestellung 1.1* mit ja zu beantworten. Es ergab sich jedoch kein sehr hoher Anteil für das Anwenden der Korrelations-Heuristik. Cronin et al. (2009) untersuchten mit verschiedenen Flussgrößen-Verläufen speziell die Korrelations-Heuristik, im Mittel wurde sie bei 71% der falschen Lösungen angewandt (ebenfalls Studien, in denen die Lösung gezeichnet wurde). In ihrer Studie gab es drei Gruppen von Aufgaben, bei den einfachen Aufgaben war die Korrelations-Heuristik seltener angewandt, in den Gruppen mit mittelschweren bis schwierigen Aufgaben war Korrelations-Heuristik ungefähr gleich häufig verwendet (wobei der Schwierigkeitsgrad

über Relationen innerhalb der Studie ermittelt wurde). Das passt zunächst nicht zur vorliegenden Studie. Hier wurde bei der leichteren Aufgabe, bei Rechteckverlauf, häufiger die Korrelations-Heuristik angewandt. Andererseits könnte der Rechteckverlauf eher im Bereich der mittelschweren Aufgaben liegen, wenn Anteile der korrekten Lösungen bei Cronin et al. betrachtet werden.

In der vorliegenden Studie zeigte sich, dass bei Rechteckverlauf mit Diagramm Versuchspersonen häufiger die Korrelations-Heuristik anwendeten als in der Bedingung ohne Diagramm. Das könnte darauf hinweisen, dass das Liniendiagramm anregt, die Korrelations-Heuristik zu nutzen. Andererseits wandten Versuchspersonen bei Dreieckverlauf die Korrelations-Heuristik häufiger an, wenn kein Diagramm, sondern eine Abbildung vorlag. Zudem ist die jeweilige Gruppengröße sehr gering, wenn die Subgruppen getrennt nach allen Bedingungen vorlagen. Deswegen sollten in den folgenden Studien weitere Hinweise zum Anwenden der Korrelations-Heuristik gesammelt werden.

Beim Betrachten der Lösungen – insbesondere der Lösungen beim Dreieckverlauf – fiel auf, dass Versuchspersonen teilweise Lösungsgraphen zeichneten, die in sich keinerlei Systematik erkennen ließen. Gemeint ist folgendes: Booth Sweeney und Sterman (2000) erwähnten in ihrer Studie Fehler, die inkorrekt waren, aber einer nachvollziehbaren Systematik folgten. In der

vorliegenden Studie wirkten einzelne fehlerhafte Lösungen, als ob sich die Vorgehensweise innerhalb einer Aufgabe in den verschiedenen Zeitabschnitten verändert. Abbildung 2.15 zeigt ein derartiges Beispiel für den Dreieckverlauf. Das könnte ein Hinweis sein, dass die Aufgaben für diese Versuchspersonen so schwierig waren, dass sie überhaupt nicht wussten, wie sie vorgehen sollten.

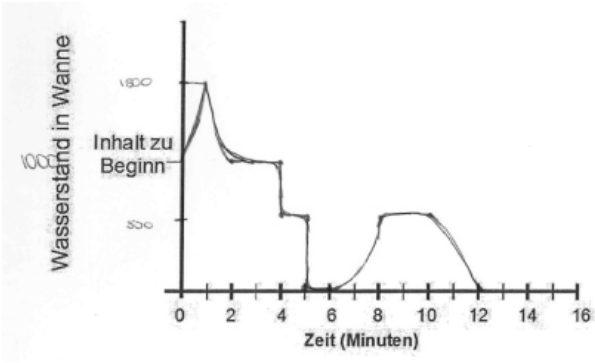


Abbildung 2.15 Fehlerhaften Lösung, in der keinerlei Systematik erkannt wurde.

Eingangs war beschrieben, dass die Versuchspersonen beim Bearbeiten des Dreieckverlaufs hier erstaunlich schlecht ab-

schnitten, während die Leistung beim Rechteckverlauf eher im normalen Bereich lag. Die Diskrepanz ist dahingehend erstaunlich, dass die selbe Stichprobe beide Aufgaben bearbeitete. Eventuell konnten Vorteile nur bei der Aufgabe mit Rechteckverlauf greifen, während die Nachteile die Leistung beim Dreieckverlauf noch stärker verschlechterten und dazu führten, dass die Versuchspersonen keinerlei Ideen für das Vorgehen hatten.

2.4.5 Fazit

Eine schematische Abbildung des Szenarios war hilfreich, also ist in allen folgenden BF-Studien eine schematische Abbildung zusätzlich gegeben. Das klassische Flussdiagramm ist nicht hilfreich. Somit soll die Darstellung von Flussgrößen modifiziert werden, um besseres BF-Verständnis herbeizuführen. Auf weitere Analysen betreffend der BF-Größe wird zunächst verzichtet. Was jedoch von nun an kontrolliert werden soll, ist der Einfluss von Geschlecht und mathematischer Vorbildung.

3 Modifikation des Flussdiagramms und Einbeziehung des Nettoflusses – Studie 2

Auch die vorhergehende Studie zeigte – so wie viele weitere Studien zuvor – dass das BF-Verständnis eher gering ausgeprägt ist. Dieses Kapitel handelt von weiteren Modifikationen, die zu verbesserter BF-Leistung führen sollten.

3.1 Einleitung

Die klassische graphische Repräsentation von Flussdiagrammen könnte problematisch sein. Cronin und Gonzalez (2007) vermuteten, dass sie zu Problemen beim Bearbeiten von BF-Szenarien führt. Cronin und Gonzalez untersuchten verschiedene diskontinuierliche Flussgrößen-Verläufe und fanden heraus, dass bestimmte Muster beziehungsweise Punkte in den Verläufen die falschen Antworten nach dem Bestand stark beeinflussen. Demnach ergaben sich Probleme beim BF-Verständnis aus der klassischen Flussdiagrammdarstellung. Wenn graphische und verbale Darstellung verglichen wurden, so führten Bedingungen mit klassischem Flussdiagramm nicht zu einer besseren Lösungsgüte (Kapitel 2.3.3 bzw. aus der Literatur vgl. Kapitel 2.4.2).

Die Forschung zu Diagrammen im Allgemeinen zeigt: Diagramme sind sehr hilfreich, wenn die Passung zwischen Darstellung und Dargestelltem günstig umgesetzt ist (Kosslyn 1994; Shah und Hoffner, 2002). Für BF-Studien existieren einige Versuche, das Flussdiagramm zu verändern. Cronin et al. (2009) zeigten ihren Versuchspersonen ein Säulendiagramm statt des klassischen Liniendiagramms für die Flussgrößen. Diese Änderung half kaum, die Verbesserung in der BF-Leistung bei gegebenem Säulendiagramm war nicht signifikant. Kann das Diagramm-Design derart geändert werden, dass der Anteil an korrekten Lösungen stark zunimmt? Falls ja, welche Änderung führt zu einer derartigen Verbesserung?

Es wird vorgeschlagen, aus systematisch auftretenden Fehlern Rückschlüsse zu ziehen, woher die fehlerhaften Lösungen kommen (z.B. Piaget, zitiert nach Ben Zeev & Star, 2001). Das Anwenden der Korrelations-Heuristik ist in bisherigen Studien das einzige systematische und häufig auftretende Fehlermuster, von dem berichtet wurde. Die Versuchspersonen orientierten sich oft am Nettofluss oder am Zufluss. Es gab erste Hinweise, dass die Versuchspersonen am häufigsten Nettofluss, seltener den Zufluss oder sehr selten den Abfluss zur Orientierung für die Korrelations-Heuristik nutzten (Cronin & Gonzalez, 2007; Cronin et al.

2009). Das folgende Experiment versucht aus diesen Fehlern hilfreiche Modifikationen für die BF-Aufgaben abzuleiten[5].

3.1.1 Modifikation des Flussdiagramms – Teilung des Koordinatensystems

Wenn Versuchspersonen sich beim Nutzen der Korrelations-Heuristik am Zufluss orientieren, scheinen sie einen Zusammenhang zwischen den Mustern des Zuflusses und des Bestands anzunehmen. Cronin und Gonzalez (2007) folgerten aus einem ihrer Experimente, dass es für die Versuchspersonen möglicherweise schwieriger ist, den Abfluss zu berücksichtigen oder zu verarbeiten. Falls einige Personen tatsächlich den Abfluss nicht berücksichtigen, könnte für sie ein im Diagramm betonter Abfluss hilfreich sein, so dass sie diesen entsprechend dem Zufluss wahrnehmen und/oder verarbeiten. Zelazny (1989) zeigte, wie es möglich ist, negative Werte in einer graphischen Zeitreihe hervorzuheben. Er betonte den Gegensatz zwischen positiven und negativen Werten, indem er die negativen Werte an der Achse „umklappte“, sie also im negativen Bereich abtrug. In der vorliegenden Studie wurde das entsprechend für die Liniendiagramme der Flussgrößen umgesetzt. Ein Diagramm zeigt die Zeitreihen

5 An dieser Stelle herzlichen Dank an Maria Hollnagel, die im Rahmen ihrer akademischen Qualifikationsarbeit die Datenerhebung und -erfassung durchführte.

für Zufluss und Abfluss. Neu ist, dass sich der Abfluss im Quadrant Nummer 4 befindet (Abbildung 3.1). Dadurch wird hervorgehoben, dass der Abfluss negativ ist. Da nur die Quadranten 1 und 4 für die vorliegende Studie genutzt werden, wird für zweiteren die vereinfachte Bezeichnung „negativer Quadrant“ genutzt.

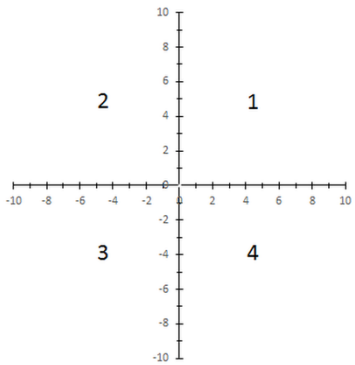


Abbildung 3.1 Kartesisches Achsensystem mit den nummerierten Quadranten von 1 bis 4.

Ossimitz (2001, 2002) betonte wiederholt, wie wichtig es ist, zwischen Fluss- und Bestandsgrößen zu unterscheiden und dass die Unfähigkeit dazu zu einem Großteil der Probleme führt. Demnach könnte es für das Lösen hilfreich sein, den Unterschied

erkenntlich zu machen. Wenn nun für das Flussdiagramm zwei Quadranten genutzt werden, sieht das Antwortdiagramm anders aus als das Flussdiagramm. Eventuell begünstigt dieser Unterschied das Erkennen, dass es sich hier um verschiedene Größen – Flüsse und Bestand – handelt.

Hypothese 2.1: Das Trennen von Zu- und Abfluss und Abtragen der Größen im „geteilten Koordinatensystem“, also im positiven und negativen Quadranten, führt zu besserer BF-Leistung.

3.1.2 Hinzufügen des Nettoflusses

Viele Versuchspersonen orientierten sich am Nettofluss, wenn sie die Korrelations-Heuristik anwendeten. Deshalb setzte hier ein weiterer Versuch an, das BF-Verständnis zu erhöhen. Eine Versuchsperson geht beim Lösen von BF-Aufgaben korrekt vor, wenn sie zunächst die Flussgrößen zum Nettofluss verrechnet und danach diesen zum Bestand hinzurechnet. Versuchspersonen, die sich jedoch am Nettofluss orientieren, verrechnen beide Flussgrößen und zeichnen diesen Verlauf als Lösungsgraphen. Eventuell sehen sie das Zwischenergebnis (den Nettofluss, also eine weitere Flussgröße) als Endergebnis nämlich als Bestand an. Das könnte beispielsweise am Anwenden eines falschen bezie-

hungsweise unvollständigen Schemas liegen (Gick, 1986). Der Begriff „Schema“ wird für die vorliegende Arbeit entsprechend genutzt wie in der Literatur zu mathematischer Intuition (Greeno, 1979). Es handelt sich um einen Rechenalgorithmus, um bei mathematischen Problemen zu einem Ergebnis zu kommen. Im vorliegenden Fall könnten die Versuchspersonen annehmen, dass die Rechenoperation „Zufluss + Abfluss“ die Lösung ergibt. In der Studie 1 aus vorliegender Arbeit schrieben einige Versuchspersonen zu ihrem Vorgehen, dass sie „Differenzen bilden“, welche dann ihren Lösungsgraphen ergaben. Möglicherweise hilft es, den Zwischenschritt Nettofluss vorzugeben, damit diese Personen weiter rechnen und das korrekte Endergebnis – den Bestand – erhalten. In anderen Bereichen ergaben sich Hinweise, dass beim Erlernen komplexer Fertigkeiten Zwischenschritte vorgeben hilfreich war (z.B. Nadolski, Kirschner, & Marriënboer, & Hummel 2001). Über diese vorgegebenen Zwischenschritte sollten Versuchspersonen günstigere Strategien entwickeln oder Operatoren wählen, die zur korrekten Lösung führen (Nadolski, Kirschner, & Marriënboer, 2005). Somit wurde der Zwischenschritt Nettofluss den Versuchspersonen gegeben. In der Aufgabenbeschreibung in dieser Bedingung war erklärt, wie der Nettofluss berechnet wird und dass es sich bei ihm um einen Zwischenschritt handelt. Es wurde jedoch nicht explizit beschrieben, wie der Nettofluss den

Bestand beeinflusst. Da der Schwerpunkt auf Diagrammdarstellung liegt, war der Nettofluss über ein weiteres Diagramm gegeben.

Es existierten verschiedene Möglichkeiten, den Nettofluss in die Aufgabenstellung zu integrieren. Entweder war der Nettofluss direkt vorgegeben oder die Versuchspersonen ermittelten ihn selber und zeichneten danach den Lösungsgraphen. Somit war vergleichbar, was günstiger ist: Den Nettofluss zu erhalten oder ihn selber zu ermitteln. Versuchspersonen, die zunächst den Nettofluss selber ermittelten, könnten die Aufgabe tiefer verarbeiten, während die Personen, denen der Nettofluss vorgegeben wurde, eventuell das Konzept nur oberflächlich verarbeiteten. Andererseits könnte die kognitive Beanspruchung steigen, wenn die Versuchspersonen den Nettofluss selber berechneten. Das könnte dazu führen, dass die Versuchspersonen schlechter abschneiden. Bei den Personen, die den Nettofluss ermittelten, überprüfte ich, ob sie ihn korrekt angaben. Zudem berechnete ich bedingte Wahrscheinlichkeiten: Die Wahrscheinlichkeit für einen korrekten Lösungsgraphen, wenn der Nettofluss korrekt versus wenn er falsch ermittelt wurde.

Hypothese 2.2: Das Vorgeben des Nettoflusses verbessert die BF-Leistung im Vergleich zur Baseline.

Hypothese 2.3: Das Selber-Ermitteln des Nettoflusses verbessert die BF- Leistung im Vergleich zur Baseline.

Fragestellung 2.1: Ergibt sich ein Unterschied in der BF-Leistung je nachdem ob der Nettofluss direkt vorgegeben ist oder ob Versuchspersonen ihn selber ermitteln?

3.1.3 Fehleranalyse

Wiederum soll eine Fehleranalyse stattfinden. Da bisher das einzige sich immer wieder zeigende Fehlermuster die Korrelations-Heuristik war, wurden Flussgrößen-Verläufe genutzt, anhand derer Cronin et al. (2009) bereits die Korrelations-Heuristik untersuchten. Wenn oben beschriebene Annahmen zutreffen, sollten die Versuchspersonen der modifizierten Bedingungen sich seltener an der Korrelations-Heuristik orientieren als die Versuchspersonen in der Baseline-Bedingung.

Hypothese 2.4: In der Baseline-Bedingung tritt die Korrelations-Heuristik als systematischer Fehler auf.

Hypothese 2.5: Es ergeben sich in den modifizierten Bedingungen weniger Lösungen orientiert an der Korrelations-Heuristik.

Fragestellung 2.2: Existieren weitere systematische Fehler?

3.1.4 Kontrollvariablen

In der vorhergehenden Studie zeigte sich ein erstaunlich schlechtes Ergebnis im Vergleich mit bisherigen Studien. An dieser Studie nahm ein sehr hoher Anteil Frauen teil, die Fächer studierten, in denen vergleichsweise wenig mathematische Inhalte gelehrt werden. Tatsächlich zeigte sich oft ein Geschlechtereffekt zugunsten der Männer, je nach Studie unterschiedlich stark (Kapitel 2.4). Außerdem scheint die mathematische Vorbildung Einfluss zu haben (ebenfalls Kapitel 2.4). Daher sollten diese Variablen in der aktuellen Studie als Kontrollvariablen erhoben werden. Weiterhin beantworteten die Versuchspersonen die Frage, ob sie schon einmal an BF-Studien teilgenommen hatten, um potenzielle Übungseffekte aufgrund mehrmaliger Teilnahme zu überprüfen.

3.2 Methoden

3.2.1 Teilnehmende

Für die damalige Abschlussarbeit wurde eine hohe Effektgröße (vgl. Kapitel 1.2.4) für den Unterschied für die Bedingung mit geteiltem Koordinatensystem erwartet (Hauptfragestellung der Abschlussarbeit, getestet mit t-Test). Mit einseitigem α von .05 und Power von 80% ergab sich für einen Vergleich von

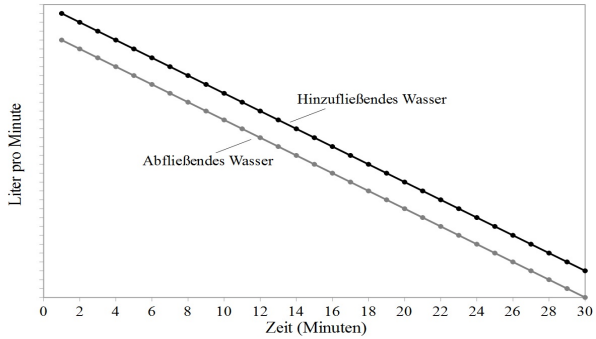
zwei Gruppen eine Anzahl von 21 Versuchspersonen pro Bedingung. Das ergab für die gesamte Stichprobe 84 Versuchspersonen. Das Alter reichte von 18 bis 41 Jahre, der Mittelwert betrug 22.58 (*SD* = .35). Für das Studienfach ergab sich folgende Verteilung: 68 der Versuchspersonen studierten Psychologie, 13 Erziehungswissenschaften und 3 Germanistik. Das bedeutet, im Vergleich sollten die Psychologie-Studierenden höhere Anteile mathematischer Inhalte im Studium haben, beispielsweise in Veranstaltungen zu Methodenlehre. Die Teilnehmenden erhielten Versuchspersonen-Stunden als Vergütung.

3.2.2 Materialien: die verwendeten Verläufe

Es handelte sich um einen Papier-und-Bleistift-Test. Die BF-Größe war in allen Fällen Wasser in Liter. Bei den Flussgrößen-Verläufen handelte es sich um drei Aufgaben aus der Studie von Cronin et al. (2009), zu denen die AutorInnen den Schwierigkeitsgrad berichteten. Für die vorliegende Studie wurden drei unterschiedlich schwierige Aufgaben ausgesucht. Zunächst sind die Aufgaben als klassische Liniendiagramme im ersten Quadranten gezeigt, wie sie in der Baseline-Bedingung vorliegen (siehe Abbildung 3.2a, 3.3a, 3.4a). Abbildung 3.2b bis 3.4b zeigen die entsprechenden Lösungen. Ihre vermutete Lösung zeichneten die

Versuchspersonen in ein Lösungsdiagramm, welches für alle Aufgaben und in allen Bedingungen gleich aussah (3.5).

(a)



(b)

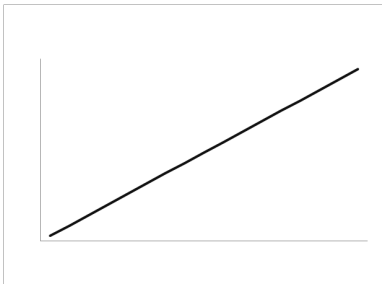
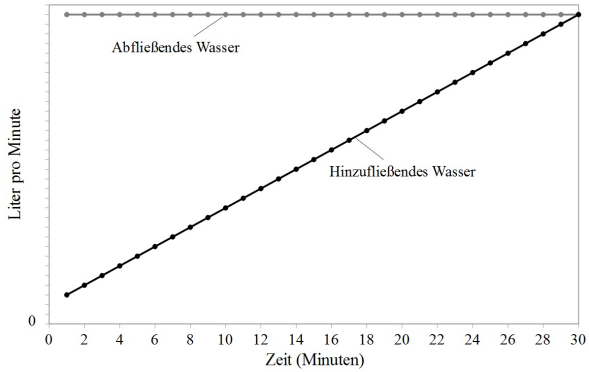


Abbildung 3.2 Flussdiagramm des leichten Flussgrößen-Verlaufes (a) und der Lösung (b) (übernommen und leicht geändert nach Cronin et al., 2009).

(a)



(b)

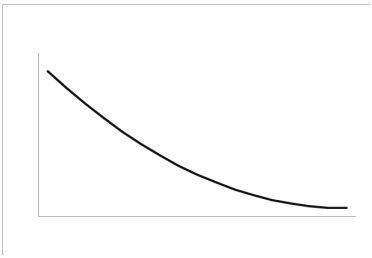
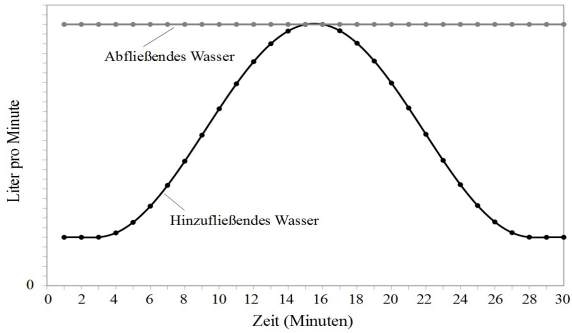


Abbildung 3.3 Flussdiagramm des mittelschweren Flussgrößen-Verlaufes (a) und der Lösung (b) (übernommen und leicht geändert nach Cronin et al., 2009)

(a)



(b)

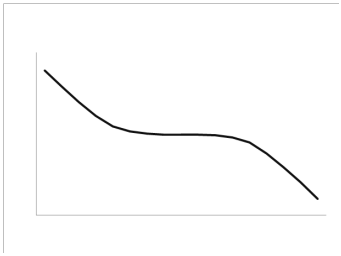


Abbildung 3.4 Flussdiagramm des schwierigen Flussgrößen-Verlaufes (a) und der Lösung (b) (übernommen und leichtgeändert nach Cronin et al., 2009).

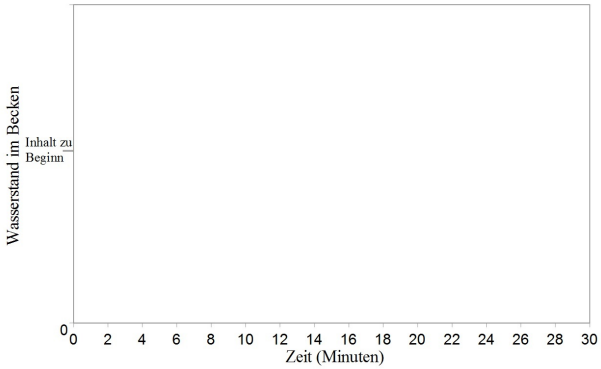


Abbildung 3.5 Lösungsdiagramm ohne eingezeichneten Lösungsgraphen für alle Flussgrößen-Verläufe.

3.2.3 Modifikationen, um eine Erleichterung des BF-Verständnisses herbeizuführen

Es existierten vier Bedingungen für ein between-subjects-Design: die Baseline-Bedingung als Kontrollgruppe mit klassischen Liniendiagramm, Koordinatensystem-geteilt, Nettofluss-gegeben und Nettofluss-ermitteln. In jeder Bedingung bearbeiteten die Versuchspersonen drei Aufgaben mit den oben gezeigten Flussgrößen-Verläufen. In allen Bedingungen war ein Flussdiagramm gegeben, aus dem die Versuchspersonen den Bestand er-

mittelten, den sie in das zugehörige Lösungsdiagramm zeichneten. Es folgt die Beschreibung der drei Bedingungen, die zu einer höheren BF-Leistung führen sollten. Anschließend gehe ich auf die abhängigen und weitere erhobene Variablen ein.

3.2.3.1 Die Bedingung „geteiltes Koordinatensystem”

Ist es hilfreich, das Flussdiagramm zu ändern? Es erfolgte eine Aufteilung des Flussdiagramms in einen positiven Bereich für den Zufluss und einen negativen Bereich für den Abfluss: die Bedingung „KS-geteilt“. Diese Aufteilung orientierte sich am klassischen Achsensystem mit zwei Quadranten für den positiven Bereich der X-Achse, auf der die Zeit abgetragen ist. Ausgehend von der X-Achse nehmen die Werte im oberen, positiven Quadranten nach oben hin zu und im negativen Quadranten nach unten hin zu. Der 0-Punkt liegt also für beide Quadranten auf derselben Linie. Abbildung 3.6 zeigt das Flussdiagramm mit geteiltem Koordinatensystem für den mittelschweren Flussgrößen-Verlauf.

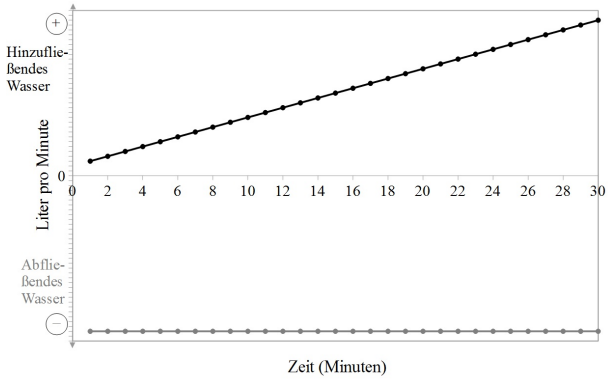


Abbildung 3.6 Flussdiagramm der Bedingung KS-geteilt für den mittelschweren Flussgrößen-Verlauf.

3.2.3.2 Die Bedingungen „Nettofluss-gegeben“ und „Nettofluss-ermitteln“

Nettofluss gegeben (NF-gegeben)

In dieser Bedingung erhielten die Versuchspersonen zusätzlich zum klassischen Flussdiagramm (Zu- und Abfluss in einem Quadranten) ein weiteres Diagramm, in dem der Nettofluss abgetragen war, inklusive einer Erklärung, was dieser bedeutet. Mittels dieser beiden Diagramme schlossen die Versuchspersonen

auf den Bestand und zeichneten den Lösungsgraphen in das ebenfalls gegebene Lösungsdiagramm.

Nettofluss ermitteln (NF-ermitteln)

In dieser Bedingung war das klassische Flussdiagramm gegeben und zusätzlich ein leeres Diagramm, in das der Nettofluss eingezeichnet wurde (das Nettofluss-Diagramm). Dieses sah grundsätzlich aus wie das Lösungsdiagramm, es gab jedoch zwei Unterschiede: Zum einen war die Beschreibung an der Y-Achse „Liter pro Minute“, zum andern fehlte der Startpunkt, also der Bestand zu Beginn. Die Versuchspersonen wurden aufgefordert, den Punkt von 0 Litern selber zu markieren. Zusätzlich erhielten die Versuchspersonen eine Erklärung, was der Nettofluss bedeutet und wie sie ihn ermitteln. Eine Versuchsperson zeichnete zunächst den Nettofluss in das Nettofluss-Diagramm und anschließend den Bestand in das Lösungsdiagramm ein.

3.2.4 Abhängige Variable

Zwei Kriterien wurden festgelegt, um die Lösung für den Bestand zu beurteilen. Diese Form von Auswertung orientierte sich an der Auswertung vom Experiment nach Cronin et al. (2009), aus dem die Aufgaben für diese Studie übernommen wurde. Zum einen gab es das Kriterium „AV-Tendenz“. Wenn eine

Versuchsperson korrekt einschätzte, wann der Bestand zu- und wann er abnimmt beziehungsweise wann er gleich bleibt, dann ist dieses Kriterium erfüllt. Wenn sie zusätzlich die Steigung korrekt einzeichnete, dann ist das Kriterium „AV-Steigung“ ebenfalls erfüllt. Ein Lösungsgraph, der hierfür einen Punkt erhält, entspricht gleichzeitig einer komplett korrekten Lösung.

Wenn eine Versuchsperson für den leichten Flussgrößen-Verlauf einen Lösungsgraphen zeichnete, der abnimmt (egal wie groß dieser Zeitraum ist), so erhielt sie 0 Punkte. Wenn eine Versuchsperson einen Lösungsgraphen zeichnete, der über den gesamten Zeitraum hinweg zunimmt, so erhielt sie einen Punkt bei AV-Tendenz. Wenn sie aber eine Kurve und keine Gerade zeichnete, so erhielt sie null Punkte für AV-Steigung. Zeichnete sie eine zunehmende gerade Linie, so erhielt sie auch den Punkt bei AV-Steigung. Gemessen wurde also die Lösungsgüte über zwei Kriterien. Diese können für jede Aufgabe aufsummiert werden, wodurch sich 0-2 Punkte für jede Aufgabe ergeben. Das wiederum wird zur „Gesamtsumme“ zusammengefasst, die 0-6 Punkte annimmt. Diese Gesamtsumme wird für die grundlegenden Auswertungen genutzt, vereinzelt finden Auswertungen getrennt für die verschiedenen Flussgrößen-Verläufe statt. Das wird für alle folgenden Studien derart gehandhabt.

3.2.4.1 Fehleranalyse

Wiederum werden die Fehler analysiert. Tritt die Korrelations-Heuristik auf, und wenn ja, welche der Flussgrößen zogen die Versuchspersonen heran? Gibt es Besonderheiten, wann sie besonders häufig oder selten auftritt? Zeigen sich weitere systematische Fehler?

3.2.4.2 Ermittelter Nettofluss und bedingte Wahrscheinlichkeiten

In der Bedingung NF-ermitteln zeichneten die Versuchspersonen den Nettofluss ein. Dieser wurde auf Korrektheit überprüft. Daraufhin ist es möglich, bedingte Wahrscheinlichkeiten zu berechnen: Die Wahrscheinlichkeit, dass Versuchspersonen den korrekten Lösungsgraph zeichnen, wenn sie vorher den Nettofluss korrekt ermittelt hatten und für den Fall, dass sie ihn falsch ermittelt hatten.

3.2.5 Weitere Variablen

Als Kontrollvariablen wurden die Variablen Geschlecht und Studienfach erhoben. Weiterhin wurde das Alter erhoben und ob die Person schon einmal an einer derartigen Studie teilgenommen hat. Zusätzlich wurden die folgenden Variablen erhoben: Frustra-

tion, Anspannung, Motivation und Verständlichkeit der Graphiken. Diese Variablen wurden für explorative Zwecke erhoben. Sie wurden ausgewertet, sind jedoch für die vorliegende Arbeit nicht dokumentiert, da sich kein relevanter Erkenntnisgewinn daraus ergab. Entsprechendes gilt für die folgenden Studien. Weiterhin wurde die wahrgenommene Aufgabenschwierigkeit erfragt, um zu überprüfen, ob sie mit der Aufgabenschwierigkeit, ermittelt über die Lösungsgüte, übereinstimmt. Außerdem sollten die Versuchspersonen die offene Frage nach ihrem Vorgehen beantworten.

3.2.6 Ablauf

Bis zu sechs Personen konnten gleichzeitig teilnehmen. Sie erhielten Informationen zum Experiment, zunächst zur Prozedur, zum gegebenen BF-Szenario und den Hinweis, sich an den/die VersuchsleiterIn zu wenden, wenn sich ein Problem oder eine Frage ergab. Für alle Versuchspersonen war das BF-Szenario ein Wasserbecken mit Zu- und Abfluss von Wasser. Die Versuchspersonen wurden zufällig einer Bedingung zugeordnet. Sie bearbeiteten die drei Aufgaben innerhalb einer Bedingung, wobei drei unterschiedliche Reihenfolgen der Aufgaben vorlagen. Das wurde über ein lateinisches Quadrat umgesetzt. Die Zuordnung zu ei-

ner Bedingung mit bestimmter Aufgabenreihenfolge ergab sich über Blockrandomisierung.

Jede Versuchsperson löste drei Aufgaben, für die sie jeweils 8 Minuten Zeit hatte. Wenn sie nach 8 Minuten nicht fertig war, unterbrach sie der/die VersuchsleiterIn und gab ihr das nächste Arbeitsblatt, unabhängig davon, ob die Aufgabe gelöst war oder nicht. Dieses Zeitlimit war aufgrund organisatorischer Begebenheiten eingebaut: Die Gesamtdauer von 30 Minuten konnte nicht überschritten werden. In allen Bedingungen erhielt die Versuchsperson für jede Aufgabe ein Aufgabenblatt mit schematischer Abbildung des Szenarios, mit einem Flussdiagramm und einem Antwortdiagramm. In den Bedingungen mit Nettofluss war auf dem Aufgabenblatt ein zusätzliches Diagramm für den Nettofluss abgebildet und im Text die Größe Nettofluss erklärt. Es wurde betont, dass es sich hierbei um eine Hilfestellung handelt. Inklusive abschließendem Fragebogen betrug die Gesamtdauer höchstens 30 Minuten.

3.2.7 Vorgehen bei der Analyse

Das Vorgehen bei der Analyse entsprach dem aus Studie 1 dieser Arbeit: Zum einen wurden Prozentzahlen für komplett korrekte Lösungen ermittelt. Da 21 Personen an jeder Bedingung teilgenommen und jeweils 3 Aufgaben bearbeitet hatten, ergab

sich pro Bedingung 63 als maximale Anzahl an Aufgaben (also 100%). Daran wurden die Häufigkeiten aller komplett korrekten Lösungsgraphen relativiert. Da die Anzahl an Versuchspersonen in jeder Bedingung gleich groß war, sind die absoluten Zahlen direkt miteinander vergleichbar. Zudem berichte ich teilweise, wie viele Personen keinen einzigen Punkt in einer der Aufgaben erhielten. Weiterhin berechnete ich den mittleren prozentualen Anteil. Hierbei handelt es sich um den Mittelwert der BF-Leistung, der in eine Prozentangabe umgerechnet ist, wobei 6 Punkte, als maximal erreichbarer Wert, 100% darstellen. Zudem liegen Abbildungen vor, und eine Regressionsanalyse wird berichtet.

3.3 Ergebnisse

Zunächst erfolgt ein erster Überblick über die Daten. Daraufhin folgen die Auswertungen für die verschiedenen Bedingungen. Anschließend beschreibe ich die Fehleranalyse. Es folgen die Analysen der zusätzlichen Variablen wie Geschlecht, Studienfach und einzelne weitere Analysen.

Die BF-Leistung wurde über die Gesamtsumme ermittelt, welche von 0 bis 6 Punkte variiert. Hierfür ergab sich ein Mittelwert von 2.68 (SD = 2.33). Das entspricht einer mittleren prozentualen Leistung von 45%. 18 Personen (21%) lösten alle drei Aufgaben komplett korrekt und 26 Personen (31%) hatten 0

Punkte für alle drei Aufgaben. Das heißt, sie konnten nicht einmal für eine Aufgabe Zu- oder Abnahme des Bestandes korrekt einordnen. Es wurde geprüft, ob das Zeitlimit zu Verzerrung der Ergebnisse hätte führen können. Abgebrochen wurde nur bei zwei Personen, bei ihnen zeigte sich kein Hinweis für daraus resultierende Probleme.

3.3.1 Überblick und Regressionsanalyse

Der Anteil derjenigen, die alle Aufgaben komplett korrekt lösten, unterschied sich stark nach den verschiedenen Bedingungen. Er lag in der Baseline bei 48%, bei KS-geteilt bei 24%, in der Bedingung NF-gegeben bei 10% und in der Bedingung NF-ermitteln bei 5%. Es folgt ein Überblick über alle vier Bedingungen für die abhängigen Variablen AV-Tendenz und AV-Steigung, jeweils zusammengefasst für alle drei Aufgaben pro Versuchsperson (Abbildung 3.7).

Nach den Medianen zu urteilen, schnitt die Baseline-Bedingung bei AV-Tendenz am besten ab, gefolgt von der Bedingung KS-geteilt, die beiden Nettofluss-Bedingungen lagen gleich auf. Für AV-Steigung haben alle Bedingungen den gleichen Median. Im weiteren Verlauf werden nur die Ergebnisse für die Summe aus diesen beiden abhängigen Variablen gezeigt. Auch wenn AV-Steigung und AV-Tendenz einzeln betrachtet wurden, ergaben

keine zusätzlichen Erkenntnisse zu den Ergebnissen aus Analysen mit der Gesamtsumme.

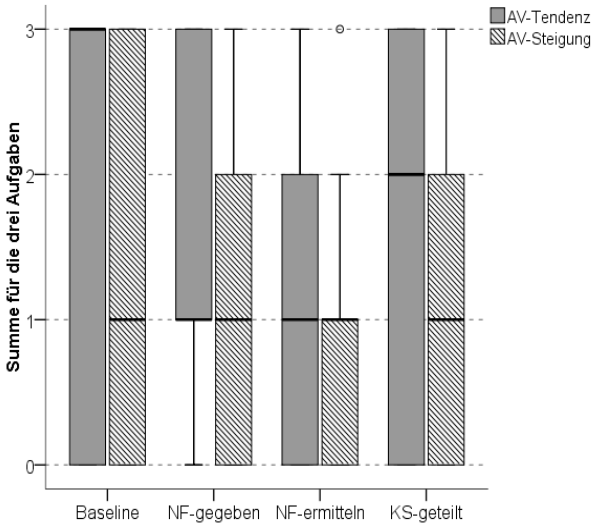


Abbildung 3.7 Boxplots für die AV-Tendenz und AV-Steigung in den vier Bedingungen.

Obwohl die Modifikationen die Leistung verschlechterten, statt sie zu verbessern, erfolgte eine Regressionsanalyse zur Veranschaulichung von Details. In die Analyse gingen die vier

Bedingungen Baseline, KS-geteilt, NF-gegeben und NF-ermitteln ein, wobei sie über drei Dummy-Variablen kodiert waren. 0 stand in allen Fällen für die Baseline, 1 für die jeweilige modifizierte Bedingung. Weiterhin enthielt das vollständige Modell die Variablen Geschlecht, Studienfach und vorherige Teilnahme. Es ergaben sich die Werte in Tabelle 3.1. Von den Bedingungen war keine hilfreich, wobei die Bedingung NF-ermitteln besonders negativ zu wirken schien.

Tabelle 3.1

Regressionsanalyse mit den drei Prädiktoren für die Bedingungen und mit den Kontrollvariablen. Berichtet werden neben β-, *t*- und *p*-Werten die Partialkorrelationen *pr*. Die abhängige Variable ist die Gesamtsumme als Indikator für BF-Leistung.

Prädiktor	β	*t*	*p*	*pr*
KS-geteilt	-.12	-.901	.19	-.10
NF-gegeben	-.13	-1.014	.16	-.12
NF-ermitteln	-.28	-2.217	.02	-.25
Geschlecht	.07	.617	.30	.07
Studienfach	.35	3.393	<.00	.36
Teilnahme	-.01	-.092	.43	-.01

Anmerkung: $R^2 = .17$, $F(6,77) = 2.705$, $p = .02$. Die Variablen sind dummy-kodiert: geteiltes Koordinatensystem (KS-geteilt: 0 nein, 1 ja); Nettofluss-gegeben (NF-gegeben: 0 nein, 1 ja); Nettofluss-ermitteln (NF-ermittlen: 0 nein, 1 ja) Geschlecht (0 weiblich, 1 männlich), Studienfach (0 sonstige, 1 Psychologie) und vorherige Teilnahme (Teilnahme: 0 nein, 1 ja).

Es folgen detailliertere Auswertungen für die einzelnen Bedingungen. Dabei werden einzelne Werte aus der Regression berichtet.

3.3.2 Bedingung „geteiltes Koordinatensystem“

Zunächst folgen die Ergebnisse zu *Hypothese 2.1*, nach der das Trennen von Zu- und Abfluss zu besserer BF-Leistung führen sollte. 56% war der mittlere prozentuale Anteil in der Leistung Baseline-Bedingung, während er für KS-geteilt bei 47% lag. In der Bedingung KS-geteilt waren 26 Aufgaben (41%) komplett korrekt gelöst. In der Baseline-Bedingung waren es 34 Aufgaben (54%).

Abbildung 3.8 zeigt die Boxplots für die Baseline-Bedingung und für die Bedingung KS-geteilt mit der Variable Gesamtsumme auf der Y-Achse. Wie in Abschnitt 3.2.7 beschrieben, handelt es sich dabei um die Summe für alle drei Aufgaben und die zwei abhängigen Variablen AV-Tendenz und AV-Steigung. Diese Summe kann somit Werte von 0 bis 6 Punkte annehmen. In der Baseline-Bedingung lag bessere Leistung vor. Das stimmt mit den Ergebnissen aus der Regressionsanalyse überein, die für das vollständige Modell mit allen Bedingungen und den drei weiteren Prädiktoren Werte von $\beta = -.12$, $pr = -.10$ mit $p = .19$ ergab.

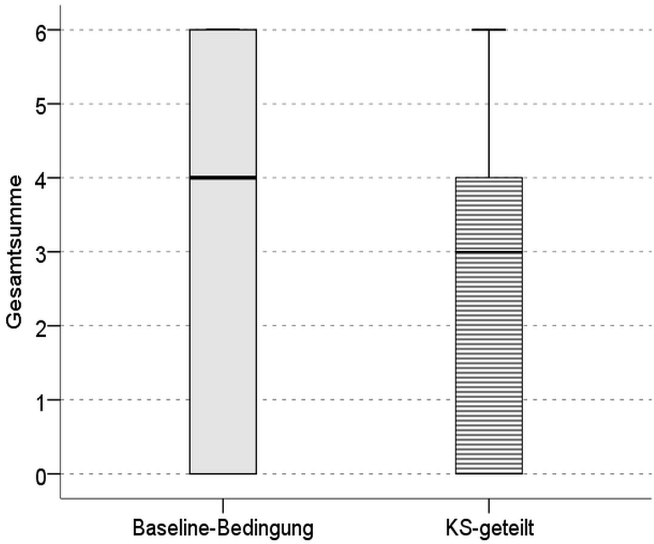


Abbildung 3.8 Boxplots der Gesamtsumme für die Bedingung geteiltes Koordinatensystem im Vergleich mit der Baseline-Bedingung.

Missinterpretation bei der neuen Darstellung

Möglicherweise ergaben sich Missverständnisse für die neue Darstellung mit dem geteilten Koordinatensystem. Beispielswei-

se könnten Versuchspersonen das Diagramm falsch verstanden haben und zu einer falsche Lösung kommen, die jedoch entsprechend ihrer Fehlinterpretation die korrekte Lösung war.

Eine mögliche Fehlinterpretation ergab sich aus dem Abfluss im negativen Quadranten. Wenn eine Versuchsperson annahm, dass der Abfluss ein negativer Wert ist, und dieser sich im negativen Bereich befindet, so könnte sie daraus folgern: Ein negativer Abfluss ist ein positiver Wert (entsprechend der Aufgabe Werte zu subtrahieren, und wenn der Subtrahend negativ ist, wird sein Betrag im Minuend addiert). Weiterhin ist es möglich, dass die Versuchsperson die Y-Achse für den Abfluss umgedreht interpretiert. In dem Fall ginge sie davon aus, dass auch für den Abfluss der 0-Punkt unten ist und nach oben hin die Werte zunehmen. Das würde dazu führen, dass die Versuchsperson einen eigentlich kleinen Wert als groß wahrnimmt und umgekehrt.

Wenn nun eine Versuchsperson die Werte falsch interpretiert, könnte es sein, dass sie mit den falschen Werten zu einer in sich stimmigen Lösung kommt. Also beispielsweise nimmt sie an, der „negative Abfluss“ ist ein Wert, dessen Betrag sie zum Zufluss addiert. Nun könnte ihre Lösung stimmen, wenn sie diesen Lösungsweg korrekt für alle Aufgaben ausführt. In dem Fall kommt sie aufgrund eines Interpretationsfehlers zu einer falschen Lösung und nicht aufgrund fehlenden BF-Verständnisses. Wenn

das der Fall ist, so hätte dies Auswirkungen auf mögliche weitere Studien (in denen derartigen Interpretationsfehlern abgeholfen wird, um zu testen, ob sich dann die Lösungsgüte verbessert). Die falschen Lösungen wurden also dahingehend untersucht, ob sie korrekt sind für den Fall, dass eine der Missinterpretationen vorlag. Hierfür wurden alle fehlerhaften Lösungen der 21 Versuchspersonen aus der Bedingung KS-geteilt analysiert.

Insgesamt existierten 6 Lösungen, die aussahen, als handele es sich um eine korrekte Lösung in Folge einer der beiden Missinterpretationen wie oben beschrieben. Jede dieser Versuchspersonen zeigte dieses Muster jedoch nur in einer Aufgabe, die anderen Aufgaben waren entweder korrekt (nach ursprünglicher Bewertung) oder falsch gelöst, jedoch mit anderem Fehlermuster. Die plausibelste Folgerung aus diesem Ergebnismuster ist, dass es keine systematische Missinterpretation gab, die zu einem korrekten Lösungsgraphen führte. Somit wurden keine weiteren Analysen diesbezüglich durchgeführt.

Insgesamt war es nicht hilfreich, Zu- und Abfluss im Flussdiagramm zu trennen mit dem Abfluss im negativen Bereich. *Hypothese 2.1* wird durch die vorliegenden Daten nicht unterstützt.

3.3.3 Die Bedingungen „NF-gegeben“ und „NF-ermitteln“

Es wurde überprüft, inwiefern das Vorliegen des Nettoflusses hilfreich ist (*Hypothese 2.2* und *Hypothese 2.3*). Der mittlere prozentuale Anteil der Kriterien liegt in der Baseline-Bedingung bei 56%, in der Bedingung NF-gegeben bei 43% und in der Bedingung NF-ermitteln bei 33%. Entsprechend fanden sich mit insgesamt 34 korrekt Aufgaben (54%) die meisten komplett korrekten Lösungen in der Baseline-Bedingung. In der Bedingung NF-gegeben waren es 23 Aufgaben (37%), und in der Bedingung NF-ermitteln waren es 16 Aufgaben (25%), die von allen Versuchspersonen komplett korrekt gelöst wurden. In Abbildung 3.9 zeigen Boxplots die Unterschiede für die Bedingungen in der Gesamtsumme.

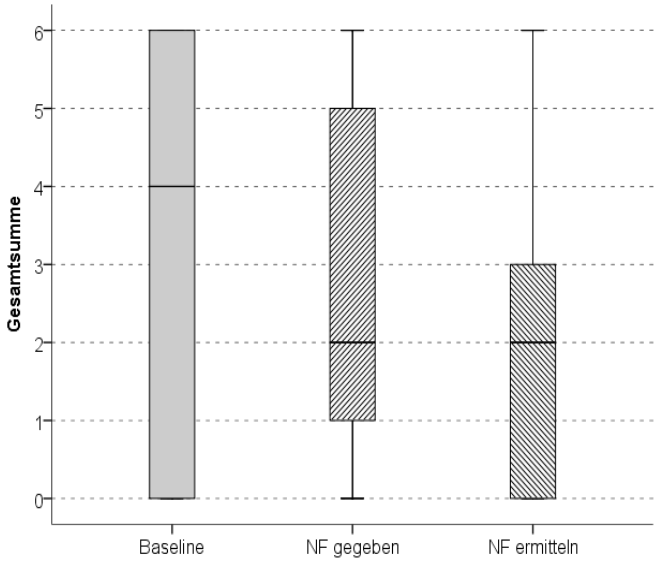


Abbildung 3.9 Boxplots der Gesamtsumme für die Bedingungen Baseline, NF-gegeben und NF-ermitteln.

Die beste Leistung lag in der Baseline-Bedingung vor. In den beiden Nettofluss Bedingungen ergab sich derselbe Median, wobei die Bedingung NF-gegeben in ihrer Gesamtheit leicht besser abschnitt. Es ist weder hilfreich den Nettofluss vorzugeben, noch ihn berechnen zu lassen. Tatsächlich verschlechterte sich

die Lösungsgüte in diesen beiden Bedingungen. In der Bedingung NF-gegeben zeigten sich folgende Werte der Regression: $\beta = -.13$. $pr = -.12$ mit $p = .16$; bei NF-ermitteln verschlechtert sich die Leistung sogar sehr stark, wie die Regressionsanalyse zeigte: $\beta = -.28$; $pr = -.25$, $p = .02$.

Hypothese 2.2 und *Hypothese 2.3* wurden verworfen Das zusätzliche Geben des Nettoflusses führte nicht zu besserer Leistung. Die Leistung verschlechterte sich tendenziell stärker, wenn der Nettofluss selbst ermittelt wurde. Jedoch erfolgten keine detaillierteren Auswertungen zur *Fragestellung 2.1* nach einem Unterschied in der BF-Leistung, je nachdem ob der Nettofluss gegeben oder selber ermittelt wurde. Aufgrund der Verschlechterung in der BF-Leistung, schien es nicht weiter relevant, ob diesbezüglich ein signifikanter Unterschied vorlag.

3.3.3.1 Korrektheit des ermittelten Nettoflusses

Wie oft haben Personen in der Bedingung NF-ermitteln den korrekten Nettofluss gefunden? Die 21 Personen dieser Bedingung mussten drei Aufgaben bearbeiten. Bei der leichten Aufgabe ermittelten 12 Versuchspersonen den Nettofluss korrekt, bei der mittelschweren Aufgabe waren es 10 Versuchspersonen und bei der schwierigen Aufgaben 6 Versuchspersonen. Insgesamt waren also 44% der Nettofluss-Lösungen korrekt angegeben.

3.3.3.2 Bedingte Wahrscheinlichkeit bei NF-ermitteln

Wie groß ist die Wahrscheinlichkeit, die korrekte Lösung zu finden, wenn eine Versuchsperson den Nettofluss korrekt ermittelt hat? Und wie groß ist die Wahrscheinlichkeit dafür, wenn sie den Nettofluss *nicht* korrekt ermittelt hat? Diese bedingten Wahrscheinlichkeiten wurden mit der Wahrscheinlichkeit verglichen, mit der die Versuchsperson in der Baseline-Bedingung die korrekte Lösung fanden, welche bei 54% lag.

Die bedingten Wahrscheinlichkeiten für AV-Steigung sind in Abbildung 3.10 gezeigt. Die Variable AV-Steigung wurde genutzt, da sie der komplett korrekten Lösung für die Aufgabe entspricht. Das heißt, wenn eine Lösung für den Bestand dieses Kriterium erfüllt, so ist der Lösungsgraph komplett korrekt gezeichnet. Die Basisrate für die bedingte Wahrscheinlichkeit ist gestrichelt in der kräftigen Säulenumrahmung abgebildet und setzt sich aus dem weißen und grauen Anteil zusammen. Die Basisrate umfasst in einem Fall alle einzelnen, korrekt gezeichneten Lösungen für den Nettofluss (Balken „Nettofluss korrekt“): 28 Mal wurde der Nettofluss korrekt gezeichnet, wiederum bezogen auf die 21 Personen, die jeweils 3 Aufgaben bearbeiteten. Der andere Balken beinhaltet alle falsch gezeichneten Lösungen für den Nettofluss: 35 Mal wurde der Nettofluss falsch ermittelt. Inner-

halb der Balken befindet sich eine graue Füllung Das entspricht dem Anteil der korrekt eingezeichneten Lösungen für den Bestand. Das heißt also die bedingte Wahrscheinlichkeit, den Bestand korrekt einzuzeichnen, wenn der Nettofluss korrekt ermittelt wurde, lag bei 40% (p(Bestand korrekt|Nettofluss korrekt)). Sie war geringer als die Baseline-Wahrscheinlichkeit, eine Aufgabe korrekt zu lösen, die 54% betrug. Die bedingte Wahrscheinlichkeit, mit falschem Nettofluss den Bestand korrekt zu zeichnen, war noch geringer mit etwa 15% (p(Bestand korrrekt| Nettofluss falsch)). Interessanterweise traten einzelne korrekte Lösungen für den Bestand auf, wenn der Nettofluss falsch ermittelt wurde.

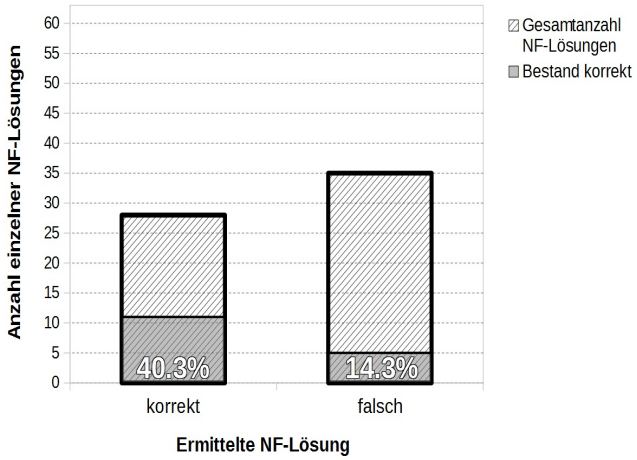


Abbildung 3.10 Bedingte Wahrscheinlichkeiten für die korrekten Lösungen (grau), wenn der Nettofluss korrekt bzw. wenn er falsch gezeichnet wurde (gestrichelt).

3.3.4 Fehleranalyse

Hypothese 2.5 erübrigt sich, da sich aufgrund der Manipulationen keine Verbesserung in der Lösungsgüte ergab und somit aufgrund der Modifikationen keinerlei Fehlermustern systematisch abgeholfen wurde. Dennoch kann gefragt werden: Wie oft und wann nutzen die Versuchspersonen die Korrelations-Heuris-

tik? Und entsprechend *Fragestellung 2.2*: Ergeben sich weitere systematische Fehlermuster?

3.3.4.1 Weiteres systematisches Fehlermuster

Tatsächlich ergab sich ein weiteres Fehlermuster, das nicht nur vereinzelt auftrat. Auch in diesem Fall zeigte der Lösungsgraph einen ähnlichen Verlauf wie eine der Flussgrößen, jedoch mit einem Unterschied zu dem klassischen Fehlermuster der Korrelations-Heuristik. Zunächst begann der Lösungsgraph an dem markierten Punkt für „Bestand zu Beginn“, dann sank der Graph in Richtung x-Achse, meist innerhalb der ersten (wenigen) Zeitpunkte. Anschließend zeichnete die Versuchsperson einen Lösungsgraphen orientiert an der Korrelations-Heuristik. Dieser Fehler zeigte sich öfter und wurde daher nicht in die Kategorie für Korrelations-Heuristik mit eingeordnet. Abbildung 3.11 zeigt ein derartiges Beispiel für die schwierige Aufgabe mit Glockenform, bei der sich ein Muster für den Bestand ergab, der aussah wie der Zu- oder Nettofluss. Dieses Fehlermuster bezeichne ich im weiteren Verlauf als Korrelations-Heuristik-Diskontinuität. Ein weiteres systematisches und häufig auftretendes Fehlermuster wurde nicht in den Daten entdeckt.

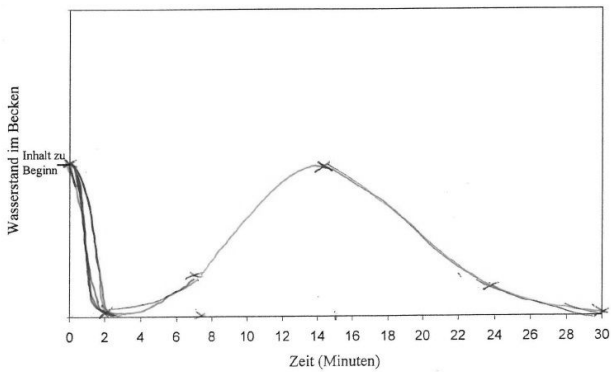


Abbildung 3.11 Lösung mit dem Fehlermuster „Korrelations-Heuristik-Diskontinuität“ für die schwierige Aufgabe.

Ein derartiges Lösungsmuster zeigte sich für die leichte Aufgabe ein Mal in der Baseline-Bedingung und ein Mal in der Bedingung NF-ermitteln. Für die mittelschwere Aufgabe zeigten sich derartige Verläufe 3 Mal in der Baseline-Bedingung, 5 Mal in NF-gegeben, 2 Mal für NF-ermitteln und kein Mal bei KS-getrennt. Für die schwierige Aufgabe zeigte es sich mit 7 Mal am häufigsten in der Bedingung NF-gegeben. In den anderen Bedingungen trat es seltener auf, nämlich ein oder 2 Mal.

3.3.4.2 Überblick über das Anwenden der Korrelations-Heuristik

Insgesamt wurden von 252 Aufgaben 153 nicht komplett korrekt gelöst. In 73 Fällen, also knapp der Hälfte der falsch gelösten Aufgaben, wurde die Korrelations-Heuristik angewandt. Abbildung 5.12 zeigt, welche Fehler in den verschiedenen Bedingungen auftraten. Die Häufigkeiten der Fehlermuster sind an den falschen Lösungen insgesamt relativiert.

Die Korrelations-Heuristik zeigte sich am häufigsten in der Baseline Bedingung und in der Bedingung NF-ermitteln, gefolgt von der Bedingung KS-geteilt. Am wenigsten wurde sie in der Bedingung NF-gegeben genutzt. *Hypothese 2.4* bestätigte sich: In der Baseline-Bedingung traten fehlerhafte Lösungen orientiert an der Korrelations-Heuristik auf. Werden die Fehlermuster Korrelations-Heuristik und Korrelations-Heuristik-Diskontinuität zusammengefasst, so ergibt sich für die Baseline-Bedingung 55%, für KS-geteilt 50%, für NF-gegeben 49% und für NF-ermitteln 38%. Es zeigten sich also keine nennenswerten Unterschiede. Tendenziell am häufigsten trat die Korrelations-Heuristik in der Baseline auf und am seltensten in der Bedingung NF-ermitteln.

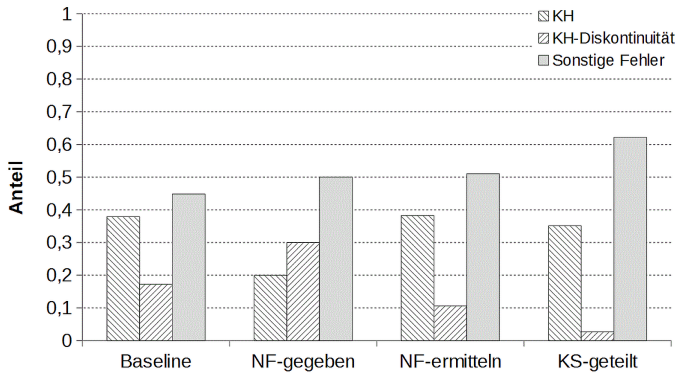


Abbildung 3.12: Anteile der Fehlermuster an den falschen Lösungen für die vier Bedingungen
Anmerkung: KH bedeutet Korrelations-Heuristik und KH-Diskontinuität bedeutet Korrelations-Heuristik-Diskontinuität.

3.3.4.3 Für Korrelations-Heuristik genutzte Flussgrößen

Weiterhin wurde analysiert, mit welcher Flussgröße der Lösungsgraph am häufigsten korrelierte, wenn die Korrelations-Heuristik angewandt wurde. Bei den Lösungen orientiert an der Korrelations-Heuristik ergab sich in den meisten Fällen eine Lösung, die wie Zufluss oder Nettofluss aussah. Beide Male nah-

men die Linien denselben Verlauf an, somit ist nicht erkennbar, welche der Flussgrößen die Versuchspersonen als Orientierung heranzogen. Außerdem nutzten einzelne Versuchspersonen den Abfluss als Orientierung für die Korrelations-Heuristik. Sechs Mal von allen 73 Fällen handelte es sich um ein Muster, das Orientierung am Abfluss nahelegte. Für die einfache Aufgabe war nicht zu unterscheiden, ob Orientierung am Ab- oder Zufluss vorlag.

In einer Aufgabe ist es möglich zu unterscheiden, ob Zufluss oder Nettofluss zur Orientierung dient. Es handelt sich um die einfache Aufgabe in der Bedingung KS-geteilt. Hier sieht das Korrelations-Heuristik-Ergebnis für alle drei Flussgrößen unterschiedlich aus. Das Flussdiagramm der einfachen Aufgabe enthielt zwei parallele, in ihrer Steigung fallende Linien. Der Nettofluss ist somit über den gesamten Verlauf hinweg gleich groß. Das ergibt einen Nettofluss mit einer waagrechten Linie. In der Bedingung KS-geteilt ergibt sich für den Abfluss im negativen Quadranten eine an der X-Achse gespiegelte Linie. Wenn eine Versuchsperson die Korrelations-Heuristik anwendet, ergibt sich für den Bestand entweder eine steigende Linie für die Korrelations-Heuristik orientiert am Zufluss, eine Linie mit fallender Steigung für die Korrelations-Heuristik orientiert am Abfluss oder eine waagrechte Linie für die Korrelations-Heuristik orientiert

am Nettofluss. Von 21 Personen in der Bedingung KS-geteilt wandten 7 die Korrelations-Heuristik an. Sie alle orientierten sich am Nettofluss.

3.3.4.4 Fazit

Zusammenfassend lässt sich sagen, es zeigten sich bei den falschen Lösungen das klassische Lösungsmuster „Korrelations-Heuristik“ und eine abgewandelte Form: Korrelations-Heuristik-Diskontinuität. Es ergaben sich keine weiteren systematischen Fehler. Die Anwendung der Korrelations-Heuristik trat über die Bedingungen hinweg ungefähr gleich häufig auf, wenn sie an gesamter Fehlermenge relativiert wurde. Es folgen Auswertungen zu den Kontrollvariablen Geschlecht und Studienfach.

3.3.5 Kontrollvariablen

3.3.5.1 Geschlecht

In der vorliegenden Studie waren 12 von den 84 Versuchspersonen männlich. Bis auf eine leichte Tendenz zum Vorteil der Männer (Median von 3 statt von 2), zeigte sich kein Einfluss von Geschlecht. Die Regressionsanalyse ergab diesbezüglich: $\beta = .07$, und $pr = .07$ mit $p = .30$.

3.3.5.2 Studienfach

Für diese Variable wurden Psychologie-Studierende den nicht Psychologie Studierenden („Sonstige“) gegenüber gestellt. Die zweite Gruppe bestand aus 13 Personen, die Erziehungswissenschaften und 3 Personen, die Germanistik als Hauptfach studierten, während ein Anteil von 81% Psychologie studierte. Abbildung 3.12 zeigt Boxplots für die BF-Leistung je nach Studienfach. Nach den Boxplots schnitten Psychologie-Studierende besser ab.

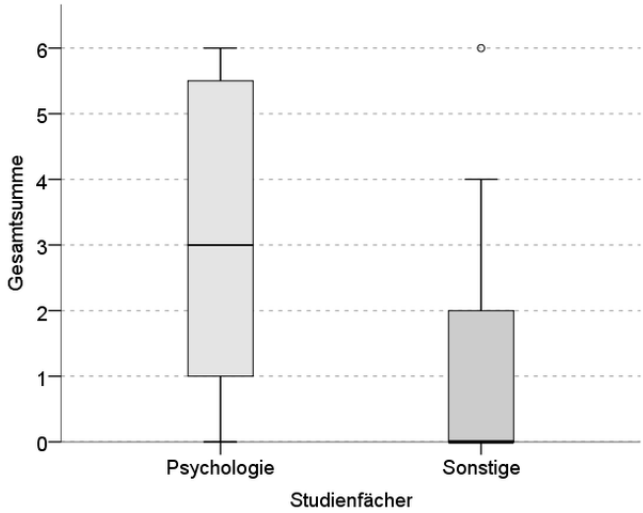


Abbildung 3.12 Boxplots der Gesamtsumme für Psychologie-Studierende und Nicht-Psychologie-Studierende (Sonstige).

Auch aus der Regressionsanalyse ergab sich $\beta = .35$, $pr = .36$, $p < .00$ und somit eine bessere BF-Leistung seitens der Psychologie-Studierenden.

3.3.5.3 Teilnahme an vorhergehender Studie

11 Personen hatten an der ersten Studie teilgenommen. Ein Vergleich dieser Gruppe mit den Versuchspersonen, die das erste

Mal an einer BF-Studie teilnahmen, zeigte keinen nennenswerten Unterschied hinsichtlich der Lösungsgüte. Die Mediane lagen für Personen, die schon teilgenommen hatten bei 2 statt bei 3 in der anderen Gruppe. Jedoch gab es in dieser Gruppe zwei Ausreißer mit voller Punktzahl. Wenn betrachtet wird, wie viele Versuchspersonen anteilsmäßig die Aufgaben komplett korrekt lösten, so zeigt sich in etwa der gleiche Anteil: Diejenigen, die schon teilgenommen haben, hatten hier einen Anteil von etwa 18%, bei den Erstteilnehmenden lag er etwa bei 22%. Berichtet sind die Werte in der Regressionsanalyse (Tabelle 3.1). Diese sprachen ebenfalls gegen einen nennenswerter Unterschied in der Leistung.

Insgesamt zeigten die Ergebnisse zu den Kontrollvariablen: Das Studienfach beeinflusste die Lösungsgüte, während Geschlecht und vorherige Teilnahme keinen Einfluss zeigten.

3.3.6 Zusätzliche Analysen

Weitere Analysen betrafen in dieser Studie die Aufgabenschwierigkeit und Positionseffekte.

3.3.6.1 Aufgabenschwierigkeit

Die Aufgabenschwierigkeit wurde über die Lösungsgüte überprüft und entsprach tendenziell der postulierten Aufgabenschwierigkeit (vgl. Cronin et al., 2009). Entsprechendes galt für

die wahrgenommene Schwierigkeit. Da sich in den folgenden Studien ein entsprechendes Bild ergab, sind diese Variablen im weiteren Verlauf nicht mehr berichtet.

3.3.6.2 Positionseffekte

Die Lösungsgüte der drei Aufgaben wurde in Abhängigkeit der Position ermitteltet, um eventuell existierende Positionseffekte aufzudecken. Diese Analyse war möglich, da unabhängig von der Bedingung jede Aufgabe an jeder Stelle gleich häufig vorkam. Der Mittelwert sollte also für jede Position, unabhängig von der Bedingung, gleich groß sein. Tatsächlich ergab sich kein Positionseffekt in den Varianzanalysen für die drei verschiedenen Aufgaben (für leicht: $F(2,81) = .974$ mit $p = .38$; für mittelschwer: $F(2,81) = 1.138$ mit $p = .33$; für schwierig $F(2,81) = .055$ mit $p = .95$).

3.4 Diskussion

Insgesamt half keine der hier untersuchten Änderungen der Aufgabendarbietung, das BF-Verständnis zu erhöhen: Weder ein geteiltes Flussdiagramm noch der zusätzlich gegebene oder zu berechnende Nettofluss verbesserte die Lösungsgüte. Tatsächlich verschlechterte sich die BF-Leistung eher in den modifizierten Bedingungen im Gegensatz zur Baseline-Bedingung. In allen

Fällen machte es keinen Unterschied, ob das Kriterium AV-Tendenz, AV-Steigung oder die Gesamtsumme betrachtet wurde.

3.4.1 Bedingung geteiltes Koordinatensystem

Die vorliegenden Daten widersprachen der Annahme, dass eine Teilung des Koordinatensystems die BF-Leistung verbesserte (Hypothese 2.1). In der Bedingung KS-geteilt schnitten die Versuchspersonen schlechter ab. Woran lag das? Mindestens eine der folgenden Erklärungen ist denkbar: Es könnte sich um eine Missinterpretation handeln oder um Probleme, die Werte zu verarbeiten. Weiterhin ist möglich, dass die Manipulation keine ist, die das korrekte BF-Verständnis anspricht. Darauf gehe ich im Folgenden ausführlich ein.

3.4.1.1 Missinterpretationen bei der neuen Darstellung?

Es besteht die Möglichkeit, dass die Versuchspersonen das Diagramm falsch interpretierten. Möglicherweise sahen sie die Werte so an, dass sie in beiden Fällen von unten nach oben hin zunahmen. Oder sie interpretierten den „negativen Abfluss“ als doppelt negativ und somit als positiven Wert. Erste Hinweise existieren jedoch, dass das keine systematische Erklärung sein kann. Wenn eine Versuchsperson eine Lösung angab, die auf ei-

ner dieser Missinterpretationen beruhte, war dies nur für eine Aufgabe der Fall. Bei anderen von ihr gelösten Aufgaben lagen entweder ein anderes Fehlermuster oder eine korrekte Antwort vor (Abschnitt 3.3.2).

Außerdem liegt das klassische Achsensystem vor, mit der X-Achse als Trennlinie zwischen positivem und negativem Bereich. Dabei handelt es sich um eine relativ bekannte Form der Diagrammdarstellung. Diese Form der Darstellung sollte den meisten Personen in der Stichprobe bekannt sein, da sie aus Studierenden bestand, die im Rahmen des Mathematikunterrichtes derartige Diagramme kennengelernt und damit gearbeitet hatten. Dieser Punkt scheint nicht die Erklärung für eine Verschlechterung in der Lösungsgüte zu sein. Die beschriebenen Missinterpretationen verursachten wohl nicht das schlechtere Abschneiden der Versuchspersonen in ihrer BF-Leistung. Also wurden weitere mögliche Ursachen in Betracht gezogen.

3.4.1.2 Probleme im Abgleich der Werte?

Auch wenn die Versuchsperson das Diagramm verstanden hatte, könnte sie Probleme haben, den Zu- und Abfluss direkt zu vergleichen. Die Zu- und Abflüsse sind im positiven und negativen Bereich abgetragen, die Werte liegen weit auseinander. Bei ähnlich großen Werten ist es auf den ersten Blick schwierig zu er-

kennen, ob der Zufluss oder der Abfluss größer ist, oder ob beide Werte gleich groß sind. Es stellt sich also die Frage: Hat die Aufteilung zwar zu einem verbesserten BF-Verständnis geführt, aber gleichzeitig vermehrte Rechenfehler verschlechterten die BF-Leistung?

Zunächst wurden die Aussagen in den Fragen nach dem Vorgehen und in den Anmerkungen der Versuchsteilnehmer durchsucht. Gibt es Hinweise, dass das Ablesen problematisch war? In drei Fällen gab es sie. Eine Person schrieb: „Das Zählen der kleinen Striche war mühsam“. Sie hat somit über das Zählen der feinen Striche genaue Werte erhalten. Ein weiterer Kommentar war: „(…) sodass es schwierig war auf einen Blick zu erkennen, ob gleich/mehr/weniger Wasser abfließt als hinzufließt.“. Eine andere Versuchsperson schrieb, dass eine Aufgabe „schlecht zu lesen“ war. Weiterhin berichtete jemand, er habe mit dem Lineal gearbeitet. Demnach waren kleine Unterschiede zwischen den Werten von Zu- und Abfluss in der Bedingung KS-geteilt schwierig zu erkennen. Einzelne Versuchspersonen ermittelten anhand der Striche die Abstände oder nutzten ein Lineal. In diesen Fällen sind die exakten Werte nicht das Problem für die falschen Lösungen. Ansonsten beschrieben mehrere Versuchspersonen, dass sie Werte oder Relationen verglichen, ohne dass sie Probleme dabei erwähnten. Da diese Aussagen die Frage nach dem Problem in

der Bedingung jedoch nicht komplett beantworten, wird an dieser Stelle zusätzlich die Aufgabenschwierigkeit und die Lösungen der Versuchspersonen betrachtet.

Die Aufgabenschwierigkeit kann folgendermaßen helfen, diese Frage zu beantworten: Zu- und Abfluss im genauen Wert zu unterscheiden, sollte gut funktionieren, wenn sich die Werte stark unterscheiden und nur bei kleinen Unterschieden schwierig sein. Somit sollten die meisten Probleme bei der mittelschweren und bei der schwierigen Aufgabe in dem Bereich auftreten, in dem ähnliche Werte von Zu- und Abfluss vorliegen. Weiterhin könnte die einfache Aufgabe insgesamt schwieriger sein, da sich Zu- und Abfluss über den gesamten Zeitraum hinweg nicht stark unterscheiden. Die Versuchspersonen nahmen die leichte Aufgabe in der Bedingung KS-geteilt als sehr leicht wahr. Im Vergleich über alle Aufgaben aus allen Bedingungen ergibt sich für sie sogar der kleinste Wert für Aufgabenschwierigkeit. Natürlich kann es auch sein, dass die Versuchspersonen den Wertunterschied für Zu- und Abfluss ermittelten und schnell erkannten, dass die Steigung für beide Größen gleichwertig ist. Das könnte ebenfalls dazu geführt haben, dass die Versuchspersonen die Aufgabe nicht als schwieriger empfanden. Also wird die Lösungsgüte ebenfalls betrachtet. Tatsächlich lösten die Versuchspersonen in der Bedingung KS-geteilt die mittelschwere von den drei Aufgaben am besten, wäh-

rend in den anderen Bedingungen die Lösungsgüte jeweils dem postulierten Schwierigkeitsgrad entsprach. Jedoch war der Unterschied zwischen der leichten und der mittelschweren Aufgabe minimal. Die leichte Aufgabe wurde also nicht zur schwierigsten Aufgabe in der Bedingung KS-geteilt.

Abschließend suchte ich nach Hinweisen, dass das BF-Verständnis zwar gestiegen war, aber die Rechenfehler zunahmen. Das heißt, vielleicht haben mehr Personen verstanden, wie sie vorgehen, um zur korrekten Lösung zu kommen. Sie zeichneten zu den leichter differenzierbaren Zeitpunkten die Lösung korrekt und nur zu den „schwierigen Zeitpunkten“ eine falsche Lösung. Dadurch wäre die gesamte Lösung falsch, auch wenn die Versuchspersonen das Konzept verstanden hatten. Jedoch zeigte sich in den Lösungen keinerlei Hinweis dafür. Das spricht gegen die Annahme, die Schwierigkeit, die genauen Werte zu erkennen, führte zu verschlechterter BF-Leistung.

3.4.1.3 Manipulation „KS-geteilt“ verschlechterte die BF-Leistung

Verschiedene oben beschriebene Probleme oder Missinterpretationen schienen nicht zu den falschen Lösungen zu führen. Somit lag die Vermutung nahe, dass die Betonung des Abflusses auf diese Art nicht erfolgreich war. Das widerspricht Hypothese

2.1. Das Trennen von Zu- und Abfluss und Abtragen der Größen in positivem und negativem Quadranten führt also nicht zu besserem BF-Verständnis. Also reicht es nicht aus, die negativen Werte zu betonen, entsprechend Zelaznys Vorschlag für Zeitreihenvergleiche (Zelazny, 1989). Er hatte dieses Vorgehen für eine Zeitreihe empfohlen, in der eine Variable abgetragen ist. Im vorliegenden Fall handelte es sich um zwei Variablen (Zu- und Abfluss). Zusätzlich sollte die korrekte Berechnung des Bestandes getriggert werden, was mit dieser Art der Diagrammänderung nicht erfolgte. Die nächste Frage war, ob es hilfreich ist, wenn der Nettofluss zusätzlich gegeben ist.

3.4.2 Die Nettofluss-Bedingungen

Es zeigte sich, dass das Hinzufügen des Nettoflusses entgegen *Hypothese 2.2* und *Hypothese 2.3* nicht hilfreich war. Ursprünglich sollte der gegebene Nettofluss die Versuchspersonen unterstützen, nicht die Korrelations-Heuristik mit dem Nettofluss anzuwenden, sondern die korrekte Lösung zu finden. Eine Vermutung war, dass viele Versuchspersonen Zu- und Abfluss verrechnen und das als Lösungsgraph ansehen. Die Idee war nun, den Nettofluss vorzugeben, damit die Versuchspersonen erkennen, dass es sich bei dem Ergebnis von Zu- und Abfluss um ein

Zwischenergebnis handelt, nämlich den Nettofluss. Das hat nicht funktioniert.

Eventuell war es für manche Versuchspersonen verwirrend, dass das Nettofluss-Diagramm aussah wie das Lösungsdiagramm. Jedoch hatte der Unterschied zwischen Flussdiagramm und Lösungsdiagramm in der Bedingung KS-geteilt auch nicht so stark gewirkt, dass sich dort die Lösungsgüte verbessert hätte. Darum wird dieser Punkt als nicht weiter relevant angesehen.

Auch in der Literatur finden sich Hinweise, dass Anleitung bei mathematischem Problemlösen (z.B. Zwischenziele vorgeben) nicht immer hilfreich ist (Tarmizi & Sweller, 1988). Möglicherweise verstanden die Versuchspersonen in der vorliegenden Studie nicht, was der Nettofluss ist. In dem Fall kann das Hinzufügen des Nettoflusses natürlich nicht hilfreich sein.

3.4.2.1 Verstanden die Versuchspersonen das Prinzip Nettofluss?

Ein Hinweis für die Antwort liefert die Bedingung NF-ermitteln. Hier zeigte sich, dass viele Versuchspersonen den Nettofluss nicht korrekt ermittelten. Hatten die Versuchspersonen die Erklärung zum Nettofluss verstanden oder nicht? Um Hinweise hierfür zu sammeln, können verschiedene Punkte näher untersucht werden.

Zunächst berichte ich, was die Versuchspersonen zu ihrem Vorgehen schrieben. In der Bedingung NF-gegeben, gab es Aussagen, dass es „schwierig“ war, die Versuchsperson „unsicher“ war (jeweils ein Mal) und Aussagen in der Art, dass der Nettofluss verwirrend war, kamen zwei Mal vor. Bei NF-ermitteln schrieben drei Personen, sie waren unsicher, ob sie die Aufgaben korrekt verstanden hatten, beziehungsweise dass sie sie nicht verstanden hatten. Eine Versuchsperson empfand die Aufgaben als seltsam formuliert. In der Baseline-Bedingung gab es keine derartigen Aussagen in den Anmerkungen oder im Vorgehen. Hieraus ergaben sich erste Hinweise, dass einigen Versuchspersonen nicht klar war, was der Nettofluss ist. Möglicherweise gab es weitere Versuchspersonen, denen es so erging, die das aber nicht aufschrieben.

Zur weiteren Analyse ist es möglich, die Fehlermuster genauer zu betrachten. Wurde das Prinzip Nettofluss nicht verstanden oder handelt es sich oft um Flüchtigkeitsfehler? Das heißt, vielleicht hatten mehr Personen verstanden, wie sie zum korrekten Nettofluss kommen, aber sie zeichneten beispielsweise zu schnell und verzeichnen sich nur minimal (z.B. das Maximum verrutscht um ein paar Zeitpunkte). Dadurch wäre die Lösung falsch, obwohl die Versuchspersonen das Konzept verstanden hatten. Jedoch sehen die Fehlermuster eher aus, als handele es

sich um Fehler aufgrund fehlenden Verständnisses für das Prinzip Nettofluss. Somit schien zumindest einem Anteil der Versuchspersonen das Prinzip Nettofluss nicht klar zu sein, was zu den fehlerhaften Nettoflussgraphen führte.

3.4.2.2 Manipulationen mit Nettofluss verschlechterten die BF-Leistung

Wie oben beschrieben schien ein Teil der Versuchspersonen das Konzept Nettofluss nicht zu verstehen. Es ist nicht überraschend, dass diese Personen in den seltensten Fällen die korrekte Lösung für den Bestand finden. Die Versuchspersonen, die den Nettofluss korrekt zeichneten, kamen nicht vermehrt zur korrekten Lösung. Tatsächlich lag bei ihnen der Anteil immer noch unter dem Anteil der korrekten Lösungen in der Baseline-Bedingung. Somit scheint die Aufgabenstellung, den Nettofluss selber zu ermitteln, die Versuchspersonen eher zu verwirren. Die genauen Gründe sind an dieser Stelle nicht zu klären.

In sich stimmig ist die Tatsache, dass bei falschem Nettofluss die bedingte Wahrscheinlichkeit, den Bestand korrekt anzugeben, noch geringer ist (also p(Bestand korrrekt|Nettofluss falsch)). Vermutlich nutzten die Versuchspersonen den ermittelten Nettofluss, um weiter zu rechnen, was aufgrund des fehler-

haften Nettoflussgraphen zu falschen Lösungsgraphen führte, oder sie wussten insgesamt nicht, wie sie vorgehen sollten.

3.4.3 Fehleranalyse

Hypothese 2.4 bestätigte sich: In der Baseline-Bedingung trat die Korrelations-Heuristik als systematischer Fehler auf. Die Manipulationen führten zu keiner Verbesserung. Somit besteht die Frage nicht mehr, ob die Korrelations-Heuristik in Abhängigkeit hilfreicher Modifikationen seltener auftritt (*Hypothese 2.5*). Die Korrelations-Heuristik trat in allen Bedingungen auf. Die Anteile für ihre Anwendung lagen eher unter dem Bereich der Anteile, die Cronin et al. (2009) fanden. Aus der Bedingung KS-geteilt ergab sich der Hinweis, dass die Versuchspersonen sich für die Korrelations-Heuristik eher am Nettofluss als am Zufluss orientierten. Das wiederum könnte dafür sprechen, dass einige Versuchspersonen den Nettofluss berechneten und ihn als Ergebnis ansahen (vgl. Kapitel 3.1.2).

Zu *Fragestellung 2.2* lässt sich sagen, es existiert eine weitere Variante der Korrelations-Heuristik. Das Phänomen von Korrelations-Heuristik-Diskontinuität lässt sich vermutlich damit erklären, dass die Versuchspersonen selber den Anfangspunkt für den Nettoflussgraphen einzeichnen sollten. Das war direkt aus

der Studie von Cronin et al. (2009) entnommen, für die die AutorInnen kein derartiges Fehlermuster beschrieben.

3.4.4 Einfluss der Kontrollvariablen

Entgegen Ergebnissen aus Resultaten vieler Studien (vgl. Kapitel 2.4), schnitten Männer nicht merklich besser ab als Frauen. Jedoch war der Anteil der Männer an der Stichprobe wiederum gering. Psychologie-Studierende schnitten in der vorliegenden Studie besser ab, was zu Ergebnissen aus der BF-Literatur passte. Dort zeigt sich teilweise ein Vorteil für Personen mit vergleichsweise mehr mathematischer Vorbildung (ebenfalls Kapitel 2.4).

3.4.5 Zusätzliche Analysen

Im Großen und Ganzen bestätigt sich die wahrgenommene Schwierigkeit der Aufgaben. Es ergab sich kein Unterschied, ob eine Versuchsperson an der vorgehenden Studie teilgenommen hatte oder nicht[6]. Es ergaben sich somit keinerlei Hinweise für Übungseffekte. Zudem zeigte sich innerhalb dieser Studie kein Positionseffekt. Das spricht ebenfalls gegen Übungseffekte. Ins-

6 Zu dem Zeitpunkt war dies am psychologischen Institut der TU Chemnitz (mit sehr hoher Wahrscheinlichkeit) die einzige Möglichkeit, derartige Aufgaben schon bearbeitet zu haben.

gesamt ist zu beanstanden, dass die Anzahl der Versuchspersonen für manche Analysen sehr gering war, so dass in diesen Fällen die Ergebnisse nur als erste Hinweise angesehen werden.

3.4.6 Fazit und Ausblick

Ausgangspunkt bei dieser Studie war es, dem häufig auftretenden Fehler-Verhalten „Anwenden der Korrelations-Heuristik“ abzuhelfen. Eine Betonung des Abflusses durch Nutzen des negativen Quadranten half nicht. Eine andere Überlegung war, dass die Versuchspersonen das Ergebnis der Addition von Zu- und Abfluss als Endergebnis ansahen. In dem Fall hätte es hilfreich sein können, wenn sie den Nettofluss als zusätzliche Größe erhalten. Leider half die Vorgabe des Nettoflusses ebenfalls nicht. Die Ursache für vermehrt falschen Lösungen in den Bedingungen NF-ermitteln und NF-gegeben waren nicht ganz klar.

Aus der vorliegenden Studie ergeben sich verschiedene, weitere Forschungsstränge. Da der Schwerpunkt dieser Dissertation bei der Diagrammdarstellung liegt, konzentriert sich die folgenden Studie auf diesen Aspekt. Wiederum soll eine Modifikation der Flussgrößen-Darstellung zu besserer BF-Leistung führen.

4 Die Darstellung von Flussgrößen als sukzessive Bestandsgrößen – Studie 3

Die Ergebnisse aus den bisherigen Studien der vorliegenden Arbeit ergaben keine nennenswerten Verbesserungen im BF-Verständnis (vgl. Kapitel 2 und 3). In diesem Kapitel erfolgt eine starke Modifikation der Diagrammdarstellung, um die BF-Leistung zu verbessern.

4.1 Einleitung

Um ein geeigneteres Darstellungsformat zu entwickeln, wandte ich mich der Frage zu, woher die Probleme bei BF-Systemen kommen. Verschiedene AutorInnen äußerten ihre Vermutungen dazu: Cronin und Gonzalez (2007) sahen die Ursache für das schlechte Abschneiden bei BF-Aufgaben eher in der graphischen Darstellung und weniger in einem kompletten Unverständnis für BF-Aufgaben. Booth Sweeney und Sterman folgerten aus ihren Ergebnissen, dass häufig intuitives Verständnis für BF-Aufgaben fehlt (2000). Cronin et al. (2009) sahen die schlechten Lösungsraten weniger als Folge ungünstiger Problem-Darbietung, sondern eher als Unwissen, was die Grundlagen von Akkumulation be-

trifft. Ossimitz (2002) vermutete, dass der Übergang von den Flussgrößen zum Bestand problematisch sei und dass viele Versuchspersonen Flussgrößen als Bestandsgrößen interpretierten. Die Idee für die aktuelle Studie ist es, das nach Ossimitz vorliegende Missverständnis zu nutzen: Die Flussgrößen werden nun in einen „temporären Bestand“ übersetzt. In einem BF-System, das ein Wasserbecken darstellt, sind die Flussgrößen ebenfalls wie eine Bestandsgröße dargestellt. Beispielsweise wird der Zufluss, der in einer Minute in das Becken fließt, zusammengefasst, und das Wasser befindet sich in einem Behälter. Derart soll der Zufluss pro Zeiteinheit eine Bestandsgröße sein, der dem „Gesamtbestand“ zugeführt wird. Anders ausgedrückt handelt es sich um Bewegungsgrößen, die jedoch für mehrere Zeiteinheiten ermittelt werden. Eine Bewegungsgröße besagt, wie groß die absolute Veränderung des Bestandes über den gesamten beobachteten Zeitraum ist (Ossimitz, 2001). Die Bewegungsgröße besitzt dieselbe Einheit wie der Bestand – beispielsweise Liter – während die Flussgröße eine Einheit pro Zeiteinheit aufweist – Liter pro Minute. Zumindest die Zusammenfassung der Flussgrößen, die in einer Minute in das Becken fließen, erfolgte schon durch eine Darstellung im Säulendiagramm. Auch in diesem Fall war beispielsweise der Zufluss pro Minute in einem Wert zusammengefasst, anders als beim Liniendiagramm, in dem sich der Zufluss

auch innerhalb einer Minute verändern kann. Das Zeigen eines Säulendiagramms statt eines Liniendiagramms führte zumindest zu einer tendenziellen Verbesserung (Cronin et al., 2009). In der vorliegenden Studie[7] wurde zusätzlich das Darstellungsformart verändert, so dass die Darstellung über Häufigkeiten erfolgte. Das ist beispielsweise gut erkenntlich, wenn es – wie beim „Kaufhaus-Problem“ – um Menschen geht. In diesem Fall ist es möglich, Piktogramme zu nutzen, wodurch natürliche Häufigkeiten vorliegen. Für das Lösen verschiedener Probleme wurde bereits der Vorteil des Verwendens natürlicher Häufigkeiten beschrieben (Kapitel 1.2.3). Ein entsprechender Versuch, das BF-Verständnis zu erhöhen, erfolgte in Kapitel 2: Dort wurden die BF-Größen Murmel und Wasser miteinander verglichen, was nicht eindeutig hilfreich war. In der aktuellen Studie sollte die Darstellung die Häufigkeiten noch mehr betonen – wie bereits erwähnt über das Nutzen von Piktogrammen.

Nach Domagks (2008, S.208) Einordnung von Bildern gehören Piktogramme zu den realistischen Bildern. Die einzige Modifikation, die in den bisherigen eigenen Studien zu einer recht eindeutigen Verbesserung führte, beinhaltete ein realistisches Bild: die schematische Abbildung (Kapitel 2.2.2.1). In der aktuel-

7 An dieser Stelle herzlichen Dank an Jana Wolf, die im Rahmen ihrer akademischen Qualifikationsarbeit die Datenerhebung und -erfassung durchführte.

len Studie enthielten die Flussdiagramme Piktogramme. Piktogramme im Alltag sollen vorhandenes Wissen aktivieren, sie sollen dazu führen, dass bestehendes Wissen beziehungsweise Schemata abgerufen werden (Weidenmann, 1994). Beispielsweise wissen Passagiere, wenn sie am Flughafen ein Piktogramm wie in Abbildung 4.1 sehen, wo die Abflüge stattfinden, wenn das entsprechende Vorwissen besteht.

Abbildung 4.1 Piktogramm für Abflüge.

In allen mir bis zu diesem Zeitpunkt bekannten Studien, in denen die Problemdarstellung verändert wurde, handelte es sich bei den Diagrammen um klassische Diagramme wie Linien- oder Säulendiagramme. Domagk (2008, S.209) spricht bei diesen Diagrammen von “konventionalisierter Darstellung“. Die Idee, für die Darstellung Piktogramme zu nutzen, weicht also stark von der bisherigen Darstellung der Flussgrößen ab.

In der vorliegenden BF-Studie waren Wasser und Personen als Bestand genutzt – so wie schon oft erfolgt (Booth Sweeney & Sterman, 2000; Cronin & Gonzalez, 2007; Cronin et al., 2009; Kainz & Ossimitz, 2002; Lungwitz et al., 2018; Schwarz, 2016; eigene Studien Kapitel 2 und 3). Die Darstellung über Häufigkeiten mit Piktogrammen von Personen als BF-Größe sollte ebenfalls helfen, das Problem besser zu verstehen, wenn es ähnlich funktioniert wie beispielsweise für Wahrscheinlichkeiten. Bei diesen führte eine Darstellung über natürliche Häufigkeiten zu verbesserter Leistung (vgl. Kapitel 1.2.3). Auf eine erneute Darstellung mit Murmeln (wie in Kapitel 2) wurde verzichtet, da die Versuchspersonen vermutlich Verständnisschwierigkeiten mit diesem Szenario hatten. Somit wurde auf die häufig genutzte BF-Größe Personen zurückgegriffen.

Wie verhält es sich nun mit der Größe Wasser? Es war nicht klar, ob für Wasser ein gleichermaßen günstiger Effekt für die BF-Leistung zu erwarten war, wenn die Darstellung über Piktogramme erfolgte – das bezeichne ich in dieser Arbeit als symbolische Darstellung. Diese sollte wiederum dazu führen, dass die Flussgrößen eher als Bestandsgröße pro Zeiteinheit wahrgenommen werden. Denn ein Behälter war für die Menge der Flussgröße pro Zeiteinheit und für die Menge des Bestands gegeben. Andererseits wurde keine Darstellung über Häufigkeit gewählt (z.B.

über Wassertropfen oder Anzahl an Wassereimern) sondern über eine Art Behälter mit Mengenangabe, der entsprechend dem Becken für den Bestand angesehen werden kann.

Welche der beiden Größen zu besserer BF-Leistung führen sollte, ist nach bisherigen Ergebnissen nicht klar vorherzusagen. Orientiert an Ergebnissen aus Forschung zu Wahrscheinlichkeiten, könnten die BF-Größe Personen einfacher verständlich sein als Wasser. Bei Wahrscheinlichkeiten führte die Darstellung über Häufigkeiten zu besseren Ergebnissen. In Kapitel 2.3.4 ergab sich kein Unterschied, ob Wasser oder Murmeln als BF-Größe genutzt wurden. Andererseits zeigte sich ein starker Positionseffekt – Murmeln an zweiter Stelle führten zu besserer BF-Leistung. Unter Umständen könnte also auch in BF-Szenarien die Darstellung über Häufigkeiten zu besseren Ergebnissen führen. Andererseits schnitten bei Cronin et al. (2009) die Versuchspersonen schlechter ab, die die Kaufhaus-Aufgabe bearbeiteten im Gegensatz zu den Versuchspersonen mit Auto- oder Badewannen-Szenario. Bei der Rechteck-Aufgabe ergab sich teilweise bessere Leistung für das Badewannen-Szenario als für das Konto-Szenario (Booth Sweeney & Sterman, 2000). In der Studie von Brunstein et al. (2010) zeigten sich wiederum die besten Lösungsraten für die Kaufhaus-Aufgabe mit der BF-Größe Personen im Gegensatz zu den anderen BF-Größen, die allesamt keine konkreten Häufigkei-

ten darstellten. Somit stellte sich die Frage, ob sich ein Unterschied für diese beiden Darstellungsformate ergibt.

Hypothese 3.1: In der symbolischen Bedingung mit der BF-Größe Wasser verbessert sich die BF-Leistung im Vergleich zu der Baseline-Bedingung.

Hypothese 3.2: In der symbolischen Bedingung mit der BF-Größe Personen verbessert sich die BF-Leistung im Vergleich zu der Baseline-Bedingung.

Fragestellung 3.1: Schneiden die Versuchspersonen in der Bedingung mit der BF-Größe Wasser und mit der BF-Größe Personen gleich gut ab?

Es fand wiederum eine Fehleranalyse statt.

Fragestellung 3.2: Welche Fehlerarten zeigen sich?

Als Kontrollvariablen waren Geschlecht und Studienfach erhoben. Für den Fall, dass diese Art der Änderung hilfreich war, sollte zusätzlich geprüft werden, ob die Versuchspersonen in einem nächsten Schritt bei der klassischen Diagramm-Darstellung (Liniendiagramm) leichter zur Lösung fanden, also ob Transfer auftrat. Reed (1999) klassifizierte verschiedene Arten von Transfer. Probleme mit gleicher Lösungsprozedur aber unterschiedlichem Kontext oder Aufgabenbeschreibung bezeichnete er als iso-

morphe Probleme. Wenn das Darstellungsformat der Flussgrößen als Aufgabenbeschreibung angesehen wird, so stellt ein Liniendiagramm bei den symbolischen Bedingungen ein isomorphes Problem dar. Ein Transfer von Beispielaufgaben zu isomorphen Problemen wurde immer wieder gefunden (Beispiele siehe Reed, 1999). Für die vorliegende Studie stellte sich die Frage nach Transfer-Leistung:

Fragestellung 3.3: Wenn die Versuchspersonen in den Bedingungen mit symbolischen Darstellungsformaten besser abschneiden als die Versuchspersonen der Baseline-Bedingung, schneiden sie dann auch in weiteren Problemen mit Liniendarstellung besser ab – findet Transfer statt?

Zusammenfassend lässt sich sagen: Im symbolischen Darstellungsformat wurden die Flussgrößen als „Bewegungsgröße für kleinere Zeitintervalle“ dargestellt, darin war alles zusammengefasst, das in der vorhergehenden kleinsten relevanten Zeiteinheit (z.B. Minute) zu- oder abgeflossen ist. Die Flussgrößen sahen wie kleine Bestandsgrößen aus und die Frage war, ob diese Veränderung zu besserem BF-Verständnis führte.

4.2 Methoden

4.2.1 Teilnehmende

Eine Poweranalyse im Rahmen der Abschlussarbeit wurde für das Nutzen einer Kontrastanalyse durchgeführt. Es sollte sich entsprechend des Vorgehens nach Sedlmeier und Renkewitz (2013) die Untergrenze der Power ergeben. Die Parameter wurden folgendermaßen festgelegt: eine Power von 90%, ein einseitiges α von .05 und ein mittelgroßer Effekt. Genutzt wurde wiederum G*Power. Es ergab sich eine erforderliche Stichprobengröße von 92 Versuchspersonen, die auf 90 Versuchspersonen abgerundet wurde, da es sich um drei Bedingungen handelte.

Neunzig Studierende mit einem mittleren Alter von 23 Jahren (*SD*: 4.2, Spannweite: 19-46 Jahre) nahmen am Experiment teil. 36 Personen (40%) waren Männer und 54 waren Frauen. 29 Versuchspersonen (32%) studierten ein Fach, das nah zu Naturwissenschaften war. Für diese Kategorien wurden die Versuchspersonen in zwei Gruppen unterteilt je nach ihrem Studienfach. Die eine Gruppe bestand aus Studierenden, die Fächer mit Schwerpunkt auf Mathematik beziehungsweise Naturwissenschaften studierten. Die andere Gruppe bestand aus Studierenden aus dem Bereich Geistes- und Sozialwissenschaften. Die Studie fand an der TU Chemnitz statt, und als Orientierung für die Un-

terteilung wurden ihre Fakultäten herangezogen. Studierende der Fakultäten für Mathematik, Informatik, Physik waren Studierende „nah den Naturwissenschaften“, Studierende der Philosophischen Fakultät waren als „fern Naturwissenschaften“ definiert. Zwar gibt es beispielsweise in der Psychologie (damals Philosophische Fakultät) auch Kurse mit mathematischen Inhalten, aber das Ausmaß ist geringer als in den anderen Fakultäten. In Anhang D sind alle Fächer und ihre Zuordnung aufgeführt. Beispiele für Fächer nahe den Naturwissenschaften sind Elektrotechnik oder Physik, während BWL, Erziehungswissenschaften oder Germanistik der Kategorie fern den Naturwissenschaften zugeordnet wurden. Die Versuchspersonen erhielten Versuchspersonenstunden oder Geld.

4.2.2 Materialien: die verwendeten Verläufe

Es handelte sich um einen Papier-und-Bleistift-Test. Insgesamt bearbeitete jede Versuchsperson drei Aufgaben derselben Darstellung, wobei insgesamt drei verschiedene Darstellungsformate existierten. Abschließend erhielten alle Versuchspersonen zwei Aufgaben, die nur in Form eines Liniendiagramms vorlagen, um Transfer-Leistung zu testen. Im folgenden beschreibe ich die Aufgaben und wie die Bewertung der Lösungsgüte erfolgte.

Die zwei kontinuierlichen Aufgaben Rechteck- und Dreieckverlauf waren gegeben und wurden leicht modifiziert: Es gab statt 16 nur 12 Zeiteinheiten. Die Anzahl der Zeitpunkte wurde aus gestalterischen Gründen geändert. In einer Studie von Cronin et al. (2009) führte eine Verringerung der Zeitpunkte bei gleichem Verlauf zu keiner nennenswerten Änderung in der Leistung. Somit wird angenommen, dass sich durch diese Veränderung der Schwierigkeitsgrad des Problems nicht oder zumindest nicht stark verändert. Weiterhin gab es eine Aufgabe mit diskontinuierlichem Flussgrößen-Verlauf, die einer Studie von Cronin et al. (2009) entstammte. Die Abbildungen 4.2, 4.3 und 4.4 zeigen die Aufgaben, die die Versuchspersonen in der Baseline-Bedingung erhielten.

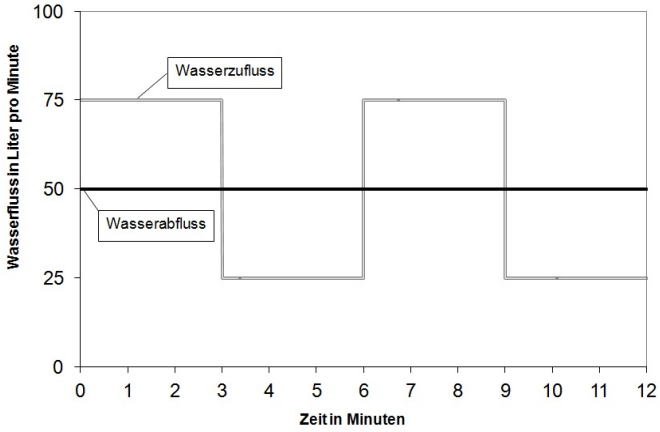


Abbildung 4.2 Rechteckverlauf in der Baseline-Bedingung (modifiziert nach Booth Sweeney & Sterman, 2000)

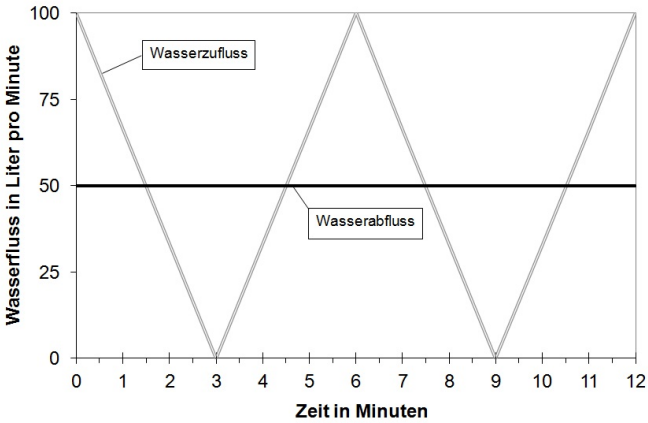


Abbildung 4.3 Dreieckverlauf in der Baseline-Bedingung (modifiziert nach Booth Sweeney & Sterman, 2000)

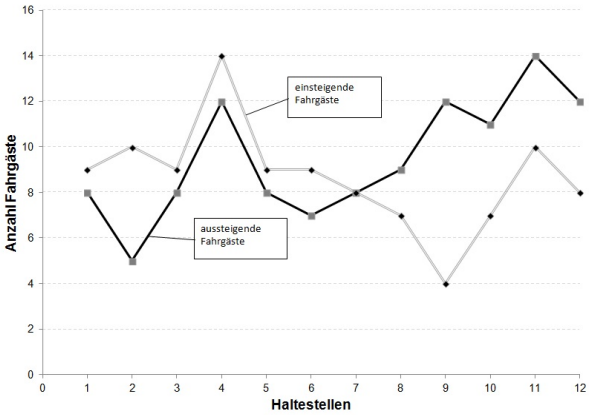


Abbildung 4.4 Der diskontinuierliche Flussgrößen-Verlauf in der Baseline-Bedingung (modifiziert nach Cronin et al., 2009)

Diese drei Aufgaben werden von nun an als “Durchgang1-Aufgaben” bezeichnet, um sie von den Transfer-Aufgaben abzugrenzen.

4.2.3 Modifikationen des Darstellungsformats

Es wurden drei verschiedene Darstellungsformate genutzt, und in jeder Bedingung war das Lösungsdiagramm dem Flussdiagramm angepasst.

4.2.3.1 Baseline-Bedingung

Die Flussgrößen wurden in der klassischen Diagramm-Darstellung als Liniendiagramm gezeigt (Abbildung 4.2, 4.3 und 4.4). Die BF-Größe war für die Probleme mit kontinuierlichen Flussgrößen Wasser und für das Problem mit diskontinuierlicher Flussgröße die Anzahl von Personen, die einen Bus nutzten. Es gab zusätzlich ein Lösungsdiagramm, in das die Versuchspersonen ihre ermittelte Lösung einzeichneten (Abbildung 4.5 für die BF-Größe Wasser).

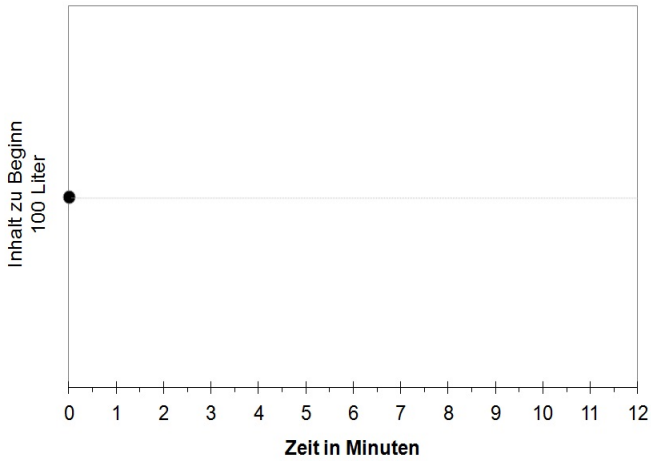


Abbildung 4.5 Lösungdiagramm für die Baseline-Bedingung.

Es existierten zwei Arten von modifiziertem Darstellungsformat. Die folgenden Abschnitte erläutern diese genauer.

4.2.3.2 Wasser-Bedingung

Bei der Aufgabe mit Wasserbecken-Szenario handelte es sich bei der BF-Größe um Wasser. Pro Zeiteinheit kam eine bestimmte Menge Wasser in das Wasserbecken hinzu oder heraus

(Abbildung 4.6[8]). Abgebildet wurden die Flussgrößen als Piktogramme in Form von schmalen Behältern mit Skalierung, an der der Wasserstand ersichtlich wurde. Die Lösung zeichneten die Versuchspersonen in ein Diagramm, das entsprechende leere Wasserbehälter enthielt; in diesen markierten die Versuchspersonen den aktuellen Wasserstand für jede Minute (Abbildung 4.7). Diese Bedingung wird im Folgenden „Wasser-Bedingung“ genannt.

8 Bestandteile dieser und folgender Abbildungen sind entweder selber erstellt oder von Seiten heruntergeladen, die Abbildungen frei anbieten [Download von Bilddokumenten] (o.D.). Ausnahmen werden extra erwähnt.

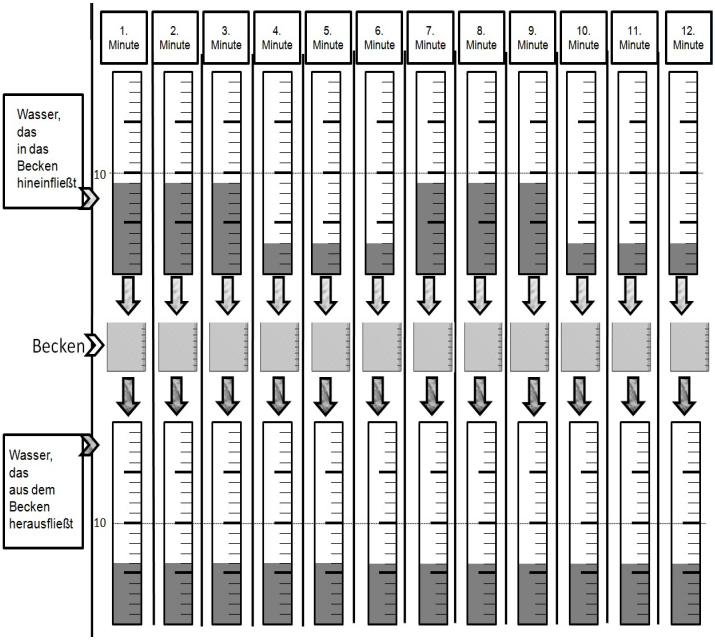


Abbildung 4.6 Rechteck-Verlauf in der Bedingung mit Wasserbecken-Szenario.

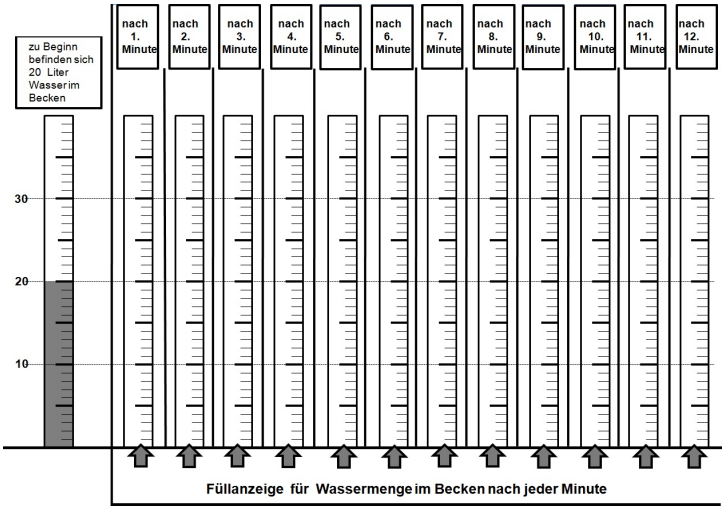


Abbildung 4.7 Lösungsdiagramm in der Bedingung mit Wasserbecken-Szenario.

4.2.3.3 Bus-Bedingung

Die Bedingung mit BF-Größe Personen wird als „Bus-Bedingung“ bezeichnet. Die Flussgrößen wurden durch Personen in Form von Piktogrammen gezeigt (Piktogramm-Menschen). Sie stehen für Personen, die in einen Bus einstiegen oder aus ihm

ausstiegen (Abbildung 4.8). Auch das zugehörige Lösungsdiagramm wurde angepasst: Die Versuchspersonen erhielten ein Lösungsdiagramm mit Kästchen, in dem sie für jede Person im Bus ein Kästchen markierten (z.B. ankreuzten oder ausmalten) (Abbildung 4.9).

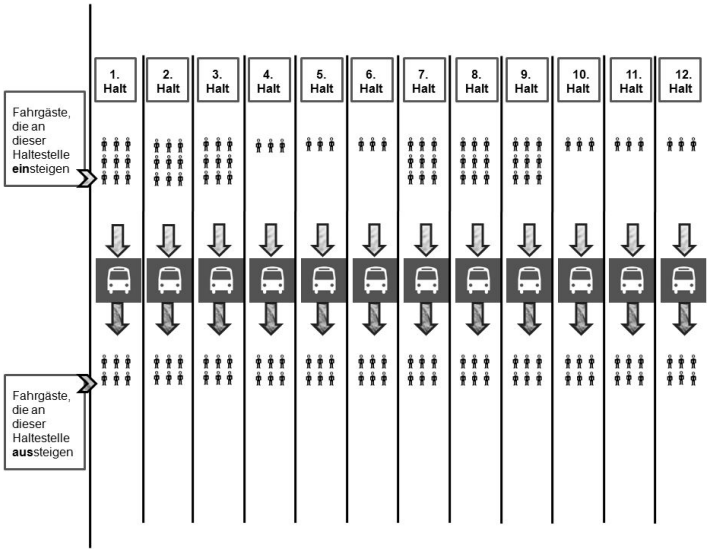


Abbildung 4.8 Rechteck-Verlauf in Bedingung mit BF-Größe Personen

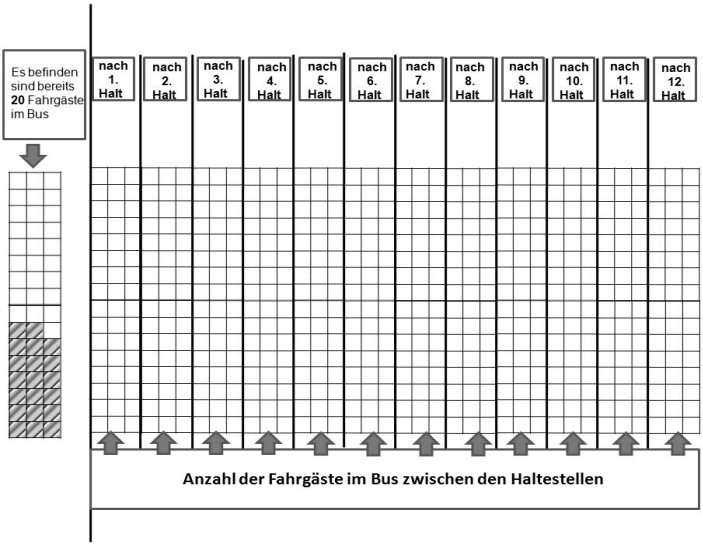


Abbildung 4.9 Lösungsdiagramm für die Bedingung mit Personen als BF-Größe

4.2.4 Abhängige Variablen

Die abhängige Variable war die Lösungsgüte, die wiederum in zwei Kriterien unterteilt war. Eines war „AV-Tendenz“: Für diese Variable erhielt eine Lösung einen Punkt, wenn Zu- und Abnahme des Bestands korrekt beurteilt waren. Für das zweite

Kriterium „AV-Steigung“ ergab sich ein Punkt, wenn eine Versuchsperson den exakten Verlauf korrekt ermittelte. Somit konnte die Lösungsgüte für *eine* Aufgabe bei einer Versuchsperson 0 bis 2 Punkte erreichen, und die Gesamtsumme konnte 0 bis 6 Punkte annehmen (vgl. Kapitel 2.2.3). Auf diese Art wurden die drei Durchgang1-Aufgaben in allen Bedingungen bewertet. Wiederum soll eine Fehleranalyse zu den Lösungen für den Bestand stattfinden. Die kontinuierliche Transfer-Aufgabe wurde entsprechend wie die Durchgang1-Aufgaben bewertet.

4.2.5 Transfer-Aufgaben

Zusätzlich sollte die Transfer-Leistung untersucht werden. Hierfür wurde einerseits die Leistung an sich in zusätzlichen Aufgaben in klassischer Diagramm-Darstellung untersucht. Das heißt, in einem zweiten Durchgang bearbeiteten alle Versuchspersonen zwei Aufgaben in Form von Liniendiagrammen, um den Transfer zu überprüfen. Es handelte sich dabei um eine Aufgabe mit kontinuierlichem Flussgrößen-Verlauf und immer Wasser als BF-Größe (Abbildung 4.10). Die andere Aufgabe zeigte einen diskontinuierlichen Flussgrößen-Verlauf und durchgängig Personen als BF-Größe (Abbildung 4.11).

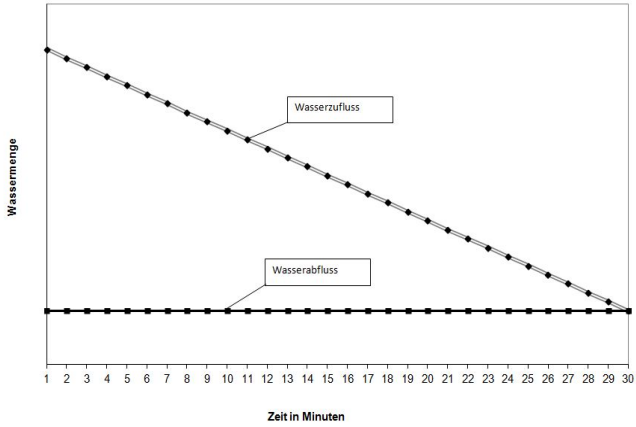


Abbildung 4.10 Kontinuierlicher Flussgrößen-Verlauf für Transfer-Aufgabe (modifiziert nach Cronin et al., 2009).

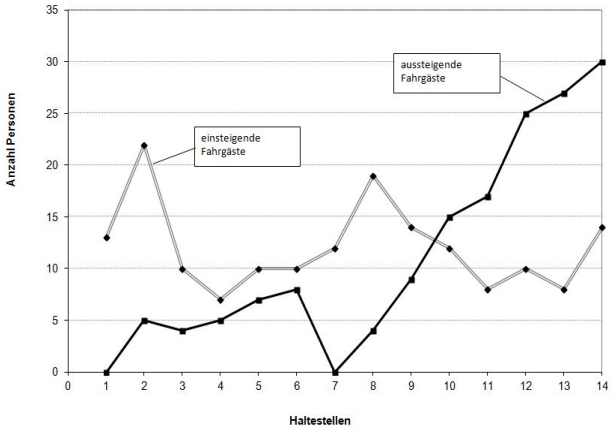


Abbildung 4.11 Diskontinuierlicher Flussgrößen-Verlauf für Transfer-Aufgabe (modifiziert nach Ossimitz et al., 2001).

Bei den Aufgaben mit den kontinuierlichen Flussgrößen sollten die Versuchspersonen einen Liniengraph für die Lösung in ein Diagramm einzeichnen. Die Lösung wurde wie oben beschrieben bewertet. Für die diskontinuierliche Transfer-Aufgabe wurde die Frage gestellt, wann der Bestand am größten war. Die Aufgabe war entweder korrekt (1 Punkt) oder falsch (0 Punkte). Diese Aufgabe war direkt von Ossimitz et al. (2001) übernom-

men. Zusätzlich war gefragt, anhand welcher Methode die Versuchspersonen feststellten, wann der Bestand am größten war. Außerdem sollten die Versuchspersonen vier Fragen zu den Flussgrößen beantworten (wann ist Zufluss bzw. Abfluss maximal und wann minimal).

Unterscheidet sich die Transfer-Leistung für die Versuchspersonen mit piktogrammartigen Darstellungsformat verglichen mit denen, die alle Aufgaben in Form von Liniendiagrammen erhielten? Hierfür wurden die korrekten Lösungen bei den Transfer-Aufgaben betrachtet und zusätzlich die bedingten Wahrscheinlichkeiten, die Transfer-Aufgaben zu lösen. Bei den bedingten Wahrscheinlichkeiten handelte es sich erstens um die Wahrscheinlichkeit, eine Transfer-Aufgabe korrekt zu lösen, wenn jeweils eine der Durchgang1-Aufgaben korrekt gelöst war. Zweitens wurden die Wahrscheinlichkeiten ermittelt, dass die Transfer-Aufgabe korrekt gelöst war, wenn die Versuchspersonen die jeweilige Durchgang1-Aufgabe falsch gelöst hatten. Diese Wahrscheinlichkeiten wurden miteinander verglichen, wobei eindeutig höhere Werte für die erstgenannten bedingten Wahrscheinlichkeiten auf Transfer-Leistung hindeuten sollten. Außerdem war eine Regressionsanalyse geplant, um zu prüfen, ob mit korrekten Lösungen im ersten Durchgang die Lösungen der Trans-

fer-Aufgaben gut vorhersagbar sind. Dabei ist der Einfluss der Bedingungen kontrolliert.

4.2.6 Weitere Variablen

Geschlecht und Studienfach wurden als mögliche Einflussgrößen untersucht. Jede Versuchsperson erhielt Fragen zu diesen Variablen und zu weiteren wie Alter und ob die Versuchsperson an vorhergehenden BF-Studien teilgenommen hatte. Außerdem kam die Frage nach dem Vorgehen bei den Aufgaben. Weiterhin sollten die Versuchspersonen Angaben machen zur Schwierigkeit aller Aufgaben insgesamt und Fragen beantworten zur Anstrengung, Frustration, Anspannung, Motivation und Verständlichkeit der Graphiken.

4.2.7 Ablauf

Für das Darstellungsformat lag ein between-subjects-Design vor: Jede Versuchsperson musste drei Durchgang1-Aufgabe lösen und erhielt alle im selben Darstellungsformat. Nachdem eine Versuchsperson die drei Aufgaben bearbeitet hatte, erhielt sie die zwei Transfer-Aufgaben. Die Reihenfolge der drei Durchgang1-Aufgaben wurde per lateinischem Quadrat ermittelt, wodurch sich drei Reihenfolgen ergaben. Diese wurden mit den zwei Rei-

henfolgen der Transfer-Aufgaben kombiniert. Also existierten 6 verschiedene Aufgabenreihenfolgen, denen die Versuchspersonen per Blockrandomisierung zugeordnet wurden. Abschließend füllte jede Versuchsperson einen Fragebogen aus zu oben erwähnten zusätzlichen Variablen.

4.3 Ergebnisse

Zunächst erfolgt ein Überblick über die Ergebnisse, dann gehe ich detailliert auf den Einfluss des Darstellungsformats ein. Es folgen Fehleranalyse und weitere Variablen.

4.3.1 Überblick

Für die BF-Leistung bei den Durchgang1-Aufgaben ergab sich einer mittlere prozentuale Leistung von 63% (als Mittelwert der Gesamtsumme: 3.76, *SD* = 2.00). 23 Versuchspersonen (26%) lösten alle Aufgaben komplett korrekt, und 10 Versuchspersonen (11%) erhielten 0 Punkte für die Gesamtsumme.

4.3.2 Einfluss des Darstellungsformats

Zunächst folgen die Ergebnisse zu *Hypothese 3.1* und *Hypothese 3.2*, die besagten, dass die symbolischen Darstellungen zu besserer Leistung führen. Die mittlere prozentuale Leistung lag

für die Baseline-Bedingung bei 36% (Mittelwert der Gesamtsumme: 2.17, *SD* = 1.95), für die Bedingung mit Wasser-Szenario bei 78% (Mittelwert der Gesamtsumme: 4.70, *SD* = 1.44) und beim Bus-Szenario bei 73% (Mittelwert der Gesamtsumme: 4.40, *SD* = 1.57). Boxplots für die Summe der Kriterien aller drei Aufgaben zeigen: Die Versuchspersonen in der Wasser-Bedingung schnitten am besten ab, gefolgt von den Personen in der Bus-Bedingung, die in ihrer Verteilung leicht schlechter abschnitten, aber mit demselben Median. Der Median in der Baseline-Bedingung war viel geringer (Abbildung 4.12).

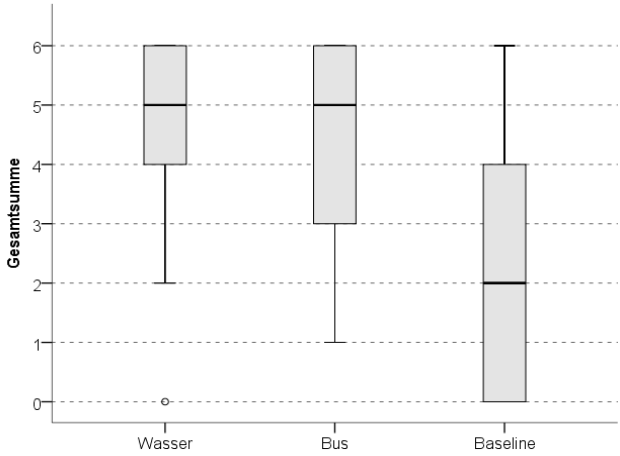


Abbildung 4.12 Boxplots der Gesamtleistung für alle drei Durchgang1-Aufgaben in den jeweiligen Bedingungen.

Es folgen die Anteile für die Personen, die alle drei Aufgaben komplett korrekt lösten: In der Baseline-Bedingung waren es 7%, in der Wasser-Bedingung 40% und in der Bus-Bedingung 30%. Abbildung 4.13 zeigt die Mittelwerte der korrekten Lösungen der Durchgang1-Aufgaben in den drei Bedingungen.

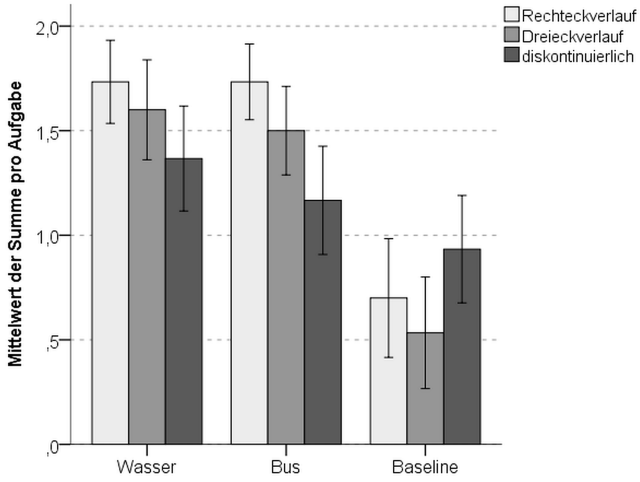


Abbildung 4.13 Leistung für die drei einzelnen Aufgaben in den verschiedenen Bedingungen (*n* = 30 pro Säule mit 90% Konfidenzintervallen).

Die Leistung in den veränderten Bedingungen stieg sehr stark an: am stärksten für die Aufgaben mit Rechteck-Verlauf und am geringsten für die diskontinuierliche Aufgabe.

4.3.2.1 Regressionsanalyse

Es erfolgte eine multiple Regressionsanalyse für korrekte Lösungen (Gesamtsumme) als abhängige Variable und mit den Prädiktoren Darstellungsformat, Geschlecht und Studienfach. Die Bedingungen der verschiedenen Darstellungsformate waren dummy-kodierte Variablen. Tabelle 4.1 zeigt die standardisierten Koeffizienten (β). Der erklärte Varianzanteil wich signifikant von 0 ab: Mehr als ein Drittel der Variabilität in der BF-Leistung wurde durch die Variablen Format, Geschlecht, Studienfach und Teilnahme erklärt.

Tabelle 4.1
Multiple Regressionsanalyse der Variablen für die Bedingungen und Personenvariablen mit der abhängigen Variable Gesamtsumme als Indikator für BF-Leistung.

Prädiktor	β	*t*	*p*	*pr*
Wasser-Darstellung	.59	5.977	<.00	.55
Bus-Darstellung	.53	5.371	<.00	.51
Geschlecht	.13	1.469	.07	.16
Studienfach	.19	2.135	.02	.23
Teilnahme	.09	1.079	.14	.12

Anmerkung: $F(5,84) = 10.857$, $p < .00$ mit $R^2 = .39$. Dummy-kodierte Variablen sind kodiert: Wasser-Darstellung: nein 0, ja 1; Bus-Darstellung: nein 0, ja 1; Geschlecht: weiblich 0, männlich 1; Studienfach: den Naturwissenschaften fern 0, nah 1; Teilnahme: noch nicht an BF-Studie teilgenommen 0, bereits teilgenommen 1.

Für die Bedingungen mit symbolischem Darstellungsformat ergaben sich hohe Effektstärken, die sich signifikant von 0 unterschieden. Somit bestätigten sich *Hypothese 3.1* und *3.2:* Die BF-Leistung verbesserte sich in den symbolischen Bedingungen mit BF-Größe Wasser und Personen im Vergleich zur Baseline-Bedingung.

4.3.2.2 Vergleich von Wasser- und Bus-Bedingung

Schneiden die Versuchspersonen in den beiden symbolischen Bedingungen gleich gut ab (*Fragestellung 3.1)*? Schon obige Abbildungen und die ähnlich großen Effektstärken der Regressionsanalyse deuteten darauf hin, dass sich die Leistung in den beiden Bedingungen kaum unterschied. Die Mittelwerte der Gesamtsumme lagen nah beieinander, und die 95%-Konfidenzintervalle überlappten sich stark (Abbildung 4.14) – 95%-Konfidenzintervalle wurden genutzt, da es sich um keine gerichtete Hypothese handelte. Aufgrund dieser Abbildungen wird von einer inferenziellen Analyse abgesehen.

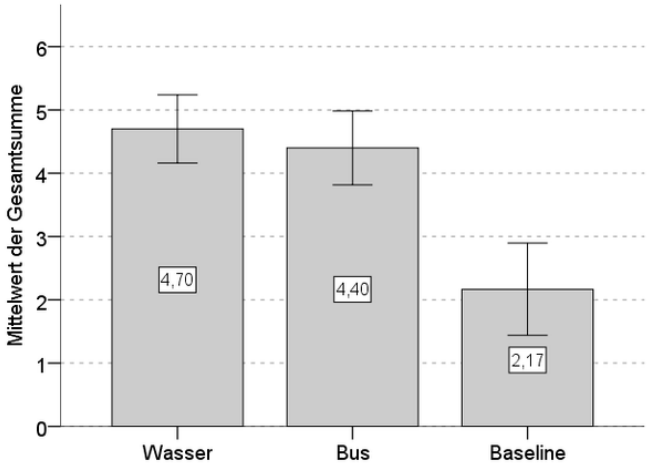


Abbildung 4.14 Mittelwerte für die Gesamtsummen in den verschiedenen Bedingungen mit 95% Konfidenzintervallen.

Die Fragestellung lässt sich verneinen: Die Leistung unterschied sich nicht bedeutsam in den symbolischen Bedingungen.

4.3.3 Fehleranalyse

Es folgen Analysen betreffend *Fragestellung 3.2* nach den Fehlerarten. Systematische Fehler ergaben sich durch das Anwenden der Korrelations-Heuristik. Für den ersten Durchgang

der Fehleranalyse nutzte ich folgende Kategorien: korrekt, Korrelations-Heuristik angewandt, Rechenfehler, mehr als zwei Fehler und sonstige Fehler. Die Kategorie „Rechenfehler“ umfasste die Lösungen, bei denen im Großen und Ganzen die Versuchsperson korrekt vorging, an einzelnen Stellen einen falschen Wert addiert hatte (z.B. +4 statt +3) oder einen Wert statt subtrahierte statt zu addieren vice versa (z.B. +3 statt -3).

Die Kategorie „mehr als 2 Fehler“ enthielt die Lösungen, in der drei oder mehr derartiger Rechenfehler vorkamen, um möglicherweise nach der Ursache hierfür oder nach Besonderheiten zu suchen. Ein derartiger Fall lag für die Aufgabe mit Rechteckverlauf vor: Vereinzelt war eine Differenz falsch berechnet und wurde dann mehrmals genutzt und zwar in dem Bereich, in dem der Zufluss konstant blieb. Beispielsweise wurde eine Differenz von -2 statt -3 ermittelt und diese für die Zeitpunkte genutzt, in denen der Zu- und Abfluss konstant waren. In der vorliegenden Studie kam „mehr als 2 Fehler“ insgesamt sehr selten vor, so dass ein Zusammenfassen der beiden Kategorien vermutlich keinen Informationsverlust darstellt. Die Kategorie „sonstige Fehler“ enthielt die Fehler, die nicht in die anderen Kategorien passten und bei denen keine Systematik entdeckt wurde.

Es gilt zu bedenken, dass ein einfacher Rechenfehler zu einer Lösungsgüte „komplett falsch“ führen kann, wenn bspw. -3

statt +3 gerechnet wurde. Somit konnte das Prinzip verstanden sein, aber aufgrund eines einfachen Rechenfehlers ergab sich eine falsche Lösung. Dennoch blieb ich bei dieser Beurteilung der Lösungen. Ein Kriterium „Prinzip verstanden“ wäre schwierig umsetzbar, da insbesondere für sonstige Fehlerarten teilweise nicht erkenntlich war, ob das Prinzip verstanden wurde oder nicht. Die folgenden Abbildungen (4.15 bis 4.17) zeigen, wie oft die verschiedenen Fehlerkategorien in den einzelnen Bedingungen auftraten.

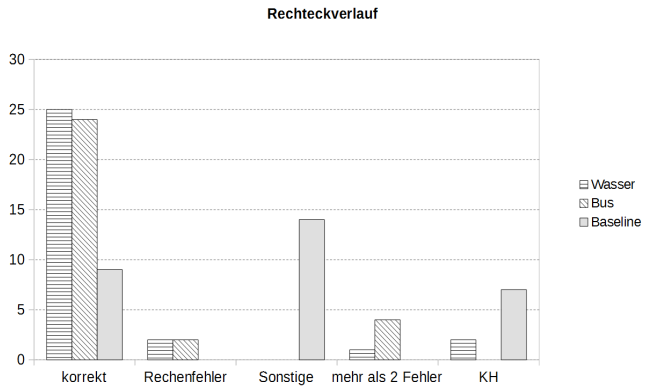


Abbildung 4.15 Häufigkeit der einzelnen Fehlerkategorien für die Aufgabe mit Rechteckverlauf.
Anmerkung: KH bedeutet Korrelations-Heuristik.

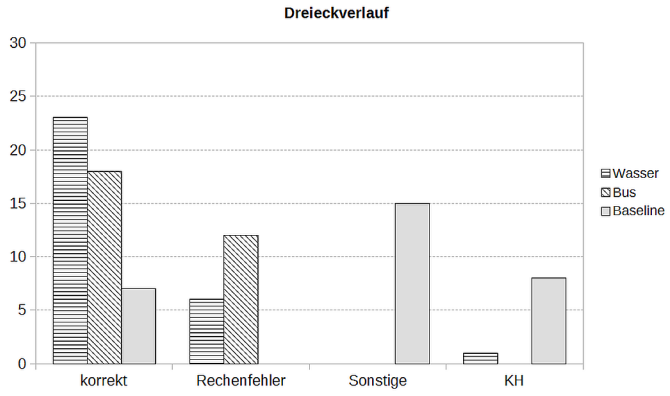


Abbildung 4.16 Häufigkeit der einzelnen Fehlerkategorien für die Aufgabe mit Dreieckverlauf.
Anmerkung: KH bedeutet Korrelations-Heuristik.

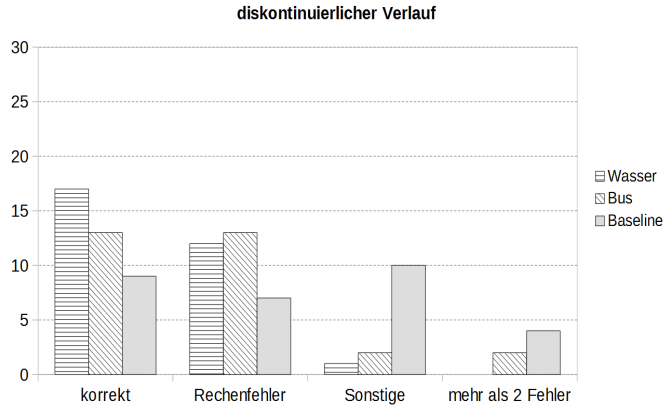


Abbildung 4.17 Häufigkeit der einzelnen Fehlerkategorien für die Aufgabe mit diskontinuierlichem Verlauf.

Beim symbolischen Darstellungsformat traten in erster Linie Rechenfehler auf, am häufigsten beim Dreieckverlauf und beim diskontinuierlichem Verlauf. In der Bus-Bedingung traten mehr Rechenfehler auf als in der Wasser-Bedingung. In der Baseline-Bedingung zeigten sich Rechenfehler nur für den diskontinuierlichen Verlauf. Die Korrelations-Heuristik wandte niemand in den symbolischen Bedingungen an, abgesehen von zwei Versuchspersonen in der Wasser-Bedingung. Eine gab bei beiden kontinuierlichen Flussgrößen-Verläufen einen gleichbleibenden Wert als Be-

stand an, was nach Korrelations-Heuristik orientiert am Abfluss aussah. In der Baseline-Bedingung verwendeten Versuchspersonen die Korrelations-Heuristik, außer beim diskontinuierlichen Verlauf. Die Kategorie „sonstige Fehler“ trat in Bedingungen mit symbolischer Darstellung nur beim diskontinuierlichen Flussgrößen-Verlauf auf, während sie in der Baseline-Bedingung für alle Verläufe und weitaus häufiger vorkam.

4.3.4 Kontrollvariablen Geschlecht und Studienfach

Zunächst gehe ich auf die Variablen Geschlecht und Studienfach ein. Daraufhin folgen Analysen zum Transfer und abschließend Analysen zu Positionseffekten. Grundlegend waren die Verteilungen der Merkmale Geschlecht und Studienfach in den verschiedenen Bedingungen ungefähr ausgeglichen.

4.3.4.1 Geschlecht

Der Anteil der Männer lag bei etwa 35% bis 45% in den einzelnen Bedingungen (detailliert in Anhang E). Männer schnitten tendenziell besser ab als Frauen, wobei die Unterschiede in der Leistung in erster Linie auf die Baseline-Bedingung zurückgingen (Abbildung 4.18). In der Regressionsanalyse ergab sich für die Variable Geschlecht ein β von .13 mit p = . 07 und pr von .16.

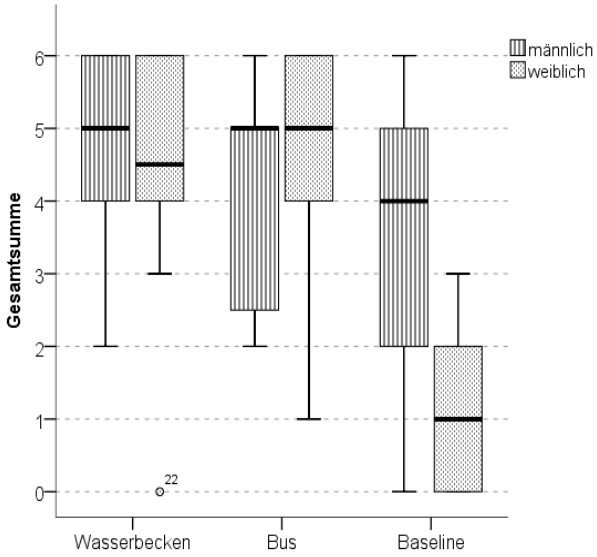


Abbildung 4.18: Boxplots für die Gesamtsumme getrennt nach Bedingung und Geschlecht.

4.3.4.2 Studienfach

Innerhalb der einzelnen Bedingungen lagen die Anteile von "nah den Naturwissenschaften" bei etwa einem Drittel der Versuchspersonen in jeder Bedingung: Sie schwanken von 30% bis

37% (detailliert in Anhang F). Nach den Medianen zu urteilen, schnitten Versuchspersonen aus naturwissenschaftlich nahen Studienfächern besser ab, außer in der Bus-Darstellung: Dort lagen die Mediane auf dem gleichen Niveau (Abbildung 4.19). In der Regressionsanalyse ergab sich bei dieser Variable für β der Wert .19 mit $p = .02$ und ein *pr* von .23.

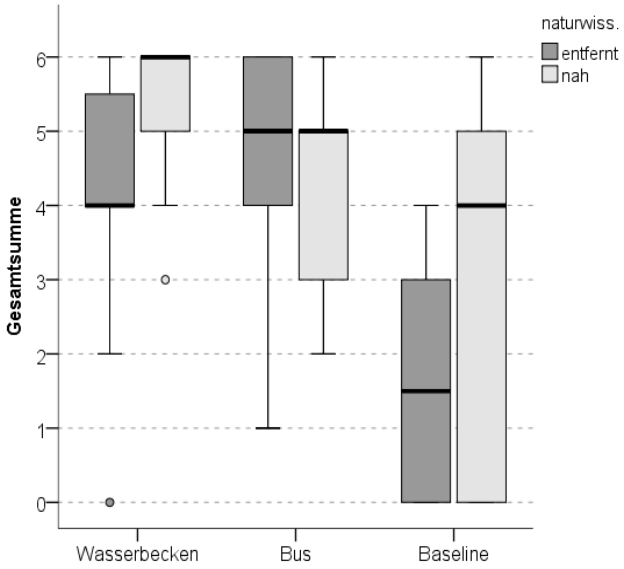


Abbildung 4.19 Boxplots für die Gesamtsumme getrennt nach Bedingung und Studienfach (naturwissenschaftlich nah versus entfernt).

4.3.4.3 Interaktionsterme

Aus der graphischen Analyse ergaben sich Hinweise, dass für Geschlecht und Bedingung eine Interaktion vorliegen könnte,

möglicherweise auch für Studienfach und Bedingung. Dies wurde mit einer zusätzlichen Regressionsanalyse überprüft. Für diese Analyse wurden die beiden symbolischen Bedingungen zusammengefasst und mit der Baseline-Bedingung verglichen. Somit ergibt sich eine dummy-kodierte Variable für die Bedingungen, weiterhin bleiben die Prädiktoren Geschlecht, Studienfach und Teilnahme. Dazu kommen die Interaktionsterme (Vorgehen beschrieben in Sedlmeier & Renkewitz, 2013).

In Tabelle 4.2 wird ersichtlich, dass tatsächlich eine signifikante Interaktion für Geschlecht und Bedingung vorlag, während keine Interaktion für Bedingung und Studienfach vorlag (diese *p*-Werte sind für die Interaktionen als ungerichtete Hypothesen berichtet).

Tabelle 4.2
Werte der Regressionsanalyse des vollständigen Modells mit einer Variable zu den Bedingungen, den Kontrollvariablen und den Interaktionen. Die abhängige Variable ist die Gesamtsumme.

Prädiktor	β	*t*	*p*	*pr*
Symbolische Bedingung	.62	7.188	< .00	.57
Geschlecht	.12	1.488	.07	.16
Studienfach	.21	2.489	.01	.26
Teilnahme	.10	1.248	.11	.14
Interaktion: Bedingung x Geschlecht	.29	-3.547	< .00	.36
Interaktion: Bedingung x Studfach	-.07	-.744	.46	-.08

Anmerkung: $R^2 = .49$ und $F(6,83) = 13.115$ mit $p < .00$. Symbolische Bedingung: 0 Baseline, 1 Bus- und Wasser-Darstellung; Geschlecht: weiblich 0, männlich 1; Studienfach: Naturwissenschaften fern 0, nah 1; Teilnahme: bisher noch nicht an BF-Studie teilgenommen 0, bereits teilgenommen 1.

4.3.5 Transfer-Leistung

Im weiteren Verlauf werden die Ergebnisse zu *Fragestellung 3.3* nach einem Transfer berichtet. Abbildung 4.20 zeigt Boxplots für die BF-Leistung der Transfer-Aufgaben in den verschiedenen Bedingungen. Bei der diskontinuierliche Transfer-

Aufgabe beantworteten die Versuchspersonen die Frage nach dem maximalen Bestand, somit erhielten sie für eine korrekte Antwort 1 Punkt. Bei der kontinuierlichen Transfer-Aufgabe zeichneten die Versuchspersonen einen Lösungsgraphen, der danach beurteilt wurde, ob AV-Tendenz und AV-Steigung korrekt waren. Somit kann die Summe der zusammengefassten Werte für beide Aufgaben 0 bis 3 Punkte annehmen. Die Leistung befand sich jeweils auf ähnlichem Niveau. Wenn der Median betrachtet wird, schnitt die Bus-Bedingung tendenziell am schlechtesten ab.

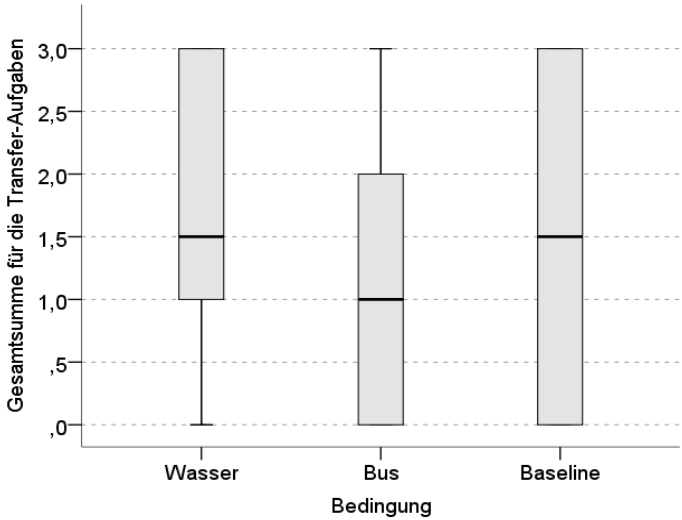


Abbildung 4.20 Leistung für die zwei Transfer-Aufgaben (Gesamtsumme: 0 – 3)

Es folgen die Anteile der Personen, die die Transfer-Aufgaben komplett korrekt lösten, getrennt nach den verschiedenen Bedingungen (Abbildung 4.21). Die Versuchspersonen in der Bus-Bedingung zeigten besonders schlechte BF-Leistung. Versuchspersonen in Wasser- und Baseline-Bedingung waren bei der kontinuierlichen Zusatz-Aufgabe etwa gleich gut, während die

Versuchspersonen in der Wasser-Bedingung beim diskontinuierlichen Problem leicht besser abschnitten.

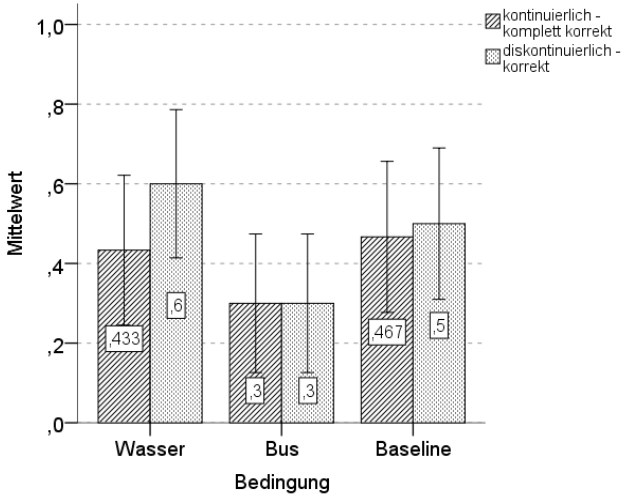


Abbildung 4.21 Leistung für die einzelnen Transfer-Aufgaben. Anmerkung: Die Mittelwerte entsprechen den Anteilen derer, die die jeweilige Transfer-Aufgabe komplett korrekt lösten; mit 95%-Konfidenz-Intervallen

Bedingte Wahrscheinlichkeiten für das korrekte Lösen der Transfer-Aufgaben

Als Zeichen für einen Transfer wurde folgendes Kriterium festgelegt: Es ist ein Zeichen für Transfer, wenn sich bei einer korrekten Durchgang1-Aufgabe und einer korrekt gelösten Transfer-Aufgabe als bedingte Wahrscheinlichkeit eindeutig höhere Werte ergeben als für die Wahrscheinlichkeit, im ersten Durchgang falsch und im zweiten korrekt zu lösen.

Zunächst bestimmte ich die Anzahl der Versuchspersonen, die die jeweilige Durchgang1-Aufgabe korrekt gelöst hatten. Dann ermittelte ich die Rate der Versuchspersonen, die zusätzlich die kontinuierliche Transfer-Aufgabe und die diskontinuierliche Transfer-Aufgabe korrekt gelöst hatten. Da die Raten der KorrektlöserInnen in ihrer Aussagekraft davon abhängen, wie viele Personen in den Anteilen verrechnet sind, folgen Abbildungen, die die Personenanzahl zusätzlich berücksichtigen. Abbildung 4.22 zeigt, wie viele der Versuchspersonen, die die Durchgang1-Aufgabe korrekt lösten, ebenfalls die kontinuierliche Transfer-Aufgabe korrekt lösen. Die gesamten Säulen zeigen die Anzahl derer, die die Durchgang1-Aufgabe komplett korrekt gelöst hatten; der graue Anteil darin zeigt die Anzahl der Personen, die zusätzlich die Transfer-Aufgabe korrekt lösten. Beispielsweise lösten in der Wasserbedingung 25 Personen die Aufgabe mit

Rechteckverlauf korrekt. 44% (11 Personen) von ihnen lösten auch die kontinuierliche Transfer-Aufgabe korrekt.

Die höchsten Werte für die bedingten Wahrscheinlichkeiten traten in der Baseline-Bedingung auf, außer bei der diskontinuierlichen Aufgabe. Dort lag etwa der gleiche Wert wie in der Wasser-Bedingung vor. Bei der Wasser-Bedingung und in der Bus-Bedingung ergaben sich insgesamt weitaus niedrigere Werte für die bedingten Wahrscheinlichkeiten. Alle Anteile lagen unter 50%.

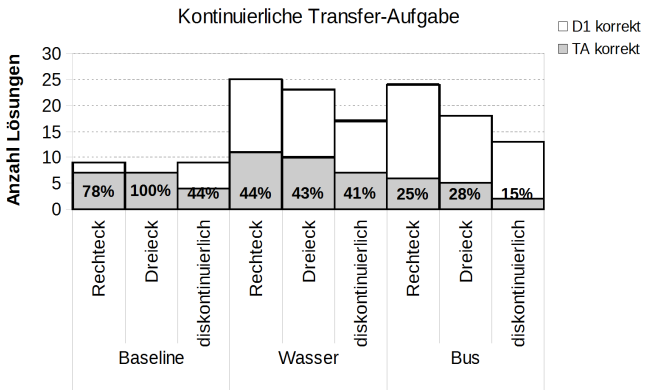


Abbildung 4.22 Anteil der Korrektlösenden der kontinuierlichen Transfer-Aufgabe (TA) an den Korrektlösenden der jeweiligen Durchgang1-Aufgabe (D1) - mit AV Steigung korrekt in Durchgang1-Aufgabe

Für die Berechnung der oben aufgeführten Werte standen in der Baseline-Bedingung die wenigsten Personen zur Verfügung. Nur weniger als 10 Versuchspersonen hatten korrekte Lösungen in den Durchgang1-Aufgaben ermittelt.

Es folgen die Wahrscheinlichkeiten für die korrekt gelösten Transfer-Aufgaben, wenn die ersten Aufgaben falsch gelöst wurden (Abbildung 4.23). Insgesamt sind die Werte in der Baseline-Bedingung eindeutig niedriger als in obiger Abbildung, bei Was-

ser-Bedingung in etwa gleich und in der Bus-Bedingung sogar etwas höher. In der Baseline-Bedingung scheint Transfer vorzuliegen und in den anderen Bedingungen nicht. Werte, die aus der Reihe fielen, lagen für die diskontinuierliche Durchgang1-Aufgabe in der Baseline-Bedingung vor und für die Durchgang1-Aufgabe mit Rechteck-Verlauf in der Bus-Bedingung. In beiden Fällen lagen die Werte ca. bei 50% und waren höher als die anderen bedingten Wahrscheinlichkeiten in der jeweiligen Bedingung.

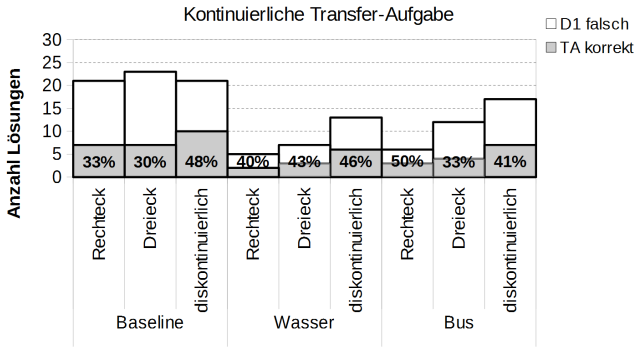


Abbildung 4.23 Anteil der Korrektlösenden der kontinuierlichen Transfer-Aufgabe (TA) an den Falschlösenden der jeweiligen Durchgang1-Aufgabe (D1)

Es folgt Abbildung 4.24 mit den bedingten Wahrscheinlichkeiten für die diskontinuierliche Transfer-Aufgabe p(Transfer-Aufgabe korrekt|Durchgang1 korrekt). Hier zeigt sich ein leicht anderes Muster. Die Basisraten entsprachen den oben beschriebenen. Wiederum schnitt die Baseline-Bedingung am besten ab: Die bedingten Wahrscheinlichkeiten lagen bei 67% und mehr. Anders als oben ergaben sich hier höhere Anteile in der Wasser-Bedingung: In der Wasser-Bedingung lagen die bedingten Wahrscheinlichkeiten um die 60%, also immer noch niedriger als in der Baseline-Bedingung. In der Bus-Bedingung lagen die Werte um die 30%.

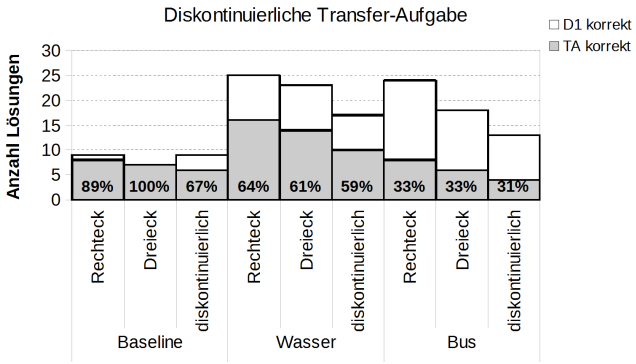


Abbildung 4.24 Anteil der Korrektlösenden der diskontinuierlichen Transfer-Aufgabe (TA) an den Korrektlösenden der jeweiligen Durchgang1-Aufgabe (D1)

Abschließend folgen die bedingten Wahrscheinlichkeiten p(Transfer-Aufgabe korrekt|Durchgang1 falsch) für die diskontinuierliche Transfer-Aufgabe (Abbildung 4.25). Die Werte der bedingten Wahrscheinlichkeiten in der Baseline-Bedingung waren wiederum durchgängig niedriger als im vorhergehenden Fall. Dasselbe gilt für die Bus-Bedingung. In der Wasser-Bedingung waren die Werte von Dreieck-Verlauf und diskontinuierlichem Flussgrößen-Verlauf ebenfalls um die 60%, außer beim Rechteck-Verlauf – hier ergab sich ein geringerer Wert.

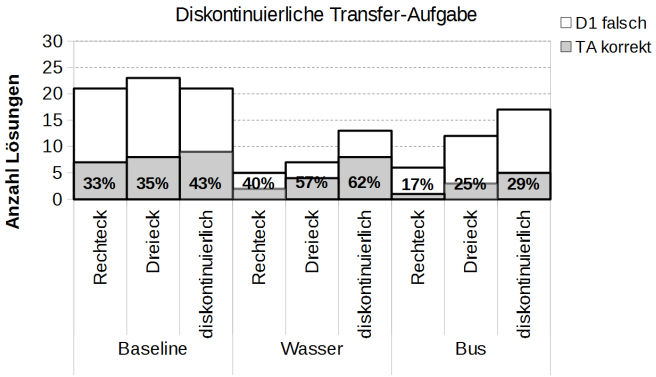


Abbildung 4.25 Anteil der Korrektlösenden der diskontinuierlichen Transfer-Aufgabe (TA) an den Falschlösenden der jeweiligen Durchgang1-Aufgabe (D1)

Diese Werte sprachen eher gegen eine Transfer-Leistung bei den neuen Darstellungsformaten: Die Prozentzahlen waren meist gleich groß wie bei der bedingten Wahrscheinlichkeit p(Transfer-Aufgabe korrekt|Durchgang1 korrekt). In der Baseline schien jedoch wieder ein Transfer vorzuliegen.

Da in den Abbildungen im Großen und Ganzen kein Hinweis für Transfer-Leistung in den symbolischen Darstellungsformaten vorlag, wurde auf eine Regressionsanalyse verzichtet. Zusammenfassend lässt sich sagen, dass die veränderten

Darstellungsformate zu verbesserter Leistung führten. Die Effektstärken waren sehr hoch verglichen mit früheren Effektstärken aus BF-Studien. Es fand jedoch kein Transfer statt.

4.3.6 Zusätzliche Analysen

4.3.6.1 Positionseffekt – bei Durchgang1-Aufgaben

Auch Positionseffekte wurden untersucht: Unterschied sich die Leistung für die jeweiligen Flussgrößen-Verläufe nur aufgrund ihrer Position, an der sie bearbeitet wurden? Varianzanalysen für die drei Aufgaben ergaben für den Rechteckverlauf: $F(2,87) = .014$ mit $p = .99$; für den Dreieckverlauf: $F(2,87) = 3.724$ mit $p = .03$ und für den diskontinuierlichen Verlauf: $F(2,87) = .975$ mit $p = .38$. Es trat somit nur für den Dreieckverlauf ein Positionseffekt auf. Post-hoc t-tests (unter Anwendung eines Bonferroni korrigierten $\alpha = .017$) deuteten darauf hin, dass die Aufgabe mit Dreieckverlauf an dritter Stelle besser abschnitt als an erster ($t = 2.482$ mit $p = .016$). Keiner der anderen Unterschiede wurde signifikant, auch nicht wenn dies für die Bedingungen einzeln ausgewertet wurde. Die Aufgaben wurden behandelt, als ob kein Positionseffekt auftrat: Da sich die Aufgabe mit Dreieckverlauf in jeder Bedingung gleich häufig an jeder

Position befand, sollte sich der Einfluss eines Positionseffektes ausgleichen.

4.3.6.2 Positionseffekt – bei Transfer-Aufgaben

Es zeigte sich kein Positionseffekt bei den Transfer-Aufgaben, jedoch schnitt die erste der beiden Transfer-Aufgaben in der Bus-Darstellung und Baseline-Bedingung leicht schlechter ab. Anders bei der Wasser-Darstellung, bei der die Leistung immer für das diskontinuierliche Problem besser war, unabhängig von der Position. Varianzanalysen ergaben für die kontinuierliche Aufgabe $F(1,88) = .629$ mit $p = .43$ und für die diskontinuierliche Transfer-Aufgabe $F(1,88) = 1.600$ mit $p = .21$.

4.4 Diskussion

Zunächst gehe ich auf die Ergebnisse zum Darstellungsformat ein, dann auf die Fehleranalyse. Im weiteren Verlauf werden die Transfer-Leistung, Ergebnisse bezüglich anderer Variablen und mögliche Wirkfaktoren diskutiert.

4.4.1 Das symbolische Darstellungsformat

Das symbolische Darstellungsformat führte zu verbesserter BF-Leistung. Die Versuchspersonen lösten in diesen Bedingungen die BF-Aufgaben durchweg besser. Die entsprechenden Ef-

fektstärken waren sehr hoch. Beim Rechteckverlauf ergaben sich in beiden Bedingungen mit symbolischer Darstellung circa 85% für die mittlere prozentuale Leistung; beim Dreieckverlauf waren es 80% (Wasser-Szenario) und 75% (Bus-Szenario). Die Spannweite aus einigen ähnlichen Angaben aus der Literatur liegt bei 23% bis 83% für den Rechteckverlauf und von 14% bis 52% beim Dreieckverlauf (Armenia et al., 2004; Booth Sweeney & Sterman, 2000; Kapmeier et al., 2014; Ossimitz, 2002; Schwarz, 2016; eigene Studien: Kapitel 2). Somit wurde für den Rechteckverlauf vereinzelt eine vergleichbar hohe BF-Leistung gefunden wie in der vorliegenden Studie, beispielsweise für das Badewannen-Szenario in der Studie von Booth Sweeney und Sterman (2000) am MIT. Jedoch handelt es sich dabei um einen ungewöhnlich hohen Wert, der in Stichproben mit anderer Zusammensetzung nicht gefunden wurde. In einer Studie von Ossimitz (2002) lag der Anteil für den Rechteckverlauf bei 42%, was eher dem sonst in dieser Studie gefundenen Leistungsniveau entsprach. Weitere Werte aus entsprechenden Studien an der TU Chemnitz sind 56% (Abschnitt 2.3.1) über alle Bedingungen gemittelt und in der Bedingung entsprechend Baseline-Bedingung knapp 50% (7 von 16 Versuchspersonen). Schwarz (2016) berichtete Werte für den Rechteckverlauf in einem Range von 23% bis 61% (erhoben an der TU Chemnitz, jedoch über verschiedene

Bedingungen hinweg, nicht nur für die Baseline). Für beide Verläufe ergab sich also in der vorliegenden Studie ein mittlerer prozentualer Anteil der Leistung, der besser ausfiel als die Leistung in den meisten bisherigen Studien. Für den Dreieckverlauf kommt das stärker zum Tragen.

Der diskontinuierliche Verlauf schnitt in den symbolischen Bedingungen schlechter ab als die kontinuierlichen Verläufe: 69% beim Wasser-Szenario und 59% beim Bus-Szenario. Cronin et al. (2009) fanden für den selben Verlauf 56% korrekte Antworten für den maximalen Bestand und 46% für den minimalen (gemittelt über verschiedene Bedingungen). In der Bedingung mit Liniendiagramm ergaben sich 52% und 41% korrekte Antworten. Bei Strohecker (2009) ergaben sich für den selben Flussgrößen-Verlauf 31% und 19%. Diese Zahlen aus der Literatur wurden über das Antwortformat Fragen zum Bestand ermittelt, während in der vorliegenden Studie die Versuchspersonen den Verlauf zeichneten. Dennoch sollte sich die Lösungsgüte in Abhängigkeit des Antwortformats zumindest nicht stark unterscheiden. Schwarz (2016) untersuchte BF-Aufgaben und ließ die Versuchspersonen Fragen beantworten und Lösungsgraphen zeichnen. Es ergaben sich hohe Korrelationen für diese beiden Antwortformate (r = .77 bis .91). In seinen Studien ergaben sich für denselben diskontinuierlichen Verlauf Lösungsraten mit einer Spannweite

von 30% bis 36% für *gezeichnete* Lösungsgraphen. Dies trat wiederum über verschiedene Bedingungen hinweg auf. Somit lagen die Lösungsraten für den diskontinuierlichen Verlauf bei der vorliegenden Studie im hohen Bereich. Grundlegend führten die symbolischen Bedingungen also zu guter bis sehr guter Leistung im BF-Bereich.

4.4.1.1 Vergleich Baseline-Bedingung und symbolische Bedingungen

Hypothese 3.1 und *Hypothese 3.*2 lassen sich mit der vorliegenden Studie bestätigen. Versuchspersonsn beider symbolischer Bedingungen schnitten besser ab als die der Baseline-Bedingung. Die Effektgrößen waren in einem sehr hohen Bereich für das BF-Forschungsgebiet.

4.4.1.2 Vergleich der beiden symbolischen Bedingungen

Zu *Fragestellung 3.1* lässt sich sagen: Versuchspersonen, die das Wasser-Szenario bearbeiteten, schnitten tendenziell besser ab als diejenigen mit Bus-Szenario. Beim Rechteck- und Dreieckverlauf schnitten die Versuchspersonen in den beiden Bedingungen ungefähr gleich gut ab, beziehungsweise es bestand ein leichter Unterschied in der Leistung beim Dreieckverlauf. Beim

diskontinuierlichen Verlauf war die BF-Leistung in der Bedingung mit Bus-Szenario vergleichbar mit der Leistung in der Baseline-Bedingung. Woran könnte das liegen? Eine mögliche Erklärung könnte sich aus der Anordnung in den Diagrammen ergeben. Im Wasser-Szenario zeigten Striche an den Wasserbehältern an, wie viel Wasser sich jeweils in dem Messbecher befand. Diese Striche waren über das Zehner-System angeordnet, indem Fünfer- und Zehner-Striche stärker markiert wurden. Hingegen erfolgte die Anordnung im Bus-Szenario in Dreier-Reihen für das Fluss- und das Lösungsdiagramm, da sich bei anderer Anordnung Platzproblemen ergeben hätten. Bei zweiterer Anordnung könnten unter Umständen Rechenfehler leichter auftreten und die BF-Leistung verschlechtern.

Insgesamt schien die symbolische Darstellung zu besserer BF-Leistung zu führen. Der Vergleich der beiden symbolischen Bedingungen zeigte keinen klaren Vorteil für eine der beiden Darstellungsformate. Ob die Beschreibungsdimension BF-Szenario/BF-Größe selber wirkt, müsste in weiteren Studien untersucht werden. Dabei sollte vermieden werden, dass die Flussdiagramme sich wie hier über die Anordnung in den Piktogrammen unterscheiden (Fünfer- vs. Dreiersystem). Im folgenden Abschnitt zur Fehleranalyse gehe ich genauer auf die verschiedenen Fehlerarten und ihr Vorkommen ein.

4.4.2 Fehleranalyse

Fragestellung 3.2 betraf die auftretenden Fehlerarten. Als Fehlerkategorien lagen vor: Korrelationsheuristik, Rechenfehler, mehr als zwei Fehler und sonstige(r) Fehler. Bei den Bedingungen in der symbolischen Darstellung waren die häufigsten Fehler Rechenfehler. Folglich könnten Versuchspersonen das Prinzip verstanden haben und dennoch keine komplett korrekten Lösungen erreichen. Die Kategorie „mehr als 2 Fehler“ kam – genauso wie Rechenfehler an sich – am häufigsten für die Bus-Bedingung vor. Sonstige Fehler und Korrelations-Heuristik traten beinahe ausschließlich in der Baseline-Bedingung auf. Gerade das Anwenden der Korrelations-Heuristik weist darauf hin, dass das Prinzip grundsätzlich nicht verstanden wurde. Es war oft nicht klar, ob sich die Versuchspersonen beim Verwenden der Korrelations-Heuristik am Zu- oder Nettofluss orientierten. Insgesamt unterschied sich das Fehlermuster für die symbolischen Bedingungen im Gegensatz zur Baseline-Bedingung.

Rechenfehler

Wenn Versuchspersonen nur einzelne Rechenfehler in ihrer Lösung aufwiesen, so spricht das dafür, dass sie wussten, wie sie vorgehen sollten und sich vereinzelt verrechnet hatten. Davon

ausgehend ergibt sich eine höhere Gesamtanzahl von denjenigen, die das Vorgehen verstanden hatten. Diese Gesamtzahl setzt sich aus der Anzahl an korrekten Lösungen addiert mit der Anzahl der Lösungen mit Rechenfehler zusammen. Nach dieser Berechnung liegen die Zahlen für die symbolischen Bedingungen knapp unter 30 beziehungsweise bei 30 Versuchspersonen, die das Vorgehen verstanden haben. In der Baseline-Bedingung hingegen ändert sich nichts bei den kontinuierlichen BF-Verläufen. Die Anzahl der Korrektlösenden lag bei unter 10 Versuchspersonen, und es existierten keine Lösungen mit Rechenfehler. Nur beim diskontinuierlichen Verlauf in der Baseline-Bedingung ergab sich eine Änderung, wenn Lösungen mit Rechenfehler einbezogen wurden: Bei diesem stieg die Gesamtzahl vom Anwenden der korrekten Lösungsprozedur von 9 auf 16 Fälle. Vermutlich verstanden die Versuchspersonen der Baseline-Bedingung das Vorgehen beim diskontinuierlichen Verlauf am ehesten. Zudem trat beim diskontinuierlichen Verlauf auch in der Baseline-Bedingung kein Mal eine Lösung orientiert an der Korrelations-Heuristik auf. Woran könnte das liegen?

Die Versuchspersonen schienen beim diskontinuierlichen Flussgrößen-Verlauf in der Baseline-Bedingung ebenfalls zeitpunktweise vorzugehen und die Flussgrößen zu verrechnen und dann zu akkumulieren, wodurch Rechenfehler möglicherweise

leichter auftreten. Bei den anderen – kontinuierlichen – Flussgrößen-Verläufen (als Liniendiagramm) könnten die Versuchspersonen vermehrt dazu neigen, Zeitpunkte zusammenzufassen, vielleicht sogar versuchen, alles auf einen Blick zu erfassen und daraus die Lösungen ermitteln. Somit lagen dort weniger Rechenfehler vor. Gleichzeitig vermehrte sich das Auftreten der Korrelations-Heuristik und sonstiger Fehler in der Baseline-Bedingung. Beim symbolischen Darstellungsformat rechneten die Versuchspersonen; jedoch konnten sie ebenfalls Zeiträume zusammenfassen. Beispielsweise war die Differenz beim Rechteckverlauf drei Mal in Folge +3 und dann drei Mal in Folge -3. Für den diskontinuierlichen Verlauf musste jeder Zeitpunkt neu berechnet werden. Somit erhöhte sich die Wahrscheinlichkeit für Rechenfehler.

Hier handelte sich um einen Spezialfall im Antwortformat beim diskontinuierlichen Verlauf: Bisher war das Antwortformat nur selten, einen Lösungsgraphen zu zeichnen. In Studien von Schwarz (2016) korrelierten die beiden Antwortformate hoch (vgl. Kapitel 4.4.1). Jedoch war nicht berichtet, ob sich beim Zeichnen des Verlaufes ähnliche Fehlermuster ergaben wie in der vorliegenden Studie. Weitere Studien könnten das Phänomen beim diskontinuierlichen Verlauf in der Baseline-Bedingung genauer untersuchen.

Insgesamt deutete die Fehleranalyse auf besseres BF-Verständnis beim symbolischen Darstellungsformat hin. Jedoch schien ebenfalls schrittweises Vorgehen zu wirken und das Finden der korrekten Lösung zu erleichtern.

4.4.3 Kontrollvariablen Geschlecht und Studienfach

Auch in der vorliegenden Studie beeinflussten Geschlecht und Studienfach die BF-Leistung. Männer und Versuchspersonen aus Fächern mit mehr naturwissenschaftlichen Inhalten in der Lehre schnitten besser ab; das zeigte sich in der Baseline-Bedingung am stärksten. Für die neue Diagrammart verlor sich die Wirkung von Geschlecht, während Studienfach noch eine leichte Wirkung zeigte. In Analysen von Interaktionen zwischen Geschlecht oder Studienfach und Bedingung ergab sich nur für Geschlecht eine signifikante Interaktion. In Einbeziehung der graphischen Analyse lässt sich dies so interpretieren, dass bei klassischem Darstellungsformat Männer bessere BF-Leistung erbrachten. Bei symbolischem schnitten die Frauen jedoch etwa gleich gut ab wie die Männer.

Bei der Variable Studienfach lag für die Auswertung eine unter Umständen willkürliche Dichotomisierung vor: Die Fächer wurden in „Naturwissenschaften nah und fern“ eingeordnet. An-

dere ForscherInnen könnten andere sinnvolle Einteilungen finden, um festzulegen, in welchem Ausmaß in einzelnen Fächern mathematische Inhalte gelehrt werden. Andererseits sollte die Einteilung nach Fakultäten zumindest eine annähernd günstige Aufteilung darstellen.

Zusammenfassend lässt sich sagen: Zusatzvariablen wie mathematische Vorbildung und Geschlecht wirkten auch in dieser Studie, wobei ihre Wirkung beim symbolischen Darstellungsformat aufgehoben oder abgeschwächt war. Dieses Darstellungsformat triggerte anscheinend BF-Verständnis für alle fast gleichermaßen. Korrekte BF-Lösungen zeigten sich weitaus seltener beim klassischen Liniendiagramm, speziell bei den „BF-schwächeren Gruppen“ wie Frauen oder Versuchspersonen aus naturwissenschaftlich fernen Studienfächern.

4.4.4 Transfer-Leistung

Betreffend *Fragestellung 3.3* zeigte sich: Beim symbolischen Darstellungsformat erfolgte keine Transfer-Leistung, sondern das symbolische Format an sich schien die Verbesserung der Leistung in den Durchgang1-Aufgaben zu triggern.

In der Baseline-Bedingung fand starker Transfer statt. Ein hoher Anteil der Personen, die die Durchgang1-Aufgaben korrekt lösten, lösten auch die Transfer-Aufgaben korrekt. Die Aufgaben

enthielten dieselbe Lösungsprozedur und dieselbe Aufgabenbeschreibung, da dasselbe Darstellungsformat vorlag. Nach der Kategorisierung verschiedener Arten von Transfer (Reed, 1999) handelte es somit um äquivalente Probleme, die nach Reeds Erkenntnissen im Allgemeinen gleichermaßen gut gelöst werden. In der vorliegenden Studie war die Transfer-Leistung bei die Aufgabe mit Dreieckverlauf maximal. Alle Versuchspersonen, die diese korrekt gelöst hatten, lösten auch beide Transfer-Aufgaben korrekt. Das ist in sich stimmig, da die Aufgabe mit Dreieckverlauf in der Baseline-Bedingung die tendenziell schwierigste war. Hohe Transfer-Leistung in der Baseline-Bedingung ist also naheliegend und trat tatsächlich auf. Eine Ausnahme stellte die diskontinuierliche Durchgang1-Aufgabe dar: Von ihren Korrektlösenden löste nur weniger als die Hälfte die kontinuierliche Transfer-Aufgabe korrekt (Überlegungen hierzu in Kapitel 4.4.6).

Nach den Ergebnissen zu den bedingten Wahrscheinlichkeiten trat in den symbolischen Bedingungen kein Transfer auf. Die Bus-Bedingung schien das Lösen der Transfer-Aufgaben regelrecht zu erschweren. Nur in der Wasser-Bedingung in Kombination aus diskontinuierlicher Transfer-Aufgabe und Rechteckverlauf könnte eine leichte Transfer-Leistung aufgetreten sein. Diese wäre jedoch im Vergleich mit der Baseline-Bedingung sehr gering. Nach Reed zeigen Personen bei isomorphen Problemen

nicht immer Transfer-Leistung. Die Expertise spielt ebenfalls eine Rolle: Ist sie geringer, so neigen Personen eher dazu, Probleme über ihre Beschreibung zu kategorisieren und somit übersehen sie, dass der Lösungsprozess derselbe ist. Möglicherweise erklärt das den fehlenden Transfer der vorliegenden Studie bei den symbolischen Bedingungen. Insgesamt ergab sich für das symbolische Darstellungsformat keine nennenswerte Transfer-Leistung, obwohl bei den Durchgang1-Aufgaben starke Verbesserungen aufgetreten waren.

4.4.5 Zusätzliche Analysen

4.4.5.1 Positionseffekte

Beim Dreieckverlauf traten Positionseffekte auf, die sich jedoch durch den Aufbau des Experimentes ausmitteln sollten. Die Versuchspersonen lösten das Problem mit Dreieckverlauf an dritter Stelle am besten. Könnte es sich um einen Übungseffekt handeln? Dagegen spricht, dass sich weder bei Rechteckverlauf noch bei diskontinuierlichem Verlauf ähnliches zeigt. Somit ist unklar, ob es sich um Zufall handelte oder ob für den Dreieckverlauf Übungseffekte auftraten.

Bei den Transfer-Aufgaben lagen keine Positionseffekte vor. Bisherige Studien mit entsprechenden Abläufen, die Übungsef-

fekte analysierten, sind mir nicht bekannt. In Studie 2 (vgl. Kapitel 3.3.6.2) trat kein Positionseffekt auf, während in der vorliegenden Studie bei den Durchgang1-Aufgaben sowie in Studie 1 (Kapitel 2.3.6) einer auftrat. Entweder entstanden sie aufgrund zufälliger Schwankungen oder es handelt sich um kleine Effekte, die somit selten signifikant werden. Weitere Studien könnten dies genauer untersuchen.

4.4.5.2 Aufgabenschwierigkeit

In der Baseline-Bedingung ergab sich folgende Rangreihe: Die Aufgabe mit Dreieckverlauf ist die schwierigste. Die diskontinuierliche Aufgabe schnitt eindeutig am besten ab, wenn die Versuchspersonen, denen Rechenfehler unterliefen, ebenfalls als Personen angesehen wurden, die das Prinzip verstanden hatten. In den symbolischen Darstellungsformaten schienen die meisten Probleme aufgrund von Rechenfehlern zu entstehen. Die Versuchspersonen lösten den diskontinuierlichen Flussgrößen-Verlauf seltener korrekt, weil sie hier die meisten Rechenfehler machten. Somit wird die diskontinuierliche Aufgabe zur schwierigsten, wenn die Lösungsgüte den Indikator für Schwierigkeitsgrad darstellt. Die Angaben zur Schwierigkeit der Aufgaben unterschieden sich nicht nach den verschiedenen Bedingungen, wobei eine Beurteilung aller Aufgaben insgesamt erfolgte. Weite-

re Studien könnten diesbezüglich detailliertere Fragen untersuchen.

4.4.6 Mögliche Wirkfaktoren

Grundlegend führten die symbolischen Darstellungsformate zu besserer BF-Leistung. Nach Ossimitz (2002) ergaben sich viele Fehler, weil Versuchspersonen die Flussgrößen als Bestandsgrößen ansehen. In der vorliegenden Studie sind die Flussgrößen wie einzelne Bestandsgrößen dargestellt, damit dieses Problem umgangen wird. Dieser Ansatz scheint zu helfen. Woran liegt dies nun?

Das schrittweise Vorgehen schien den Versuchspersonen zu helfen, die korrekten Lösung zu finden (vgl. Kapitel 4.4.2). Die Werte zu den einzelnen Zeitpunkten werden verrechnet, was für analytisches Vorgehen beim Lösen der Aufgaben spricht. Dies passt zu Hammonds Annahmen (1988): Sequentielle Darstellung führt eher zu analytischem Vorgehen. Zu Beginn der vorliegenden Arbeit bestand die Idee, dass durch das passende Darstellungsformat intuitives BF-Verständnis getriggert werden sollte. Zunächst scheint das also für die aktuelle Studie nicht erfolgt zu sein.

Jedoch stellte das schrittweise Vorgehen wohl nicht den einzigen Wirkfaktor dar: In der Baseline-Bedingung wird die Aufga-

be mit diskontinuierlichem Verlauf besser gelöst als die anderen beiden Aufgaben. Vermutlich wird aufgrund des Antwortformats eher schrittweises Vorgehen genutzt. Dennoch ist die Leistung schlechter als in den symbolischen Darstellungsformaten. Sie bleibt auch schlechter, wenn die Versuchspersonen mit Rechenfehlern als Personen angesehen werden, die das Prinzip verstanden haben: Es ergab sich eine Steigerung von 30% auf 53%, wenn Rechenfehler einbezogen werden. Sie stieg sogar auf 67%, wenn auch die Kategorie „mehr als 2 Rechenfehler“ hinzugenommen wurde. In beiden Fällen lag jedoch der Anteil am vermutlich korrekten Vorgehen in den symbolischen Bedingungen bei 87% bis 97% und war somit höher. Zusätzlich ergaben sich nur in der Baseline-Bedingung Fehler-Arten, die nicht systematisch erschienen und die darauf hindeuten, dass die Versuchspersonen das Prinzip nicht verstanden hatten. Derartige Fehler-Arten traten (fast) nicht in den symbolischen Bedingungen auf, ein weiterer Hinweis, dass das schrittweise Vorgehen nicht den einzigen Wirkfaktor darstellte. Weiterhin fanden Versuchspersonen in der klassischen Diagrammdarstellung, auch wenn sie rechneten, seltener zu den korrekten Lösungen (Booth Sweeney & Sterman, 2000; Kapitel 2). Denn in diesen Studien bestand die Möglichkeit schrittweise vorzugehen, und dennoch zeigten sich weitaus schlechtere Lösungsraten. Es muss also beim symbolischen Dar-

stellungsformat noch etwas wirken, wodurch die Versuchspersonen korrekt vorgingen und akkumulierten. Welche weiteren Unterschiede – außer der Nutzung von Piktogrammen – bestanden zwischen den Darstellungsformaten? Entsprechend wie in der vorhergehenden Studie (Kapitel 3) waren Zufluss und Abfluss getrennt, wobei beide Flussgrößen von unten nach oben hin zunahmen. Eventuell könnte diese Veränderung hilfreich sein. Meines Wissens nach wurde das bisher nicht untersucht und könnte in weiteren BF-Studien geprüft werden.

Ein Vorteil der symbolischen Darstellung könnte im Senken der kognitiven Beanspruchung liegen. Zwar untersuchten Cronin et al. (2009) diese Fragestellung und kamen zu dem Ergebnis, dass ein Vereinfachen der Aufgabe und somit gesenkte kognitive Beanspruchung nicht die BF-Leistung verbesserte. Jedoch nutzten sie durchgängig konventionalisierte Diagramme in dieser Fragestellung (wie Linien- oder Säulendiagramme). Eine Studie zu Diagrammdarstellung zeigte bessere Erinnerungsleistung bei Diagrammen mit Piktogrammen, wenn eine Belastung des Arbeitsgedächtnisses vorlag (Haroz, Kosara, & Franconeri, 2015).

Die symbolische Darstellung könnte zusätzlich auf andere Art wirken: Personen aktivieren über Piktogramme vorhandenes Wissen, beziehungsweise sie rufen Schemata leichter ab (Weidenmann, 1994, S.31). Weidenmann beschreibt diesbezüglich

verschiedene Funktionen von pädagogischen Bildern. Im vorliegenden Fall könnten die Versuchspersonen ein Schema zum Vorgehen beim Bearbeiten von BF-Aufgaben besitzen. Eventuell ist dies unvollständig oder nicht ganz korrekt, oder es kann noch differenziert werden. Ein inkorrektes Schema könnte beispielsweise das Anwenden der Korrelations-Heuristik sein. Das Ausdifferenzieren oder Berichtigen kann über ein Bild mit *Fokusfunktion* erfolgen. Der in vorliegender Studie wiederholt abgebildete Bus oder Wasserbehälter könnte diese Funktion übernehmen. Denn in der symbolischen Darstellung ist der Bestand zu jedem Zeitpunkt betont: In den Abbildungen existierte zwischen Zu- und Abfluss immer eine Abbildung vom Wasserbecken oder vom Bus. Eventuell hilft schon das, um einfacher zum korrekten Vorgehen zu kommen. Die vom Autor gegebenen Beispiele sind zwar weniger abstrakt: Zum Beispiel handelte es sich um eine Abbildung, in der Gefahrenquellen hervorgehoben werden. Jedoch könnte die Fokusfunktion zumindest zusätzlich für die BF-Aufgaben wirken.

Das symbolische Darstellungsformat schien das neu gefundene Vorgehen oder das korrekt aktivierte Schema zu triggern, da keine beziehungsweise kaum Transfer-Leistung stattfand.

Schlechte Transfer-Leistung in den symbolischen Bedingungen

Schon beim Betrachten der grundlegenden Anteile, das heißt für die korrekten Lösungen bei den Transfer-Aufgaben unabhängig von den Durchgang1-Aufgaben, zeigte sich keine besonders gute Leistung. In der Bus-Bedingung war die Leistung am schlechtesten mit circa einem Drittel der Versuchspersonen, das die jeweilige Aufgabe korrekt löste. In der Baseline-Bedingung war es jeweils etwa die Hälfte. In den Studien, denen die zwei Aufgaben entnommen waren, lagen für beide ungefähr zwei Drittel an korrekten Lösungen vor (Cronin et al., 2009; Ossimitz et al., 2001). In der Wasser-Bedingung war der Unterschied zwischen den beiden BF-Verläufen am größten: 40% lösten die kontinuierliche und um die 60% die diskontinuierliche Aufgabe korrekt. Es ergaben sich Unterschiede für die Lösungsgüte bei den Transfer-Aufgaben in den verschiedenen Bedingungen. Warum ist das so? Eine mögliche Erklärung ist der Zufall. Es kann sein, dass die Gruppen sich von vornherein unterschieden. Jedoch zeigten sich in der Lösungsgüte weder bei den Durchgang1- noch bei den Transfer-Aufgaben offensichtliche Hinweise hierfür. Zudem bestand in vorliegender Studie kein nennenswerter Unterschied zwischen den Gruppen, was die Verteilung von Geschlecht und Studienfach betraf. Somit ist auszuschließen, dass Unter-

schiede in der BF-Leistung zwischen den Gruppen aufgrund unterschiedlicher Verteilung dieser Personenvariablen auftreten. Ob es sich doch um Zufallsbefunde handelte oder eine Systematik dem zugrunde lag, müssten weitere Studien klären.

Die BF-Leistung in den Transfer-Aufgaben fiel also erstaunlich schlecht aus für Versuchspersonen der Bedingungen mit symbolischen Darstellungsformat. Woran könnte das liegen? Eventuell existieren bei Personen verschiedene Strategien zum Vorgehen. Die symbolischen Darstellungsformate könnten eine erfolgreiche Strategie triggern, die das schrittweise Vorgehen enthält. Möglicherweise gibt es zusätzlich Personen, die nicht schrittweise vorgehen müssen, da sie ein grundlegenden BF-Verständnis besitzen. Diese Personen verstehen, dass wenn der Zu- über Abfluss liegt, der Bestand zunimmt (und vice versa). Sie sind in der Lage vom Flussdiagramm ausgehend den Bestand korrekt zu bestimmen und sollten in der Baseline-Bedingung erfolgreich sein. Eventuell haben Personen mit derartigem Verständnis Probleme, wenn sie zunächst aufgrund des symbolischen Darstellungsformats zum schrittweisen Vorgehen gezwungen werden, und anschließend bei den Transfer-Aufgaben über das oben beschriebene „grundlegende BF-Verständnis“ den Bestand ermitteln müssen.

Weiterhin zeigte sich in der Baseline-Bedingung eine auffällig gute Leistung für die diskontinuierliche Durchgang1-Aufgabe. Es könnten einige Personen durch den Zwang zum schrittweisen Vorgehen eher zur korrekten Lösung gekommen sein, die ansonsten nicht über ein grundlegendes BF-Verständnis verfügen. Für diese Aufgabe schien die Transfer-Leistung niedriger in Kombination mit kontinuierlicher Transfer-Aufgabe. Das könnte einen weiteren Hinweis für ein Vorhandensein verschiedener Strategien darstellen. Das Nutzen verschiedener Strategien allgemein könnte wiederum die tendenziellen Positionseffekte bei den Transfer-Aufgaben erklären. Das sind jedoch nur erste Überlegungen, die mit Studien untersucht werden müssten, da die Datenlage diese Annahmen weder eindeutig unterstützen noch widerlegen kann.

4.4.7 Fazit

Idee der vorliegenden Arbeit war es, intuitives Vorgehen für BF-Aufgaben zu triggern, um zu besseren Lösungsraten zu kommen. Nach bisherigen Erkenntnissen sollte eine bildliche Darstellung von Inhalten zu intuitivem Vorgehen führen (Hammond, 1988; Hogarth, 2001). Da es sich bei Liniendiagrammen auch um eine bildliche Darstellung handelt, muss ein weiterer Vorteil in der symbolischen Darstellung liegen, der zum korrekten Vorgehen führt. Das wiederum passt zum Bereich der mathematischen

Intuition, wo in der Aufgabe Hinweisreize liegen, die zum korrekten Vorgehen führen (Kapitel 1.2.2). Eventuell erfolgte in der vorliegenden Studie eine Art intuitives Triggern des korrekten, analytischen, schrittweisen Vorgehens. Weitere Studien sollten genauer erkunden, inwiefern die oben vermuteten Wirkfaktoren für sich genommen hilfreich sind. Weiterhin wären Studien zu Positionseffekten oder zu einer Förderung von Transfer-Leistung spannend.

Das Abändern der Flussdiagramme und des Lösungsdiagramms führt zu erhöhtem Verständnis der BF-Aufgaben. Gleichzeitig ist es mit großem Aufwand verbunden, einerseits für die GestalterInnen der Diagramme, andererseits für die Versuchspersonen, die alle Zeitpunkte einzeln berechnen. Es ist also eher für Trainings oder im Unterricht geeignet, und weniger als Kommunikationsmittel wie beispielsweise in Zeitungen oder anderen Medien, die derartige Sachverhalte darstellen. Die letzte Studie dieser Arbeit untersucht erneut die symbolische Darstellung. Einerseits soll dem großen Zeitaufwand abgeholfen werden. Weiterhin soll die symbolische Darstellung weiterentwickelt werden.

5 Erweiterung der symbolischen Darstellung – Studie 4

Die vorangegangene Studie (Kapitel 4) zeigte eine starke Verbesserung der BF-Leistung. Vermutlich war die Darstellung von Flussgrößen als „Bestandsgrößen für einzelne Zeiteinheiten“ hilfreich. Dies wurde über symbolische Darstellung umgesetzt. Das führte zu einer starken Betonung der einzelnen Zeitpunkte, was vermutlich ebenfalls hilfreich war. Eine weitere Hilfe lag möglicherweise in der symbolischen Darstellung im Gegensatz zu konventionalisierten Diagrammen. Darum werden in der folgenden Studie die Untersuchungen mit symbolischem Darstellungsformat fortgeführt. Wiederum werden die Flussgrößen wie kleine Bestandseinheiten pro Zeiteinheit dargestellt.

5.1 Einleitung

Im vorhergehenden Diskussionsteil (Kapitel 4.4) wurden einige Probleme beim Bearbeiten der Aufgaben im symbolischen Darstellungsformat angeführt. Beispielsweise ist das Vorgehen für die Versuchspersonen sehr zeitaufwändig, wenn sie konkrete Werte für die Lösung angeben sollen. Außerdem ist ein möglicher Einwand, dass die verbesserte Leistung ausschließlich auf den

Zwang zu rechnen zurückzuführen ist. Die BF-Aufgaben wurden für die aktuelle Studie derart verändert, dass es nicht mehr notwendig war, den Bestand für jeden einzelnen Zeitpunkt zu ermitteln. Es sollte genügen, das Flussdiagramm zu betrachten und durch einen Gesamtüberblick der Flussverläufe auf die korrekte Lösung zu kommen. Das ist möglich, wenn die Fragen zum maximalen und minimalen Bestand gestellt werden (vgl. Sterman, 2002 bzw. Kapitel 1.1.3). Dort war dies anhand der Kaufhaus-Aufgabe mit diskontinuierlichem Verlauf beschrieben: Vor dem Schnittpunkt der Flussgrößen lag der Zufluss über dem Abfluss, danach war es umgekehrt. Die hinzugeflossene Gesamtmenge war größer als die abgeflossene Gesamtmenge, somit lag beim Schnittpunkt der maximale Bestand vor. Ähnlich einfach ist es, den Zeitpunkt für den minimalen Bestand zu ermitteln. Auch bei diesem Aufgabenverlauf können Versuchspersonen die Werte für die einzelnen Zeitpunkte berechnen. Jedoch forderte die Problemdarstellung hier nicht so stark dazu auf wie in der vorhergegangenen Studie. Über die zusätzliche gegebene Frage nach dem Vorgehen ist prüfbar, ob viele Versuchspersonen angeben, dass sie für alle Zeitpunkte den Bestand berechnen.

Zu Beginn der vorliegenden Arbeit hatte sich gezeigt, dass eine *schematische Abbildung* des BF-Szenarios hilfreich war (Kapitel 2.3.2). Dabei handelte es sich um ein realistisches Bild,

das das BF-Szenario darstellt. Die symbolischen Darstellungsformate der letzten Studie könnten als Abwandlung von logischen Bildern angesehen werden (Domagk, 2008, S.209, f.) zu logischen Bildern kombiniert mit realistischen Bildern. Ähnlich gestaltete Neurath Diagramme (Weidenmann, 1994). Neurath entwickelte sogenannte Isotype-Diagramme, mit denen er statistische Inhalte leicht verständlich kommunizieren wollte. Beispielsweise stellte er Anteile schon damals über Häufigkeiten dar (1936). Weidenmann (1994) gibt eine Zusammenfassung von Regeln für das Isotype-System:

a) es handelt sich um eine Darstellung über „ikonische Zeichen“. Das heißt, dass beispielsweise ein bestimmter Beruf durch ein entsprechendes Zeichen an oder in einem stilisierten Menschen gezeigt wird;
b) genutzte Zeichen repräsentieren eine bestimmte Anzahl der Objekte, die sie darstellen;
c) verschiedene Mengen werden konkret über die Menge von Objekten und nicht über deren Fläche abgebildet;
d) die Anordnung der Zeichen ergeben ein sogenanntes „visuelles Argument“. Das kann beispielsweise der Vergleich von Importen und Exporten verschiedener Länder sein.

Für die Gestaltung von Flussdiagrammen der vorliegenden Studie wurden Eigenschaften von Isotype-Diagrammen berücksichtigt.

5.1.1 Hypothesen

Eine Annahme ist, dass die symbolische Darstellung auch zu besserer Leistung führt, wenn das Rechnen nicht so stark forciert wird.

Hypothese 4.1: Die BF-Leistung ist besser in den Bedingungen mit den symbolischen Darstellungsformaten.

Zusätzlich sollten einzelne Aspekte aus Neuraths Bildersprache Isotype genutzt werden. Im Sinne der Nutzung von „ikonischen Zeichen“ werden die Unterschiede der beiden Flussgrößen stärker betont, indem unterschiedliche Zeichen für Zufluss- oder Abflussgrößen stehen.

Hypothese 4.2: Bei den symbolischen Darstellungsformaten erhöht sich die Leistung stärker durch eine stärkere Differenzierung von Zu-und Abfluss mittels ikonischer Zeichen statt gleichartiger Piktogramm-Zeichen.

Im Folgenden unterscheide ich zwischen „piktogrammartig“ und ikonisch als Begriffe für das verwendete Darstellungsformat.

Piktogrammartige Zeichen zeigen für Zufluss und Abfluss gleichartige gestaltete Piktogramme, die für eine bestimmte Anzahl der Flussgrößen stehen. Ikonische Zeichen sind stärker an die Bedeutung angepasst: Beispielsweise steht das Bild eines Babys für eine bestimmte Anzahl an Geburten, die wiederum eine Zuflussgröße darstellen. Natürlich könnten die Begriffe anders verwendet werden – letztendlich handelt es sich bei den Babys auch um Piktogramme. Für die vorliegende Studie drücken diese Begriffen einen Unterschied in der Darstellung aus, und als Oberbegriff für beide Versionen wird „symbolisch“ verwendet.

Weiterhin wurde geprüft, ob das Betonen der Differenzen von Zu- und Abfluss im symbolischen Darstellungsformat hilfreich war. Im Bereich Multimedia Learning hatte sich gezeigt, dass die Integration unterschiedlicher Informationen in einer Abbildung positive Auswirkungen haben kann. Beispielsweise kann es zu hoher kognitiver Beanspruchung kommen, wenn Informationen voneinander getrennt dargeboten werden. Der Rezipient muss seine Aufmerksamkeit zwischen verschiedenen Quellen zunächst aufteilen und diese dann mental zusammenfügen. Das wird als „Split-Attention“ bezeichnet (Van Merriënboer & Ayres, 2005). Um den Split-Attention-Effekt zu umgehen, sollten zusammengehörige Informationsquellen gemeinsam präsentiert werden (Sweller, 1999). Das könnte bei den BF-Szenarien umge-

setzt werden, indem die Flussgrößen und der Nettofluss in einer Abbildung vorliegen. Das kann ähnlich der Bedingung „Nettofluss gegeben“ (Kapitel 3) gesehen werden. Zwar hatte sich gezeigt, dass zusätzliches Vorgeben des Nettoflusses das BF-Verständnis nicht erhöhte (Kapitel 3.3.3, Schwarz, 2016). Jedoch lag in den bisherigen Studien mit konventionalisierten Diagrammen eine Extra-Abbildung für den Nettofluss als Zwischenschritt vor. Im symbolischen Darstellungsformat ist es möglich, alle drei Flussgrößen in einer Abbildung zu integrieren und dabei den Nettofluss zu betonen. Möglicherweise hilft die zusätzliche Information des Nettoflusses, wenn die kognitive Beanspruchung gering gehalten wird, und alle Informationen in ein symbolisches Flussdiagramm integriert sind.

Fragestellung 4.1: Hilft es hier, den Nettofluss zu betonen?

Es ist wiederum von Bedeutung, welche Fehler sich für die BF-Lösungen ergeben.

Fragestellung 4.2: Welche Fehlerarten zeigen sich?

Außerdem wurden Geschlecht und mathematisches Vorwissen als Kontrollvariablen erhoben. Bisher schnitten in den meisten Fällen Männer besser ab als Frauen, wenn die BF-Probleme im klassischen Format dargeboten werden. Zusätzlich ist mathe-

matische Vorbildung ein wichtiger Faktor (ausführlich beschrieben Kapitel 3.1.4). In einigen Studien stellte die Mathematiknote einen guten Prädiktor für BF-Verständnis dar (Lungwitz et al., 2018; Schwarz, 2016). In der vorliegenden Studie sind diese Variablen zur Kontrolle in der Regressionsanalyse erhoben.

Da für das Lösen von BF-Aufgaben mathematische Fähigkeiten hilfreich sind, können Ergebnisse aus der Forschung zum Mathematikverständnis helfen, einen Geschlechtereffekt besser einzuordnen. Grundlegend zeigte sich beim Erforschen von Geschlechtsunterschieden immer wieder, dass Männer besser in Mathematik abschneiden als Frauen (Kimura, 1999). Es wird häufig angenommen, dass ein allein aufgrund des biologischen Geschlechts bestehender Unterschied in der Mathematikleistung unwahrscheinlich ist (Beermann, 1992; Budde, 2009). Spencer, Steele und Quinn (1999) beschreiben ein generelles Muster: Frauen schneiden gleich gut ab wie Männer, außer bei Inhalten in höheren und schwierigen Kursen an High School und College oder wenn es sich um einen sehr anspruchsvollen Test handelt – anspruchsvoll bezogen auf den aktuellen, mathematischen Kenntnisstand. In diesen Fällen schneiden Männer besser ab. Weiterhin beschreiben die AutorInnen, dass es dazu unterschiedliche Begründungen gibt. Einerseits wird (noch immer) das biologische Geschlecht als Ursache angesehen. Andererseits könnten diese

Unterschiede die Sozialisation beziehungsweise die erworbene Geschlechterrolle widerspiegeln. Nach einer Studie von Selkow (1985) führte das Geschlecht zu keinem signifikanten Unterschied in mathematischer Leistung im Gegensatz zur Geschlechterrolle: Personen mit hoher maskuliner Geschlechterrolle schnitten besser ab als die Gruppen mit anderen Geschlechterrollen.

Ein weiterer Aspekt kommt aus der Sozialpsychologie. Es handelt sich um das Konzept „Stereotype threat“ oder Bedrohung durch Stereotype. Das bedeutet, einer Gruppe wird eine negatives Vorurteil zugeschrieben, und das ist den Gruppenmitgliedern bekannt. In Situationen, für die das Stereotyp gilt, besteht die Bedrohung, dieses Stereotyp zu verstärken, wenn sich die betreffende Person entsprechend verhält (Spencer et al., 1999). Das wiederum kann in der jeweiligen Situation zu verstärktem Druck oder Stress führen, was zu schlechterer Leistung führen kann. Im Bereich Mathematik zeigte sich dies häufig. Beispielsweise schnitten asiatische Frauen besser ab, wenn das Vorurteil "Asiatische Personen sind besser in Mathematik" salient war. Sie schnitten schlechter ab bei erzeugter Salienz von "Frauen sind schlechter in Mathematik" (Shih, Pittinsky & Ambady, 1999). Nach Spencer et al. (1999) reicht es aus, dass sich eine Person in einer Situation befindet, in der sie ein negatives Stereotyp durch ihr Verhalten unterstützen könnte, um diese Bedrohung bei ihr aus-

zulösen. Schon die Teilnahme an einem Mathematik-Test könnte also als derartige Situation wahrgenommen werden. Weiterhin argumentierten sie, dieses Phänomen galt nur für schwierige Probleme.

BF-Aufgaben können zu mathematischen Aufgaben gezählt werden, insbesondere wenn Flussdiagramme gegeben sind (Schwarz, 2016). Zudem handelt es sich um schwierige Aufgaben. Somit können bei BF-Aufgaben diese Bedingungen als gegeben angesehen werden. Das heißt, auch wenn keine Bedrohung durch Stereotype aktiv herbeigeführt wird, könnte sie wirken. Schmader (2001) stellte fest, dass Bedrohung durch Stereotype für Frauen in Mathematik-Tests eher besteht, wenn sie sich stark mit ihrer Zugehörigkeit zur „weiblichen Gruppe“ identifizierten. Aufgrund dieser Befunde sollte in der aktuellen Studie die Geschlechterrolle erhoben werden, um zu testen, ob ein Zusammenhang mit BF-Leistung bestand, beziehungsweise ob die BF-Leistung aufgrund der Geschlechterrolle vorhersagbar war. Beim „Bem Sex-Role-Inventory“ (BSRI, Bem 1974) wird die Geschlechterrolle nicht als ein Kontinuum mit zwei Polen angesehen, sondern Maskulinität und Femininität stellen unabhängige Dimensionen dar. In diesem Inventar sind Aussagen genutzt, denen die Personen auf einer 7-Punkte-Skala zu unterschiedlichem Ausmaß zustimmen. Diese Aussagen sind für Männer bezie-

hungsweise Frauen mehr oder weniger sozial erwünscht. Aufgrund der Werte auf diesen Dimensionen und der Differenz der beiden Werte können verschiedene Aspekte der Geschlechterrolle beschrieben werden.

Fragestellung 4.3: Besteht ein Zusammenhang zwischen BF-Leistung und den Ausprägungen der Geschlechterrollen?

5.1.2 Materialien: die verwendeten Verläufe

Als Flussgrößen-Verläufe wurden der Dreieckverlauf und zwei diskontinuierliche Verläufe verwendet. Der Dreieckverlauf scheint im Allgemeinen schwierig lösbar zu sein und führte immer zu schlechteren Lösungsraten als der Rechteckverlauf (Armenia, Onori, Bertini, 2004; Booth Sweeney und Sterman, 2000; Ossimitz, 2002; Schwarz, 2016; Kapitel 2.3.1; Kapitel 4.3.2 jeweils die Baseline-Bedingung; eine einzelne Ausnahme enthielt Kapmeier et al., 2014). Das symbolische Darstellungsformat fiel diesbezüglich aus der Reihe: In der vorigen Studie wurden Aufgaben mit Dreieckverlauf in symbolischer Darstellung besser gelöst als die Aufgabe mit diskontinuierlichem Flussgrößen-Verlauf. Vermutlich lag das jedoch an einer erhöhten Wahrscheinlichkeit sich beim diskontinuierlichen Verlauf zu ver-

rechnen. Der Dreieckverlauf wird also als guter Indikator für stabiles BF-Verständnis angesehen.

In der vorhergegangenen Studie fiel die Aufgabe mit diskontinuierlichem Verlauf aus der Reihe. In den symbolischen Bedingungen war die Leistung für diese Aufgabe am schlechtesten, während sie in der Baseline-Bedingung zum höchsten Anteil an korrekten Lösungen führte. Für die aktuelle Studie sollten daher diskontinuierliche Verläufe mit unterschiedlichem Schwierigkeitsgrad verglichen werden. So war die Lösungsrate bei diesen Verläufen untersuchbar, wenn vermutlich weniger Versuchspersonen die Lösung berechneten. Es gab in dieser Studie zwei verschiedene Versionen von diskontinuierlichem Verlauf: Eine sehr einfache Version und eine etwas schwierigere. Der Schwierigkeitsgrad sollte darüber variiert werden, ob sich die Verläufe der Flussgrößen schneiden oder nicht. Beispielsweise beschrieben Cronin et al. (2009) einen diskontinuierlichen Verlauf ohne Schnittpunkt als besonders einfach, dieses Problem wurde (annähernd identisch) für diese Studie übernommen.

5.2 Methoden

5.2.1 Teilnehmende

Für die Poweranalyse wurde insgesamt eine konservative Schätzung für den Effekt genutzt, gesetzt den Fall, dass die Ergebnisse aus Kapitel 4 den Effekt überschätzten. Die Bus-Bedingung schnitt schlechter ab als die Wasser-Bedingung. Somit wurden Baseline und Bus-Bedingung herangezogen für die Schätzung einer Effektstärke aus der vorhergehenden Studie. Es ergab sich $g = 1.48$ (g ist kleiner/gleich d), und das Konfidenzintervall für g (approximativ) reichte von .90 bis 2.05. Für den Effekt wurde die Untergrenze des Konfidenzintervalls verwendet. In eine Poweranalyse für den Vergleich zweier Gruppen (bezogen auf symbolische vs. konventionalisierte Darstellung) gingen die Werte ein: $d = .9$; $\alpha = .05$; eine Power von 90%. Das ergab eine Gruppengröße von 22 Versuchspersonen. Abgebrochen wurde die Studie, die online durchgeführt wurde, als 149 Studierende der TU Chemnitz die Materialien komplett ausgefüllt hatten. Es lagen pro Bedingung 25 bis 35 Versuchspersonen vor.

Die Versuchspersonen waren im Mittel 22 Jahre alt ($SD = 3.8$, Spannweite 18 bis 45). Sie erhielten Versuchspersonen-Stunden für ihre Teilnahme. 97 Personen studierten Psychologie (65%), und 48 Personen studierten Sensorik und kognitive

Psychologie (SeKo) (32%). Eine Person gab an, Erzieherin zu sein. 32 Männer (21%) und 113 (76%) Frauen nahmen an der Studie teil. Vier Personen gaben ihr Geschlecht nicht an. Zwei weitere Versuchspersonen gaben ihr Studienfach und ihr Alter nicht an.

5.2.2 Materialien

Diese Studie wurde als Online-Studie konzipiert. Der Aufgabenkontext war in allen Fällen die Bevölkerung einer mittelalterlichen Stadt. Diese war in einzelnen Zeiträumen von der Außenwelt abgeschnitten, und die Bevölkerung änderte sich ausschließlich aufgrund von Geburten und Todesfällen. Jede Versuchsperson bearbeitete drei Aufgaben im selben Darstellungsformat, es handelte sich um ein between-subjects-Design. Für das Darstellungsformat ergaben sich fünf Bedingungen. Zunächst folgen die drei verschiedenen Aufgabenmuster, wie sie in der Baseline-Bedingung vorliegen. Im weiteren Verlauf erläutere ich die einzelnen Bedingungen.

Genutzte BF-Verläufe

Es handelte sich um drei verschiedene Flussgrößenverläufe: Zwei waren diskontinuierlich, einmal ohne und einmal mit Schnittpunkt. Gezeigt werden die Verläufe zunächst im klassi-

schen Darstellungsformat. Der diskontinuierliche Verlauf ohne Schnittpunkt (Abbildung 5.1) wurde fast identisch aus dem vierten Experiment von Cronin et al. (2009) entnommen. Die korrekten Lösungen sind für Bestand maximal das Jahr 615 und für minimal 611. In den Original-Darstellungen der Online-Studie war der Zufluss als durchgängige Linie in grün abgebildet; das gilt für diese und die folgenden Abbildungen.

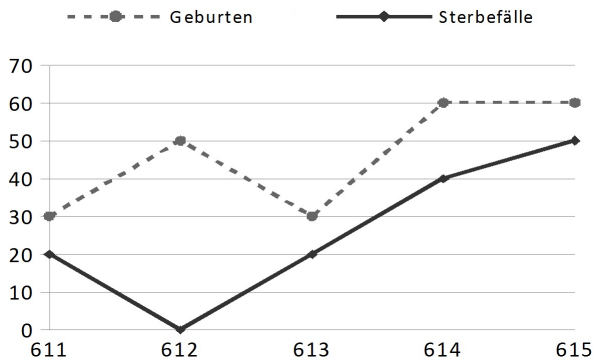


Abbildung 5.1 Diskontinuierlicher Verlauf ohne Schnittpunkt: d-oS (übernommen und leicht abgewandelt von Cronin et al., 2009).

Der andere diskontinuierliche Verlauf (Abbildung 5.2) ist orientiert an dem Verlauf aus dem ersten Experiment von Cronin

et al. (2009). In diesem Fall ist der Verlauf um einige Zeitpunkte gekürzt (6 statt 12), da es sonst Probleme mit der Gestaltung der Diagramme in den symbolischen Bedingungen gegeben hätte. Die korrekten Lösungen sind für den maximalen Bestand 672 und 673; für den minimalen Bestand 675.

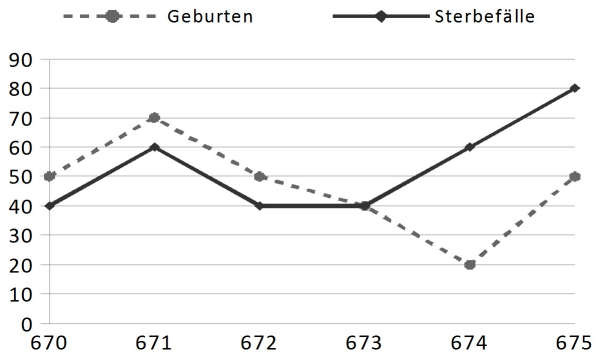


Abbildung 5.2 Diskontinuierlicher Verlauf mit Schnittpunkt: d-mitS (orientiert an Cronin et al., 2009).

Der dritte Verlauf war der Dreieckverlauf. Auch dieser enthielt im Vergleich zu bisherigen Studien weniger Zeitpunkte, und zwar aus denselben Gründen wie oben (Abbildung 5.3). Korrekte Lösungen sind für den maximalen Bestand die Jahre 701, 702,

705, 706, 709 und für den minimalen Bestand die Jahre 703, 704, 707, 708.

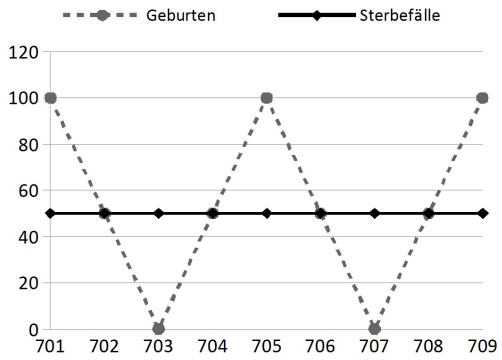


Abbildung 5.3 Dreieckverlauf (übernommen und leicht abgewandelt von Booth Sweeney & Sterman, 2000).

5.2.3 Modifikationen des Darstellungsformats

Insgesamt ergaben sich fünf verschiedene Arten von Darstellungsformat, die im Folgenden beschrieben sind.

5.2.3.1 Baseline-Bedingung

In der Baseline-Bedingung waren die Probleme im klassischen Liniendiagramm abgetragen. Die oben gezeigten Abbildungen (Abbildung 5.1; Abbildung 5.2; Abbildung 5.3) zeigen somit die Aufgaben für die Baseline-Bedingung. Es folgen die vier verschiedenen symbolischen Bedingungen: die Kombinationen aus ikonischer versus piktogrammartiger Darstellung und mit betontem versus ohne betonten Nettofluss.

5.2.3.2 Bedingung „Piktogramm-Menschen, ohne Differenz“ (Pikt-oDiff)

Hierbei handelt es sich in der Darstellung um eine Replikation der vorhergehenden Studie, Zu- und Abfluss werden gleichermaßen über Piktogramm-Menschen dargestellt (Abbildung 5.4). In der Online-Studie war der hellere Grau-Ton im Bereich des Zuflusses hellgrün und der dunklere Grau-Ton beim Abfluss war grau. Dies gilt entsprechend für alle folgenden Bedingungen.

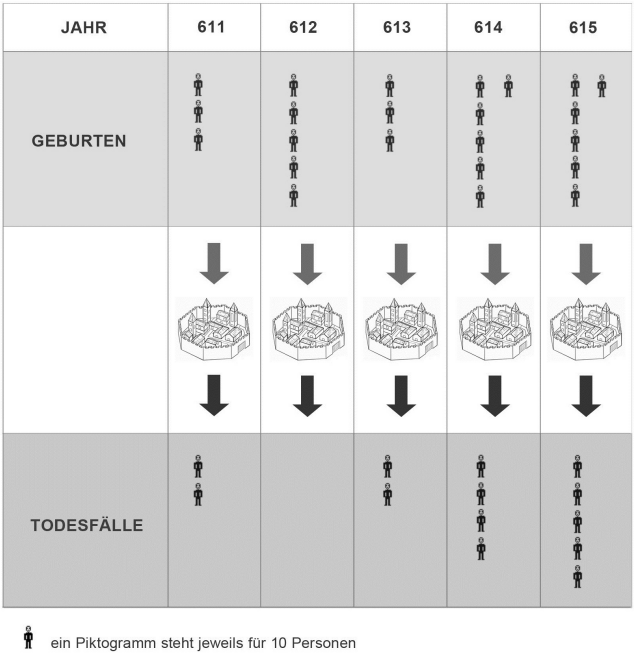


Abbildung 5.4 Darstellung des Flussdiagramms für die Bedingung Pikt-oDiff

5.2.3.3 Bedingung „ikonische Symbole ohne Differenz“ (ikon-oDiff)

In dieser Bedingung wird der Unterschied zwischen Zu- und Abfluss stärker betont. Babys als Piktogramme im Sinne ikonischer Symbole stehen für Geburten, ikonische Symbole von Särgen für Todesfälle. Wiederum sind Zu- und Abfluss räumlich getrennt voneinander abgetragen (Abbildung 5.5).

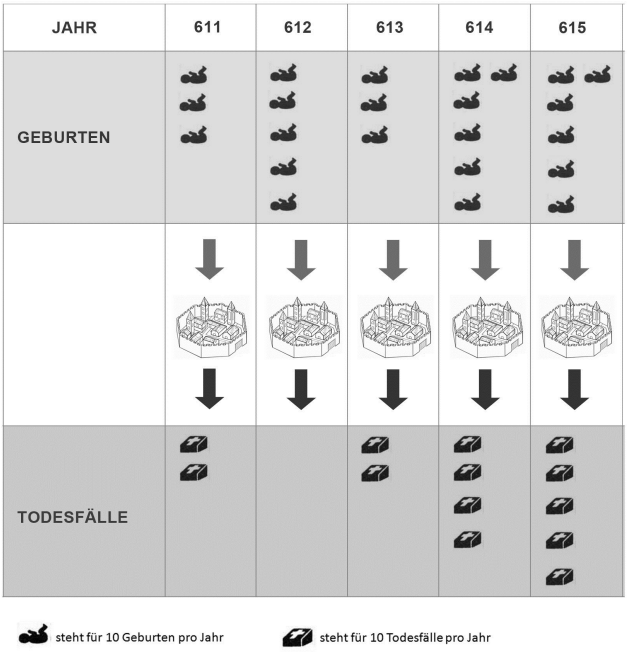


Abbildung 5.5 Darstellung des Flussdiagramm für die Bedingung ikon-oDiff (Gerd Arntz "Piktogramme“ (Babies und Särge) / © VG Bild-Kunst, Bonn 2018).

5.2.3.4 Bedingung „Piktogramm-Menschen, mit Differenz“ (Pikt-mitDiff)

Die Bedingung mit Piktogramm-Menschen wurde etwas abgeändert. Nun werden Zu- und Abfluss so nebeneinander gestellt, dass sie quasi schon verrechnet sind. In der Mitte der Abbildung befinden sich die sich ausgleichenden Werte von Zu- und Abfluss (Ausgleichswerte). Somit ist der Nettofluss explizit abgebildet, wenn auch nicht erklärt oder als solcher benannt (Abbildung 5.6). Im Jahr 670 gibt es beispielsweise 5 Piktogramm-Menschen für Geburten und 4 Piktogramm-Menschen für Todesfälle. Das sind 4 Ausgleichswerte und ergibt einen Nettofluss von einem Piktogramm-Menschen und somit 10 Personen.

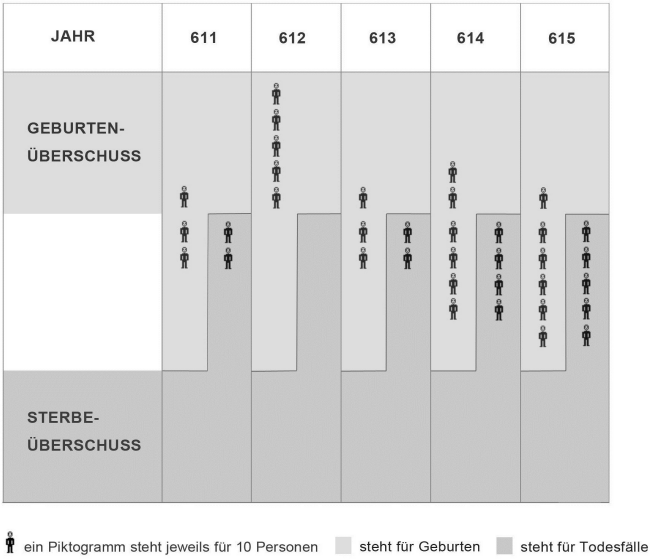


Abbildung 5.6 Darstellung des Flussdiagramms für die Bedingung Pikt-mitDiff

5.2.3.5 Bedingung „ikonische Symbole mit Differenz“ (ikon-mitDiff)

Die Bedingung mit den ikonischen Symbolen wird ebenso erweitert wie die Bedingung mit Piktogramm-Menschen (Abbildung 5.7). Wiederum werden Zu- und Abfluss derart nebeneinan-

der gestellt, dass gleich ablesbar ist, ob es mehr Geburten oder Sterbefälle gibt.

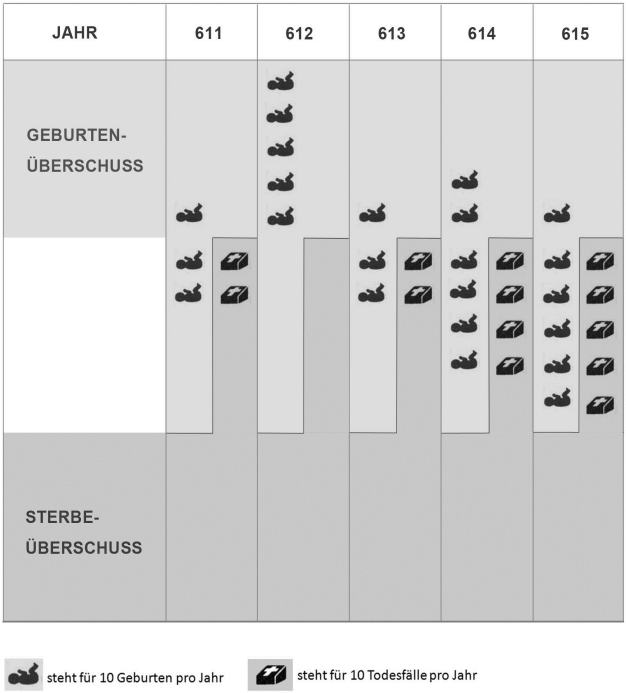


Abbildung 5.7 Darstellung des Flussdiagramms für die Bedingung ikon-mitDiff (Gerd Arntz "Piktogramme“ (Babies und Särge) / © VG Bild-Kunst, Bonn 2018).

Jede Versuchsperson bearbeitete die drei oben beschriebenen Verläufe in einer Bedingung, also immer im selben Darstellungsformat. Zu jeder einzelnen Aufgabe waren Fragen gestellt, um das BF-Verständnis zu prüfen.

5.2.4 Abhängige Variablen und Zusatzvariablen

Bei den abhängigen Variablen handelte es sich um zwei Fragen zum Bestand, das heißt zur Bevölkerungszahl. Zum einen wurde erfragt, wann die meisten Personen in der Stadt lebten. Dabei handelte es sich um die Variable Bestand-maximal. Zum anderen gab es die Frage, wann die wenigsten Personen in der Stadt lebten: Das war die Variable Bestand-minimal. Zusätzliche Variablen waren Fragen nach den Flussgrößen: Wann die meisten und die wenigsten Personen geboren wurden beziehungsweise wann die meisten und wann die wenigsten Personen verstarben. Dabei handelte es sich um maximalen und minimalen Zufluss beziehungsweise um maximalen und minimalen Abfluss. Hierbei wurde geprüft, ob die Versuchspersonen die Flussgrößen im symbolischen Darstellungsformat genauso gut ablesen konnten wie in der Baseline-Bedingung. Alle diese Fragen wurden mit korrekt/nicht korrekt kodiert. Die Fragen zum Bestand auf der einen Seite und die zu den Flussgrößen auf der anderen Seite bildeten jeweils einen sogenannten Fragenkomplex. Diese beiden Fragenkomple-

xe waren getrennt angeordnet. Zu jedem dieser Fragenkomplexe kam die Frage, wie schwierig die Versuchspersonen die Fragen fanden (Schwierigkeit). Zusätzlich erhielten die Versuchspersonen Items aus dem BSRI-R („Bem Sex-Role-Inventory revisited“, Troche & Rammsayer, 2011), um ihre Geschlechterrolle zu ermitteln. Troche und Rammsayser hatten die ursprüngliche Version des deutschen BSRI (Schneider-Düker & Kohler, 1988) überarbeitetet. In der aktuellen Studie wurden die Items der Skalen für Maskulinität und Femininität gemischt abgefragt. Die dritte Skala für soziale Erwünschtheit wurde nicht erhoben. Die Reihenfolge der Items wurde zufällig variiert, unabhängig von ihrer Zugehörigkeit zur Skala für Maskulinität oder für Femininität. Weiterhin beantworteten die Versuchspersonen Fragen zu demographischen Variablen wie Geschlecht, Alter, Studienfach, Mathematiknote. Letztere bezog sich auf die Abiturnote in Mathematik (als Punkte). Abgefragt wurde dies, indem Versuchspersonen in einer Liste (mit Punkten von 0 bis 15) die Note auswählen konnten. Das sollte gewährleisten, dass keine Versuchsperson aus Versehen die Note aus dem Notensystem 1 bis 6 angab. Außerdem gaben sie eine allgemeine Einschätzung ihrer Fähigkeiten zum mathematischen Denken/Problemlösen. Sie beantworteten, ob sie bereits an einer BF-Studie teilgenommen hatten, und sie gaben ihre Vertrautheit mit den bearbeiteten

Diagrammen an. Weiterhin beantworteten die Versuchspersonen den Fragenkomplex nach Motivation und Anspannung, wie anstrengend, frustrierend, langweilig die Aufgaben waren, wie verständlich die Graphiken waren und ob die Versuchspersonen Schwierigkeiten hatten, die Linien zu erkennen – also als schwarz und grün zu unterscheiden. Die Versuchspersonen beschrieben, wie sie beim Lösen der Aufgaben vorgegangen waren und woran sie erkannt hatten, dass die Menge der Personen zu einem bestimmten Zeitpunkt zu- oder abnahm (Frage nach „Kriterium“). Weitere Anmerkungen konnten sie ebenfalls machen.

5.2.5 Ablauf

Es existierten 5 Bedingungen. Die Studie erfolgte online in LimeSurvey (Version 1.92). Ein Link für die Online-Studie wurde an Studierende der TU Chemnitz über einen Verteiler geschickt. Nachdem eine Versuchsperson den Link anklickte, erschien ein Willkommenstext, mit der Bitte die Aufgaben ohne Hilfestellung und ohne Ablenkung zu bearbeiten. Jede Versuchsperson bearbeitete drei Probleme im selben Darbietungsformat. Die Darstellung war also between variiert. Es gab drei verschiedene Reihenfolgen für die Aufgaben, ermittelt über ein lateinisches Quadrat. Somit ergaben sich 15 verschiedene Subgruppen in LimeSurvey (fünf Bedingungen, mit jeweils drei Reihenfol-

gen). Die Versuchspersonen wurden zufällig – blockrandomisiert – den Bedingungen zugeordnet. Dennoch enthielten die Subgruppen verschieden viele Versuchspersonen. Das lag daran, dass einige Versuchspersonen die Umfrage aufgerufen hatten, somit einer Subgruppe zugeordnet wurden und dann die Umfrage nicht oder nicht komplett ausfüllten. Die verschiedenen Subgruppen wurden je nach Bedingung zusammengefasst. Daraufhin umfassten die Bedingungen 25 bis 35 Versuchspersonen. In Abschnitt 5.3.5.4 gehe ich darauf genauer ein. Zunächst bearbeiteten die Versuchspersonen die Fragen zu den Flussgrößen, dann die Frage nach der Schwierigkeit. Daraufhin folgte der nächste Fragenkomplex mit den Fragen nach Bestand und diesbezüglich nach der Schwierigkeit. Die Fragen zu Flussgrößen und Bestand waren jeweils in zufälliger Reihenfolge dargeboten. Anschließend folgten die weiteren obengenannten Fragen. Als letztes konnten die Versuchspersonen eine Kodierung eingeben, über die sie Versuchspersonen-Stunden erhielten. Diese Kodierung wurde getrennt von den anderen Variablen abgespeichert. Die Auswertungen erfolgten wiederum in den meisten Fällen über Regressionsanalysen und EDA. Die falschen Lösungen wurden ebenfalls analysiert.

5.3 Ergebnisse

Nach einem Überblick folgt die Darstellung der Ergebnisse für die verschiedenen Bedingungen. Anschließend folgt die Fehleranalyse. Im weiteren Verlauf werden Ergebnisse der weiteren Variablen wie Geschlecht, Mathematiknote und BSRI-R berichtet.

5.3.1 Überblick

Wiederum wird die Gesamtsumme für einen Großteil der Auswertungen genutzt. Sie setzt sich zusammen aus den einzelnen Variablen für Maximum und Minimum des Bestandes, die mit 0 für falsch und mit 1 für korrekt[9] kodiert waren. Die beiden Variablen existierten für die drei verschiedenen Flussgrößen-Verläufe, also kann die Gesamtsumme 0 bis 6 Punkte annehmen. Das arithmetische Mittel für die BF-Leistung betrug 2.58 (SD = 2.48). Dieser Mittelwert entsprach einer mittleren prozentualen Leistung von 43%. Komplett korrekt gelöst hatten 30 Versuchspersonen (20%) alle Aufgaben zum Bestand, und 57 Versuchspersonen (38%) erhielten für die Gesamtsumme 0 Punkte.

9 Teilweise waren die Jahreszahlen nicht korrekt angegeben, wirkten jedoch wie Tippfehler. Wenn ein Tippfehler als sehr eindeutig erschien (z.B. „701 705 509“) wurde es als korrekt kategorisiert (in diesem Fall wie: „701 705 709“).

Eine Versuchsperson zeigte einerseits bei den Fragen nach den *Flussgrößen* und beim BSRI-R Ausreißer-Werte. Bei den Fragen nach den Flussgrößen beantwortete die Person keine einzige Frage korrekt. Beim BSRI-R zeigte sie insgesamt sehr geringe Werte: 1.47 für Maskulinität und 1.53 für Femininität. Für den Rest der Stichprobe liegen die Mittelwerte bei 4.26 für Maskulinität und bei 4.97 für Femininität. Auch für die Variable Alter stellte die Versuchsperson mit 45 Jahren einen Ausreißer dar. Beim Rest der Stichprobe lag der Median bei 21. Es gab jedoch zwei weitere Ausreißer mit 35 und 31 Jahren. Es war unklar, ob die zuvor genannte Versuchsperson tatsächlich ein Ausreißer betreffend der Fragen nach den Flussgrößen und betreffend des BSRI-R war, der die angegebenen Werte als korrekt empfand oder ob fehlende Compliance zu diesen Ausprägungen führte. Auch die offenen Fragen geben hierfür keine klaren Hinweise. Die einzige beantwortete offene Frage war die nach dem Kriterium, woran erkennbar war, ob die Bevölkerung zu- oder abnahm: Dort schrieb die Person „Augenschein“. Aufgrund dieser Unklarheit wurden die Regressionsanalysen ohne diese Ausreißer-Versuchsperson durchgeführt. In den Fehleranalysen ist die Ausreißer-Versuchsperson wieder mit aufgenommen, denn es ist unwahrscheinlich, dass möglicherweise fehlende Compliance zu verzerrten Fehlermustern führt.

5.3.2 Einfluss der Aufgaben-Darstellung

Es folgen die Ergebnisse zu *Hypothese 4.1*, nach der die BF-Leistung in den symbolischen Bedingungen besser sein sollte und zu *Fragestellung 4.1*, ob das Betonen des Nettoflusses die Leistung verbesserte. Der mittlere prozentuale Anteil ergab 30% für die Baseline-Bedingung, 48% für Pikt-oDiff (Piktogramm-ohne Differenz), 41% für Pikt-mitDiff (Piktogramm-mit-Differenz), 53% für ikon-oDiff (ikonisch-ohne-Differenz) und 44% für ikon-mitDiff (ikonisch-mit-Differenz). Es folgen Angaben über den Anteil der KorrektlöserInnen für alle drei Aufgaben, also Versuchspersonen die in der Gesamtsumme 6 Punkte erhielten. Dieser lag bei 7% für Baseline, für ikon-oDiff bei 32%, für ikon-mit-Diff bei 16%, für Pikt-oDiff bei 28% und für Pikt-mitDiff bei 20%. Die Bedingung ikon-oDiff schnitt durchweg am besten ab, während Versuchspersonen der Baseline-Bedingung durchgängig die schlechteste BF-Leistung zeigten. Bei den anderen Bedingungen bestand teilweise ein Zusammenhang mit Flussgrößen-Verlauf.

Abbildung 5.8 zeigt bessere Leistung in den symbolischen Bedingungen im Vergleich zur Baseline-Bedingung, wobei der Unterschied nicht so stark wie in der vorhergegangenen Studie war. Nur die Bedingung Pikt-mitDiff fiel aus der Reihe und wies

denselben Median auf wie die Baseline-Bedingung (Abbildung 5.8). Das zusätzliche Betonen des Nettoflusses führte zu keiner Verbesserung im Gegensatz zu den anderen symbolischen Bedingungen.

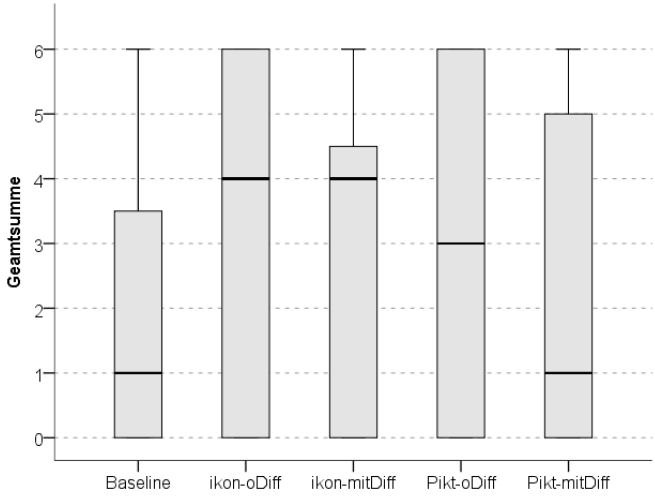


Abbildung 5.8 Boxplots für die BF-Leistung (aufsummierte Anzahl von korrekt beantworteten Fragen nach dem Bestand) je nach Bedingung
Anmerkung: ikon-oDiff: ikonisch-ohne-Differenz; ikon-mitDiff: ikonisch-mit-Differenz; Pikt-oDiff: Piktogramm-ohne-Differenz; Pikt-mitDiff Piktogramm-mit-Differenz

Zunächst wurde eine Regressionsanalyse für Prädiktoren durchgeführt. Für die 5 Bedingungen ergaben sich 3 jeweils dummy-kodierte Variablen: Eine Variable für Baseline versus symbolische Darstellungsformate, eine Variable für piktogrammartig versus ikonisch und eine Variable für mit versus ohne Differenzen. Die weiteren Variablen waren Geschlecht, Studienfach, Mathematiknote, Mittelwert der Maskulinitäts-Skala, Mittelwert der Femininitäts-Skala, Interaktion der beiden Skalen (MW-M x MW-F) und vorherige Teilnahme. Das vollständige Modell war signifikant und erklärte etwa 20% der Gesamtvarianz (Tabelle 5.1). Dort sind auch die Werte für die einzelnen Prädiktoren berichtet. Für das vollständige Modell mit der Ausreißer-Versuchsperson ergab sich $R^2 = .22$; $F(10,121) = 3.390$; $p < .00$. Im weiteren Verlauf sind die Werte der Regressionsanalyse ohne die Ausreißer-Versuchsperson berichtet (vgl. Kapitel 5.3.1).

Für den Vergleich von Baseline und symbolischen Bedingungen ergab $\beta = .20$ ($p = .02$): Die Versuchspersonen der Baseline-Bedingung schnitten schlechter ab. Für piktogrammartiges versus ikonisches Darstellungsformat ergab sich kein Unterschied je nach Format ($\beta = .04$, mit $p = .33$). Beim Vergleich der Bedingungen mit versus ohne Differenzen ergab sich $\beta = -.08$ ($p = .19$), also ebenfalls kein Unterschied.

Tabelle 5.1

Regressionsanalyse für das vollständige Modell mit Gesamtsumme als abhängige Variable. Prädiktoren waren die Bedingungen, die Kontrollvariablen, diesmal ergänzt um Mathematiknote, und die Skalen des BSRI-R bzw. die Interaktion der beiden Skalen.

Prädiktor	β	*t*	*p*	*pr*
Baseline	-.20	-2.097	.02	-.19
ikonisch	.04	.444	.33	.04
Differenz	-.08	-.871	.19	-.08
Geschlecht	-.01	-.122	.45	-.01
Studienfach	.10	1.194	.12	.11
Mathematik-Note	.26	3.119	<.00	.27
Femininitäts-Skala	-.27	-2.833	.01	-.25
Maskulinitäts-Skala	-.04	-.462	.65	-.04
Interaktion BSRI-R	-.05	-.568	.57	-.05
Teilnahme	.12	1.394	.08	.13

Anmerkung: Die Analyse ergab $R^2 = .22$; mit $F(10,120) = 3.424$; $p < .00$ (ohne die Werte der Ausreißer-Versuchsperson). Dummy-kodierte Variablen sind folgendermaßen kodiert:

Baseline 1, symbolische Bedingungen 0; ikonische Zeichen 1, alle anderen Bedingungen 0; Differenz: mit 1, ohne 0; Geschlecht: weiblich 0, männlich 1; Studienfach: SeKo 0, Psychologie 1; Teilnahme: bisher noch nicht teilgenommen 0, vorherige Teilnahme 1. Alle *p*-werte wurden halbiert, mit Ausnahme der Werte für die Interaktion und für die BSRI-R-Skalen.

Beim vollständigen Modell fehlten 8 Versuchspersonen, die entweder ihr Geschlecht oder das Studienfach nicht angegeben hatten, und es entfiel die Versuchsperson mit Berufsangabe Erzieherin. Vier dieser Personen entstammten der Bedingung Pikt-mit-Diff, zwei der Baseline und jeweils eine der Bedingung ikon-mit-Diff und ikon-oDiff. Von diesen Versuchspersonen hatten 6 als Gesamtsumme das Ergebnis 0. Um zu prüfen, inwiefern das Fehlen dieser Versuchspersonen die BF-Leistung beeinflusste, wurde in einem Modell ohne Geschlecht und Studienfach geprüft, wie die Leistung ausfiel. Es ergab sich $R^2 = .19$ und $F(8,130) = 3.854$ mit $p < .00$. Die gesamte Regressionstabelle (Anhang H) zeigt, dass für die meisten Werte keine starken Veränderungen auftraten. Für eine einfachere Vergleichbarkeit zwischen den Studien werden im weiteren Verlauf die Werte aus dem vollständigen Modell berichtet. Es folgen die Auswertungen zu den verschiedenen Fehlern.

5.3.3 Fehleranalyse

Zunächst zu den Ergebnisse zu den Fehlerarten (*Fragestellung 4.2*) für die einzelnen Gruppen: Abbildung 5.9 zeigt die Fehler für den Flussgrößen-Verlauf „diskontinuierlich ohne Schnittpunkt“ (d-oS) für die Frage nach maximalem Bestand. Tatsächlich korrekt war der Zeitpunkt 615. Die meisten fehlerhaf-

ten Lösungen gingen auf die Korrelations-Heuristik orientiert am Nettofluss zurück. Die Werte schwankten ungefähr von 40% bis 50%. Falsche Lösungen der Kategorie „sonstige Fehler“ traten selten auf. In der Baseline-Bedingung waren es etwa 15% und in der Bedingung ikon-mitDiff sogar weniger als 10%.

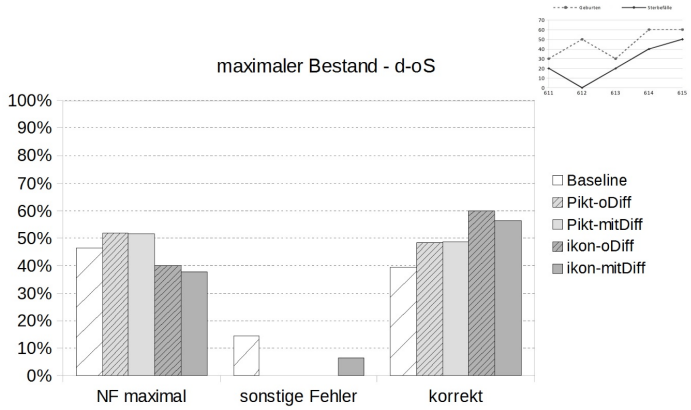


Abbildung 5.9 Der Anteil an korrekten Lösungen und Fehlerarten je nach Bedingung für die Frage nach dem maximalen Bestand für den Flussgrößen-Verlauf d-oS (d-oS ist der diskontinuierliche Verlauf ohne Schnittpunkt, vgl. Abbildung in rechter oberer Ecke).
Anmerkung: NF bedeutet Nettofluss; Bedeutung der Abkürzungen für die Bedingungen in der Legende siehe Abbildung 5.8.

Die tatsächlich korrekte Lösung für den minimalen Bestand ist das Jahr 611. Bei dieser Frage trat der Fehler „Korrelations-Heuristik orientiert am Nettofluss“ ebenfalls am häufigsten auf (Abbildung 5.10). Es war der Zeitpunkt angegeben, wann der Nettofluss am geringsten ist. Es folgte der Fehler „AF maximal“: In diesem Fall gaben Personen als Zeitpunkt des minimalen Bestandes den Zeitpunkt an, zu dem der Abfluss maximal ist.

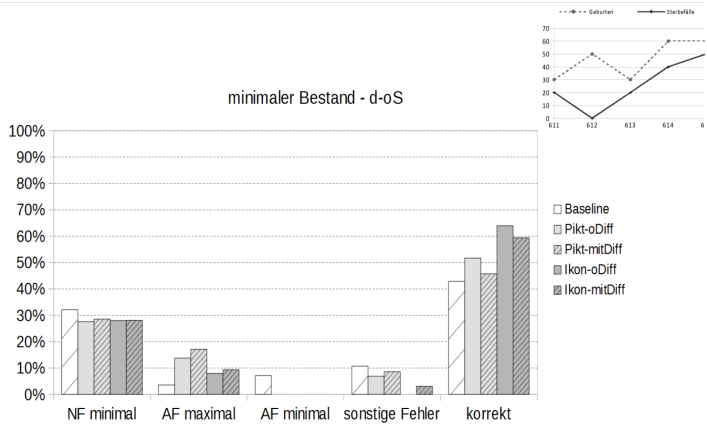


Abbildung 5.10 Der Anteil an korrekten Lösungen und Fehlerarten je nach Bedingung für den Flussgrößen-Verlauf d-oS für die Frage nach minimalem Bestand.
Anmerkung: NF bedeutet Nettofluss, AF: Abfluss. Bedeutung der Abkürzungen für die Bedingungen siehe Abbildung 5.8.

Nun zu den entsprechenden Abbildungen für den Flussgrößen-Verlauf „diskontinuierlich mit Schnittpunkt“ (d-mitS, Abbildung 5.11 und 5.12): Die tatsächlich korrekte Lösung ist 672 und 673 für den maximalen Bestand; falsche Lösungen verteilten sich die auf viele verschiedene Fehler. Am häufigsten trat die Korrelations-Heuristik orientiert an Nettofluss oder Zufluss auf. Weiterhin gaben die Versuchspersonen oft nur einen der beiden korrek-

ten Zeitpunkte an, wobei in einem Fall der Nettofluss 0 war, während im zweiten Fall der Zeitpunkt mit einem Nettofluss von 0 fehlte („NF=0“ bzw. „ohne NF=0“ in Abbildung 5.11). Der Zeitpunkt mit maximalem Zufluss war oft als Orientierung in der Bedingung Pikt-mitDiff genutzt. „NF positiv/ZF minimal“ bedeutet: Die Zeitpunkte 670 und 672 wurden als maximaler Bestand angesehen. Dabei handelte es sich um zwei der Zeitpunkte, an denen der Nettofluss maximal ist. Das dritte Jahr mit maximalem Nettofluss und gleichzeitig höherem Zufluss war nicht enthalten. In der Kategorie „sonstige Fehler“ sind unter anderem drei Lösungen enthalten, bei denen sich die Personen am Abfluss orientierten.

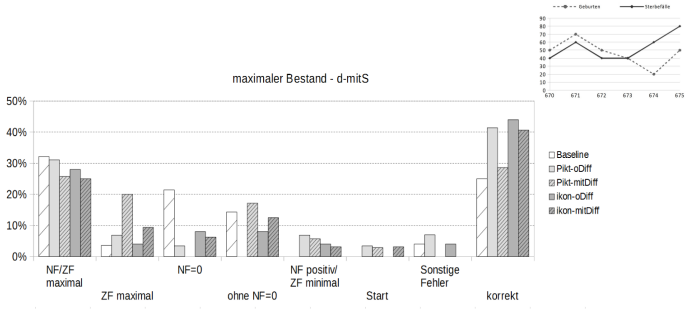


Abbildung 5.11 Der Anteil an korrekten Lösungen und Fehlerarten je nach Bedingung für den Flussgrößen-Verlauf d-mitS für die Frage nach maximalem Bestand (diesmal ist bei der Y-Achse 50% der maximale Wert).
Anmerkung: d-mitS ist der diskontinuierliche Verlauf mit Schnittpunkt (vgl. Abbildung in rechter oberer Ecke); NF: Nettofluss; ZF: Zufluss; weitere Erläuterungen im Text. Bedeutung der Abkürzungen für die Bedingungen siehe Abbildung 5.8.

Für den minimalen Bestand war beim Verlauf d-mitS das Jahr 675 die korrekte Lösung. Es zeigte sich nur ein systematischer Fehler: Korrelations-Heuristik orientiert am minimalen Nettofluss. (Abbildung 5.12).

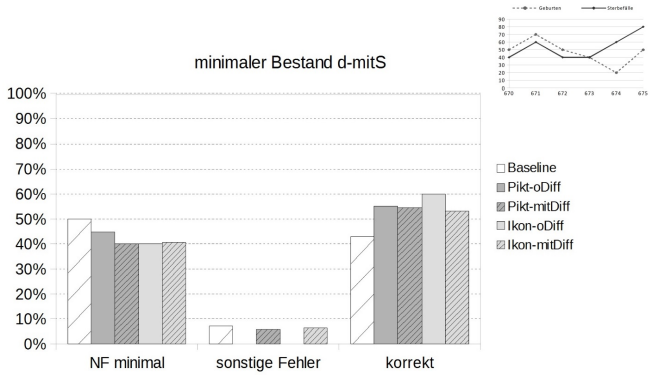


Abbildung 5.12 Der Anteil an korrekten Lösungen und Fehlerarten je nach Bedingung für den Flussgrößen-Verlauf d-mitS für die Frage nach minimalen Bestand.
Anmerkung: Bedeutung der Abkürzungen für die Bedingungen siehe Abbildung 5.8.

Im weiteren Verlauf sind die entsprechenden Abbildungen für den Dreieckverlauf gezeigt (Abbildung 5.13 und 5.14). Für den maximalen Bestand waren die korrekten Lösungen 701, 702, 705, 706 und 709. Bei maximalem Bestand zeigte sich wiederum als häufigster Fehler die Korrelations-Heuristik orientiert an Netto-/Zufluss maximal (Abbildung 5.13). Keiner der weiteren Zeitpunkte, die als Orientierung für die Korrelations-Heuristik ge-

nutzt wurden, kamen häufig vor. In der Kategorie „sonstige Fehler“ gab es noch zwei fehlerhafte Lösungen, bei denen vier der fünf korrekten Werte angegeben waren.

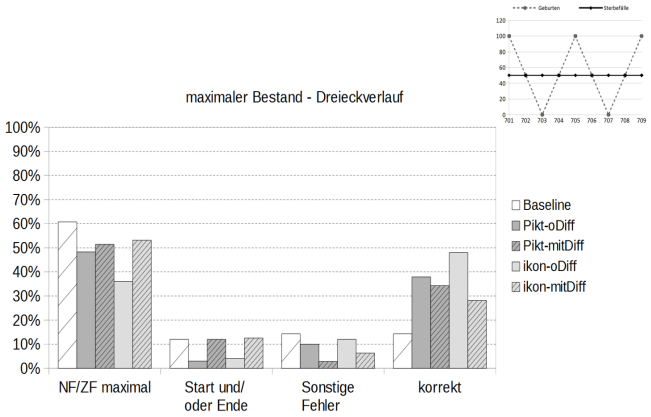


Abbildung 5.13 Der Anteil an korrekten Lösungen und Fehlerarten je nach Bedingung für den Dreieckverlauf (vgl. Abbildung in rechter oberer Ecke) für die Frage nach maximalem Bestand.
Anmerkung: NF: Nettofluss; ZF: Zufluss; weitere Erläuterungen im Text. Bedeutung der Abkürzungen für die Bedingungen siehe Abbildung 5.8

Für den minimalen Bestand entsprachen die Jahre 703, 704, 707, 708 der korrekten Lösung. Unter den falschen Lösungen war beim minimalen Bestand Netto-/Zufluss der häufigste Orientierungspunkt für die Korrelations-Heuristik (Abbildung 5.14). Die Kategorie „NF minimal ein Zeitpunkt“ bedeutet: Die Versuchsperson gab nur einen der Zeitpunkte an, zu denen der Nettofluss minimal war, obwohl es davon zwei gab. Zum Beispiel wurde nur 703 angegeben, obwohl 707 der Nettofluss ebenfalls minimal war.

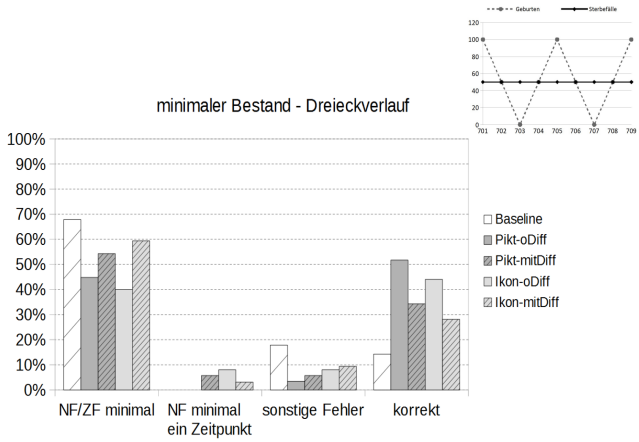


Abbildung 5.14 Der Anteil an korrekten Lösungen und Fehlerarten je nach Bedingung für den Dreieckverlauf (vgl. Abbildung in rechter oberer Ecke) für die Frage nach minimalem Bestand.
Anmerkung: NF: Nettofluss; ZF: Zufluss; weitere Erläuterungen im Text. Bedeutung der Abkürzungen für die Bedingungen siehe Abbildung 5.8

Es zeigte sich, dass die Korrelations-Heuristik orientiert am Nettofluss die meisten Fehler ausmachte.

5.3.4 Analysen zum Einfluss von Kontrollvariablen

Nicht alle Versuchspersonen haben diese Variablen komplett angegeben, weshalb einige Versuchspersonen für die Analysen entfielen.

5.3.4.1 Geschlecht

Männer und Frauen waren in dieser Stichprobe ungleichmäßig auf die einzelnen Bedingungen verteilt (Anhang G). Grundlegend ergab sich kein Effekt für Geschlecht: $\beta = -.01$ mit $p = .45$. Die folgende Abbildung (5.15) zeigt die Variable Gesamtsumme als Indikatoren für BF-Leistung für Männer und Frauen je nach Bedingung, um die Median-Werte der Gesamtsumme für Männer und Frauen zu vergleichen. Es zeigte sich kein Geschlechtereffekt: Entweder ergab sich für den Median beider Gruppen derselbe Wert (Baseline-Bedingung, ikon-mitDiff) oder die Werte lagen nah beieinander (ikon-oDiff und Pikt-mitDiff). Nur in der Bedingung Pikt-oDiff unterschieden sie sich stark, wobei die Männer besser abschnitten. Jedoch handelte es sich hierbei um eine der Bedingungen, an der nur drei Männer teilnahmen. Weiterhin schnitten die Frauen den Medianen nach besser in den ikonischen Bedingungen ab als in den restlichen Bedingungen. Bei den Män-

nern ergab sich kein so klares Bild: Am besten war die Leistung in der Bedingung Pikt-oDiff, gefolgt von den ikonischen Bedingungen. In der Baseline-Bedingung war die Leistung am schlechtesten.

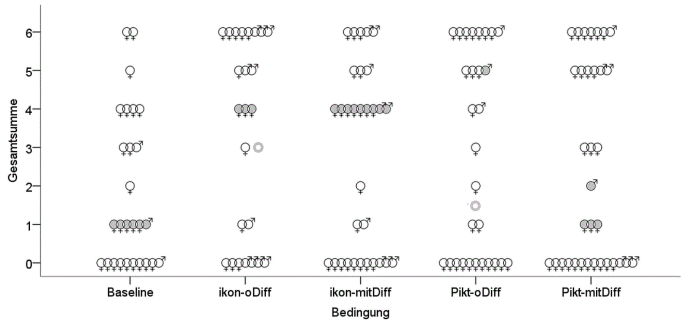


Abbildung 5.15: Streudiagramm für Gesamtsumme unterteilt nach Bedingung und Geschlecht, grau markierte Zeichen für Personen liegen auf dem Median. In zwei Fällen ist für den Median ein grauer Kreis eingezeichnet, da dort keine Person mit diesem Wert vorlag: einer in ikon-oDiff für die Männer und einer in Pikt-oDiff für die Frauen.
Bedeutung der Abkürzungen für die Bedingungen siehe Abbildung 5.8.

5.3.4.2 Studienfach

In der vorliegenden Stichprobe waren nur Studierende von Psychologie und SeKo enthalten, bis auf eine Erzieherin. Es ergab sich kein signifikanter Effekt für Studienfach: $\beta = .10$ mit $p = .12$. Da in dieser Studie ebenfalls Mathematiknote als Indikator für mathematisches Vorwissen erhoben worden war, fanden keine weiteren Analysen für Studienfach statt. Im Folgenden werden die Ergebnisse zur Variable Mathematiknote detailliert beschrieben.

5.3.4.3 Mathematiknote

Zunächst wurde geprüft, wie die Mathematiknoten in den verschiedenen Bedingungen ausfielen. Nach erster Analyse über Boxplot lagen hierfür keine schwerwiegenden Unterschiede vor. Dies wurde mit einer ANOVA überprüft, bei der sich kein signifikanter Unterschied für Mathematiknote ergab ($F(4,135) = .53$ und $p = .72$). In der Regressionsanalyse (vollständiges Modell) ergab sich $\beta = .26$ ($p < .00$). Die Mathematiknote stellte einen guten Prädiktor für die BF-Leistung dar.

5.3.4.4 Geschlechterrolle

Es folgen die Ergebnisse zum Zusammenhang zwischen BF-Leistung und der Ausprägung der Geschlechterrollen (*Fragestellung 4.3*). Hierfür beinhaltete die Regressionsanalyse als Prädiktoren die Werte von Maskulinitäts-, Femininitäts-Skala und ihre Interaktion. Für die Maskulinitäts-Skala ergab sich $\beta = -.04$ ($p = .65$), für die Femininitäts-Skala $\beta = -.27$ ($p = .01$) und für die Interaktion der beiden Skalen $\beta = -.05$ ($p = .57$) (alle p-Werte für BSRI-R zweiseitig). Nur die Variable Femininität hatte Vorhersagekraft für die BF-Leistung: Höhere Werte der Femininitäts-Skala hingen mit schlechterer Leistung zusammen.

5.3.4.5 Vorherige Teilnahme

Eine weitere Analyse betraf die vorherige Teilnahme: 8 Personen gaben an, an einer vorhergegangenen BF-Studie teilgenommen zu haben. In ihrer Leistung schnitten sie tendenziell besser ab. Die Hälfte von ihnen löste alle Aufgaben komplett korrekt, während es bei der anderen Gruppe 18% waren. Der Mittelwert lag bei 3.63 mit $SD = 3.02$ (schonmal teilgenommen) und 2.52 mit $SD = 2.44$ (noch nicht teilgenommen). Der Unterschied war jedoch nicht signifikant (siehe Tabelle 5.1).

5.3.5 Zusätzliche Analysen

5.3.5.1 Positionseffekte

Positionseffekte wurden keine gefunden. Eine Varianzanalyse ergab für die einzelnen Aufgaben: bei d-oS: $F(2,146) = .874$ ($p = 42$); bei d-mitS: $F(2,146) = 1.630$ $p = 20$); beim Dreieckverlauf: $F(2,146) = .997$ ($p = 37$).

5.3.5.2 Aufgabenschwierigkeit

Es folgen die Anteile der einzelnen korrekt beantworteten Fragen, unterteilt nach Flussgrößen-Verläufen und den Bedingungen (Abbildung 5.16). Insgesamt war wiederum die Aufgabe mit Dreieckverlauf am schwierigsten zu lösen. Eine Ausnahme bildete die Frage nach dem minimalen Bestand bei Pikt-oDiff; diese wurde vergleichsweise gut gelöst. Die Fragen zu den diskontinuierlichen Verläufen schienen innerhalb der verschiedenen Bedingungen tendenziell gleich gut gelöst zu werden. Eine Ausnahme stellte die Frage nach maximalem Bestand beim Verlauf mit Schnittpunkt dar; diese wurde durchgängig am schlechtesten gelöst.

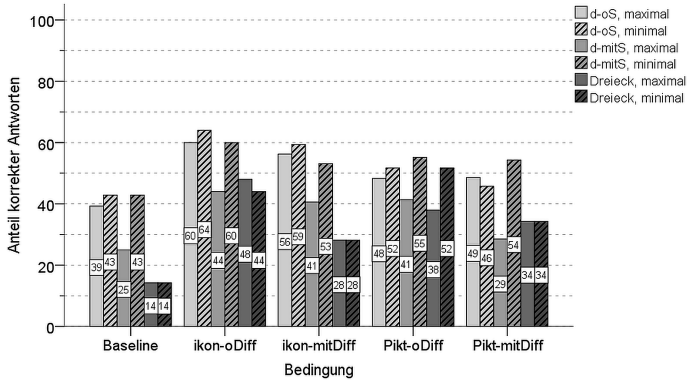


Abbildung 5.16: Anteile der korrekt beantworteten Fragen zu den verschiedenen Flussgrößen-Verläufen, unterteilt für die verschiedenen Bedingungen.
Anmerkung: Bedeutung der Abkürzungen für die Bedingungen siehe Abbildung 5.8.

5.3.5.3 Fragenkomplex zu den Flussgrößen

Bei den Fragen nach den Flussgrößen wurden keine Besonderheiten gefunden, abgesehen vom außergewöhnlichen Abschneiden der Ausreißer-Versuchsperson (Abschnitt 5.3.1). Für die einzelnen Bedingungen ergaben sich folgende Werte für die mittlere prozentuale Leistung: 93% für die Baseline-Bedingung,

96% für Pikt-oDiff, 82% für Pikt-mitDiff, 95% für ikon-oDiff und 90% für ikon-mitDiff. In allen Bedingungen lag der Median für die Gesamtsumme der Flussgrößen bei 12, also bei der maximal möglichen Punktezahl. Die einzige Ausnahme stellte die Bedingung Pikt-mitDiff dar, mit einem Median von 11. Neben dem eingangs beschriebenen Ausreißer, der hier 0 Punkte aufwies, stellten 6 weitere Personen der Baseline-Bedingung und der ikonischen Bedingungen Ausreißer dar. Sie hatten nur 6 bis 9 Punkte erreicht.

Beim diskontinuierlichen Verlauf ohne Schnittpunkt wählten Versuchspersonen häufig nur einen der beiden korrekten Werte für den maximalen Zufluss. Dieser Fehler trat in mehreren Bedingungen auf. Ein anderes Beispiel ist die Frage nach dem minimalen *Abfluss* bei dieser Aufgabe. Als Antwort wählten 8 Versuchspersonen Tage mit minimalem *Zufluss*. Sie alle hatten eine symbolische Bedingung durchlaufen. Ähnliches zeigte sich für die Fehler bei den anderen Flussgrößen-Verläufen. Beispielsweise beim Dreieckverlauf in der Bedingung Pikt-mitDiff: Bei der Frage nach maximalem *Abfluss* gaben 7 Versuchspersonen Tage mit minimalem *Zufluss* an; ein Fehler, der auch bei 3 Versuchspersonen aus den ikonischen Bedingungen vorlag. Ein weiterer systematischer Fehler zeigte sich in der Bedingung Pikt-mitDiff beim diskontinuierlichen Verlauf ohne Schnittpunkt: 9 Versuchs-

personen gaben für den maximalen *Zufluss* den Zeitpunkt des maximalen *Nettoflusses* an. Viele Fehler schienen auf einer Verwechslung der Flussgrößen zu basieren, wobei die meisten Fehler in der Bedingung Pikt-mitDiff auftraten.

5.3.5.4 Abbruchquote

In den unterschiedlichen Bedingungen brachen verschieden viele Versuchspersonen ab. Aufgrund der Blockrandomisierung sollten in allen Bedingungen gleich viele Versuchspersonen vorliegen, somit zeigen die Anzahl der Teilnehmenden in den einzelnen Bedingungen, wo die meisten AbbrecherInnen vorlagen. Es ergab sich folgende Verteilung für die fünf verschiedenen Bedingungen: In der Baseline waren es 28, in ikon-oDiff 25, in ikon-mitDiff 32, in Pikt-oDiff 29 und in Pikt-mitDiff 35 Versuchspersonen, die teilnahmen. Die niedrigste Anzahl an AbbrecherInnen ergab sich mit 5 Personen in der Bedingung Pikt-mitDiff. Die höchste Anzahl an AbbrecherInnen lag in der Bedingung ikon-oDiff vor, gefolgt von Baseline-Bedingung und Pikt-oDiff. Zusatzvariablen wurden geprüft, ob sie unterschiedlicher Ausprägung für die verschiedenen Bedingungen waren, wie die Verständlichkeit der Diagramme oder Motivation seitens der Versuchspersonen, wobei diese Daten nur für die Personen vorlagen, die an der Studie teilgenommen hatten und nicht für die Ab-

brecherInnen. Die Zusatzvariablen wurden dennoch geprüft: Aus ihnen ergab sich zumindest kein erster Hinweis, dass sich die Bedingungen bezüglich dieser Variablen unterschieden.

5.4 Diskussion

5.4.1 Einfluss der Aufgaben-Darstellung

Entsprechend *Hypothese 4.1* unterstützte das symbolische Darstellungsformat das BF-Verständnis, und die Versuchspersonen der Baseline-Bedingung schnitten eindeutig schlechter ab. Insgesamt jedoch verbesserte sich die Leistung in den symbolischen Bedingungen weniger stark als in der vorhergegangenen Studie. Eventuell ist eine starke Aufforderung für schrittweises Vorgehen erforderlich, um zur korrekten Lösung zu kommen. Das spricht dafür, dass die in Kapitel 4.4.6 vermuteten Wirkfaktoren „symbolische Darstellung“ und „schrittweises Vorgehen“ additiv wirkten.

5.4.2 Unterschied piktogrammartige und ikonische Darstellung

Nach *Hypothese 4.2* sollten die Bedingungen mit ikonischer Darstellung besser abschneiden als die mit piktogrammartiger. Zwar schnitten nach der explorativen Datenanalyse die Versuchs-

personen in der Bedingung „ikonisch-ohne-Differenz“ (ikon-oDiff) am besten ab, gefolgt von den Versuchspersonen der Bedingung „ikonisch-mit-Differenz“ (ikon-mitDiff). Jedoch ergab die Regressionsanalyse keinen bedeutsamen Unterschied. In einer Reihe von Experimenten ergab sich ein anderes Bild. Haroz et al. (2015) verglichen unter anderem Piktogramme mit einfachen „Formen“ (Kreis statt Piktogramm). Somit waren in beiden Fällen Häufigkeiten gegeben. Bei den Piktogrammbedingungen stellten jedoch die Piktogramme direkt den Inhalt dar, den die Diagramme thematisierten. Die Piktogrammdarstellung führte teilweise zu Vorteilen in der Erinnerungsleistung. Direkt auf die vorliegende Studie sind diese Ergebnisse nicht übertragbar, da hier ikonische Zeichen mit Piktogrammen verglichen wurden. Zudem ist bei BF-Aufgaben mehr verlangt als Erinnerungsleistung. Dennoch kann man es als Hinweis deuten, dass verschiedene symbolische Darstellungen zu unterschiedlichen Leistungen führen können. Dies zeigte sich in den Studien von Haroz et al. ebenfalls, wenn eine höhere kognitive Beanspruchung vorlag, was bei BF-Aufgaben gegeben sein sollte. Jedoch fanden sich in der vorliegenden Studie höchstens Tendenzen für ein besseres Abschneiden der ikonischen Bedingungen.

Nach den Regeln für Isotype sollten die Zeichen entsprechend dessen, wofür sie stehen, dargestellt werden (Neurath,

1936). Das sollte dann zu einem besseren Verständnis führen, was hier jedoch nicht der Fall war. Möglicherweise lag das an Unterschieden zwischen der hier genutzten Darstellung und den klassischen Isotype-Diagrammen: In der vorliegenden Studie ist die Zeitachse horizontal ausgerichtet, während bei Isotype die Zeitachsen meist vertikal vorliegen. Zu- und Abfluss sind voneinander getrennt, während bei Isotype in entsprechenden Fällen Zu- und Abfluss beieinander abgebildet sind (Neurath, 1936). In der vorhergehenden Studie schienen diese Eigenschaften jedoch nicht von Nachteil zu sein. Weitere Forschung sollte die offenen Fragen klären. Einerseits könnten andere Modifikationen in der Darstellung zu entsprechenden Ergebnissen führen wie, nach den oben beschriebenen Aspekten, erwartet. Andererseits könnte diese Form der Modifikation bei BF-Aufgaben nicht wirken, weil diese komplexer sind als die Aufgaben in oben beschriebenen Studien.

5.4.3 Betonung des Nettoflusses

Fragestellung 4.1 war zu verneinen: Die Bedingungen mit Betonung der Differenzen verbesserten nicht die BF-Leistung, sondern war eher ungünstig. Das war auch in anderen Studien der Fall, in denen zusätzlich der Nettofluss gegeben war (Schwarz, 2016; Kapitel 3.3.3), wobei sich dort teilweise die Leistung sogar

substantiell verschlechterte. Jedoch war in diesen Studien eine zusätzliche Abbildung für den Nettofluss gegeben, oder die Versuchspersonen mussten ihn ermitteln, während hier der Nettofluss in der Abbildung mit Zu- und Abfluss integriert war. Für die vorliegende Studie kann nur vermutet werden, was zu einer weniger guten Lösungsleistung führte. Möglicherweise war den Versuchspersonen nicht klar, was die sich ausgleichenden Werte in der Mitte der Abbildung bedeuten sollten. Eventuell war die Abbildung durch das Ergänzen dieser Ausgleichswerte schwieriger zu lesen und führte somit zu schlechterer BF-Leistung.

Es existiert eine BF-Studie, in der der Nettofluss indirekt in einer Abbildung vorlag und zumindest zu keiner Verschlechterung in der Lösungsleistung führte (Allenstein, 2013). In dieser Studie war der Nettofluss in Form von Pfeilen in Säulendiagramme eingezeichnet. Diese Pfeile zeigten die Differenzen von Zu- und Abfluss an. Somit war einerseits die Größe des Nettoflusses ersichtlich (hoher Nettofluss-Wert ergab einen längeren Pfeil), und andererseits war die Richtung betont: Grün für positiven und rot für negativen Nettofluss. Diese Pfeile zeigten im zusätzlichen Bestandsdiagramm, wie sich die Werte für den Bestand durch den Nettofluss verändern. Verglichen wurde dies unter anderem mit einer Bedingung, die genau so aufgebaut war, nur ohne die Nettofluss-Pfeile. In dieser Studie ergab sich kein Unterschied zwi-

schen den Bedingungen mit Nettofluss und ohne. Jedoch war in beiden Bedingungen ebenfalls der Bestand gegeben, wodurch möglicherweise die gegebene Nettofluss-Information nicht verarbeitet wurde. Denn Versuchspersonen könnten ausschließlich auf das gegebene Bestandsdiagramm achten, um die Fragen nach dem Bestand zu beantworten.

Insgesamt scheint also die Information über den Nettofluss nicht zu besserer BF-Leistung zu führen, oder bisher wurden nur ungünstige Darstellungen vom Nettofluss gezeigt. Weitere Studien könnten diesbezüglich genauere Erkenntnisse liefern.

5.4.4 Fehleranalyse

Zu *Fragestellung 4.2* nach den genutzten Fehlerarten ergab sich, dass die Versuchspersonen in allen Bedingungen die Korrelations-Heuristik verwendeten. Das entspricht den Ergebnissen bisheriger Studien (s. Auch Kapitel 1.1.5). Durchgängig war der häufigste Fehler das Anwenden der Korrelations-Heuristik orientiert am Nettofluss beziehungsweise am Netto- oder Zufluss, wenn das nicht zu unterscheiden war. Entsprechendes berichteten Cronin und Gonzalez (2007). Interessanterweise gaben in der Baseline-Bedingung viele Versuchspersonen beim diskontinuierlichen Verlauf mit Schnittpunkt nur einen der beiden korrekten Zeitpunkte an, wenn nach dem maximalen Bestand gefragt war.

Wäre dies als korrekt eingestuft worden, würden die Baseline-Bedingung und die Bedingung ikon-oDiff in diesem Fall etwa gleich gut abschneiden. Da in der vorliegenden Arbeit insgesamt eher nach strengeren Maßstäben[10] ausgewertet wurde, wurde auch hier die Bewertung für die Lösungsgüte nicht derart angepasst. Außerdem läge eine derartige Verbesserung der Leistung nicht bei den anderen Aufgaben vor.

Es traten weitere systematische Fehler auf. In der Bedingung Piktogramm-mit-Differenz (Pikt-mitDiff) bei den diskontinuierlichen Verläufen orientierte sich fast ein Fünftel der Versuchspersonen für die Antwort auf die Frage nach *minimalem Bestand* am maximalen Abfluss (Verlauf ohne Schnittpunkt). Die Anteile der anderen Bedingungen lagen unter diesem Wert. Bei der Frage nach dem *maximalen Bestand* gaben etwa ein Fünftel der Versuchspersonen einen Zeitpunkt mit *maximalem Zufluss* an (Verlauf mit Schnittpunkt). Der entsprechende Anteil für die anderen Bedingungen lag unter 10%. Eventuell hat die Bedingung Pikt-mitDiff die Versuchspersonen dazu verleitet, sich eher am Zu- oder Abfluss zu orientieren als am Nettofluss. Weitere Studien müssten klären, ob das tatsächlich der Fall war.

10 Beispielsweise werteten Cronin et al. (2009) bei der Kaufhaus-Aufgabe zusätzlich die Zeitpunkte 1 Minute vor bzw. nach der tatsächlichen Lösung als korrekt, was in der vorliegenden Arbeit nicht derart erweitert wurde.

Weitere systematische Fehler zeigten sich nur sehr selten, beispielsweise der Start- oder Endpunkt als fehlerhafte Lösung. Es ist nicht klar, wie die Versuchspersonen dazu kamen. In der Studie wurde nach dem Kriterium gefragt, woran die Versuchspersonen sich orientierten, um zu der Lösung zu kommen. Auch in diesen Antworten lässt sich diesbezüglich nichts ableiten. Eine Antwort war beispielsweise: "Das Verhältnis zwischen Geburten- und Sterberate". Ähnliches, was eher nach Korrelations-Heuristik mit Nettofluss klingt, wurde auch von anderen Versuchspersonen erwähnt. Das Fehlermuster mit Start- oder Endpunkt zeigte sich auch bei Cronin et al. (2009). Das vereinzelte Auftreten weiterer systematischer Fehler passt zu den Ergebnissen von Booth Sweeney und Sterman (2000), die ebenfalls von Fehlern berichteten, die von ihrer Idee nachvollziehbar waren, jedoch nur vereinzelt auftraten.

Grundlegend basierten also die meisten systematischen Fehler in irgendeiner Form auf einer Orientierung an den Flussgrößen. Das passt wiederum zu Ossimitz (2002) Annahme, das Unterscheiden von Flussgrößen und Bestandsgrößen führe zu Problemen beim Lösen der Aufgaben. Entsprechend hohe Anteile für das Nutzen der Korrelations-Heuristik zeigten sich bereits in anderen Studien. Beispielsweise untersuchten Cronin et al. (2009) ihr Auftreten genauer und fanden für 8 verschiedene

Flussgrößen-Verläufe Anteile von ca. 30% bis ca. 90%, bei denen die Korrelations-Heuristik genutzt wurde (Basisrate waren die falschen Lösungen). Im Mittel wurden bei etwa 70% der falschen Lösungen die Korrelations-Heuristik als Orientierung genutzt. Insgesamt zeigte sich in der vorliegenden Fehleranalyse, im Gegensatz zur vorhergegangenen Studie, kein grundlegender Unterschied zwischen der Baseline- und den symbolischen Bedingungen.

5.4.5 Analysen zum Einfluss von Geschlecht, Studienfach und Mathematiknote

5.4.5.1 Geschlecht

Zunächst ließ sich der Geschlechtereffekt zugunsten der Männer nicht bestätigen. Jedoch ist zu bedenken, dass in einzelnen Bedingungen extrem wenig Männer teilgenommen hatten. In der klassischen Diagrammdarstellung sind die Geschlechterunterschiede oft zu finden (vgl. Kapitel 2.4). Es war hier nicht möglich, dies gesondert zu untersuchen, da in der Baseline-Bedingung nur 3 Männer teilnahmen. Insgesamt zeigte sich kein Unterschied in der BF-Leistung von Männern und Frauen – weder in der Regressionsanalyse, noch bei den Einzelvergleichen für die Bedingungen (EDA). In diesem Fall unterschieden sich

nur in einem Fall die Mediane von Männern und Frauen stark: in der Bedingung Pikt-oDiff (Piktogramm-ohne-Differenz), an der jedoch nur 3 Männer teilnahmen. In der Baseline-Bedingung ist die Leistung der Frauen am schlechtesten, während sie im Darstellungsformat ikon-oDiff am besten abschneiden. Die restlichen Bedingungen liegen nach dieser Form der EDA in dem Leistungsbereich zwischen diesen beiden Bedingungen.

Ein problematischer Punkt war die Tatsache, dass in der Bedingung mit bester BF-Leistung (ikon-oDiff) der höchste Anteil an Männern vorlag. In der Regressionsanalyse war der Prädiktor Geschlecht weiterhin enthalten, so dass diese Variable für die Analysen des Darstellungsformats kontrolliert war. Das gilt gleichermaßen für den Einwand, dass in der Baseline-Bedingung sehr wenig Männer waren, was zu einer schwächeren Leistung in dieser Gruppe führen könnte. Jedoch sollte auch hier die Regressionsanalyse diesen Einfluss kontrollieren.

5.4.5.2 Studienfach

Für die Variable Studienfach fiel die Verteilung günstiger aus als für Geschlecht. Es handelte sich jedoch nur um Studierende zweier Fachrichtungen, SeKo und Psychologie. Die Leistung unterschied sich kaum für die beiden Gruppen und wurde daher nicht detailliert analysiert. Tendenziell schnitten Psychologiestu-

dierende besser ab. SeKo ist eigentlich naturwissenschaftlich näher (angesiedelt an der Fakultät für Naturwissenschaften, zudem physikalische Inhalte im Curriculum). Andererseits besteht an der TU Chemnitz für Psychologie ein Numerus Clausus, für SeKo jedoch nicht. Tatsächlich zeigten EDA und Mittelwerte, dass Psychologiestudierende bessere Mathematiknoten angaben.

5.4.5.3 Mathematiknote

Als weiterer Indikator für mathematisches Vorwissen beziehungsweise Fähigkeiten oder Fertigkeiten wurde die Mathematiknote erhoben. Sie stellte einen der besten Prädiktoren in der Regressionsanalyse dar. Mathematisches Vorwissen hat also großen Einfluss auf die Leistung. Das stimmt mit Ergebnissen aus anderen Studien überein (Lungwitz et al., 2018; Schwarz, 2016).

5.4.6 Analysen zur Geschlechterrolle

Fragestellung 4.3 betraf die Frage, ob die Geschlechterrolle die BF-Leistung beeinflusst. Es bestand ein Zusammenhang zwischen Femininitäts-Skala und BF-Leistung: Je höher die Werte der Femininitäts-Skala, desto schlechter die Leistung. Dieser Zusammenhang war stärker im vollständigen Modell.

In bisherigen Studien mit mathematischen Aufgaben schnitten Personen der Kategorie „maskulin“ besser ab als die anderen

Kategorien-Gruppen (z.B. Selkow, 1985). Als maskulin werden die Personen bezeichnet, die nach Median-Split hohe Werte auf der Maskulinitäts-Skala und niedrige auf der Femininitäts-Skala aufweisen. Das ist eine Form der Auswertung, die hier nicht für die Erklärung der BF-Leistung herangezogen wurde, da es von Vorteil erschien, die BSRI-R-Skalen als Prädiktoren zu verwenden und somit das Vorliegen der feineren Abstufungen zu nutzen. Zumindest ist das Ergebnis kein Widerspruch zum Ergebnis der vorliegenden Studie, denn beim Vorgehen nach Median-Split liegen in der Kategorie „maskulin“ eher niedrige Werte für die Femininitäts-Skala vor. Somit weist das Ergebnis in dieselbe Richtung, wie das Ergebnis für die vorliegende Regression.

Femininitäts-Skala ist ein signifikanter Prädiktor, während sich bei Maskulinitäts-Skala und bei der Interaktion keine signifikanten Ergebnisse ergaben. Das spricht für einen Einfluss der Geschlechterrolle: Je stärker die weibliche Geschlechterrolle ausgeprägt ist, desto schlechter ist die BF-Leistung. In diesem Bereich könnte es sich lohnen, weitere Studien durchzuführen, um den Einfluss der Geschlechterrolle genauer zu untersuchen.

5.4.7 Ablesen der Flussgrößen

Insgesamt traten Fehler beim Ablesen der Flussgrößen auf. Tendenziell lag die Leistung unabhängig von den Bedingungen

etwa im gleichen Bereich, nämlich mit 90% und mehr an mittlerer prozentualer Leistung. Eindeutig aus der Reihe fiel die Bedingung Pikt-mitDiff mit 82%. Bei Ergebnissen aus den Studien von Cronin et al. (2009) und weiteren lagen die korrekten Angaben zu diesen Fragen fast immer über 90% (Brunstein, Gonzalez & Kanter, 2010; Cronin & Gonzalez, 2007; Pala & Vennix, 2005; Gonzalez & Wong, 2012; Pala & Vennix, 2005; Sterman, 2002; Strohhecker, 2009). Vereinzelt trat jedoch auch in diesen Studien schlechtere Leistung auf, wie beispielsweise 84% korrekte Antworten bei der Frage nach dem maximalen Abfluss (Gonzalez & Wong, 2012). Somit lagen die korrekten Antwortraten zu den Flussgrößen in einem Bereich entsprechend der Leistung in der Literatur. Ob das schlechtere Abschneiden der Bedingung Pikt-mitDiff zufällig war oder sich aufgrund des Darstellungsformats ergab, müssten weitere Studien klären. Ein Fehler, der nur für diese Bedingung auftrat, war die Angabe des maximalen Nettoflusses statt maximalen Zuflusses. Möglicherweise wurde dieser Fehler durch die Darstellung forciert: In der Bedingung Pikt-mitDiff war der Nettofluss stark betont. Das würde mit Erkenntnissen aus der Literatur zur Diagramm-Darstellung übereinstimmen, wonach die Darstellung beeinflusst, welche Informationen leicht erkenntlich sind und welche Folgerungen gezogen werden (vgl. Kapitel 1.2.4).

5.4.8 Vorherige Teilnahme, Abbruchquote und fehlende Werte

5.4.8.1 Vorherige Teilnahme

In dieser Studie zeigten die Versuchspersonen, die schon mal an einer BF-Studie teilgenommen hatten, eine bessere BF-Leistung. Zwar ist das im Rahmen der Regressionsanalyse nicht signifikant, jedoch ist zu bedenken, dass es sich nur um 8 Versuchspersonen handelte. Das könnte einerseits bedeuten, dass ein tatsächlich vorliegender Effekt nicht signifikant wird; andererseits kann es sich einfach um einen Zufall handeln, dass diese Versuchspersonen besser abschnitten. Möglicherweise besteht ein Zusammenhang zwischen diesem Ergebnis und der Abbruchquote.

5.4.8.2 Abbruchquote

Eventuell brachen Versuchspersonen die Studie ab, wenn sie bereits an einer BF-Studie teilgenommen hatten und diese als problematisch oder unangenehm zu bearbeiten empfunden hatten. Das könnte zu höheren Abbruchquoten geführt haben als bei denjenigen, die höheres BF-Verständnis aufwiesen. Vermutlich haben auch andere Personen (ohne vorhergehende Teilnahme an

BF-Studien) abgebrochen, wenn es für sie zu mühsam war. Das ist allgemein ein Problem bei Online-Studien.

Zudem liegen unterschiedlich hohe Abbruchquoten in den verschiedenen Bedingungen vor. Es ist unklar, warum die Quote in Bedingung ikon-oDiff am höchsten ist. Die Zusatzvariablen unterschieden sich nicht für die verschiedenen Bedingungen, wobei derartige Variablen nicht von den AbbrecherInnen erhoben werden konnten. Jedoch ist zumindest auszuschließen, dass ein sehr grundlegender Effekt bezüglich einer dieser Variablen vorlag. Außerdem ist davon auszugehen, dass aufgrund der Möglichkeit einer Entlohnung über Versuchspersonen-Stunden auch Personen weitermachten, die die BF-Aufgaben langweilig oder anstrengend fanden. Es ist nicht nachvollziehbar, warum die Versuchspersonen abbrachen: ob sie die Aufgaben zu anstrengend fanden oder ob äußere Umstände (Störung oder ähnliches) dazu führten. Möglicherweise wollten Versuchspersonen neu starten, wenn sie dachten, sie hätten bei den vorhergehenden Aufgaben Fehler gemacht. Denn es war nicht möglich gewesen, zu vorhergehenden Aufgaben zurückzugehen und dort Antworten zu ändern. Die ursprüngliche Idee, die IP-Adressen zu loggen, um derartiges zu erfassen, wurde aufgrund von Bedenken betreffend des Datenschutzes verworfen.

Eine weitaus wichtigere Frage ist, ob das unterschiedliche Abbruchverhalten Folgen für die Ergebnisse hat. Im ungünstigsten Fall kam die beste BF-Leistung in der Bedingung ikon-oDiff zustande, weil hier mehr Personen mit niedrigem oder keinem BF-Verständnis abbrachen. Andererseits hat in der Baseline-Bedingung fast die gleiche Anzahl an Personen abgebrochen und die Leistung war dort einiges schlechter bei den Versuchspersonen, die die Studie komplett bearbeiteten.

Eine Möglichkeit herauszufinden, ob durch die Abbrecher die Leistung in der Studie überschätzt wurde, könnte über einen Vergleich von vorliegenden Ergebnissen mit Ergebnissen aus der Literatur erfolgen. Wenn Werte in der vorliegenden Baseline-Bedingung stark von den Werten der Baseline-Bedingungen anderer Studien abweichen, so könnte das ein Hinweis für eine ungewöhnliche Stichprobe sein. Der diskontinuierliche Verlauf ohne Schnittpunkt war ziemlich gleichartig von Cronin et al. (2009) übernommen. In seiner Studie ergab sich in der Baseline-Bedingung für maximalen Bestand 8% und für minimalen Bestand 16% korrekte Antworten. In der vorliegenden Studie waren es 39% (Bestand-maximal) und 43% (Bestand-minimal). Bis zum aktuellen Zeitpunkt sind mir keine weiteren Studien bekannt, die diesen Verlauf nutzten. Das spricht eher für ein besseres Abschneiden der Versuchspersonen in der vorliegenden Arbeit. Für

den diskontinuierlichen Verlauf mit Schnittpunkt ergab sich durch die Verkürzung ein stärker abgewandeltes Muster als wenn beispielsweise beim Dreieckverlauf weniger Zeitpunkte gegeben sind. Die grundlegende Struktur ist im zweiten Fall weiterhin erhalten. Daher werden die Werte für diskontinuierlich mit Schnittpunkt nicht für den Vergleich mit Ergebnissen aus anderen Studien genutzt.

Beim Dreieckverlauf ergaben sich in vorliegender Studie 14% korrekte Antworten für beide Fragen nach den Bestand. In der Literatur ist mir kein berichtetes Ergebnis bekannt zu Fragen beim Dreieckverlauf. Schwarz (2016) berichtete, dass in seinen Studien die Werte aus gezeichneter Lösung und Antworten mit etwa .8 bis .9 korrelierten. Wenn der Verlauf zu zeichnen ist, schwanken die mittleren prozentualen Anteile im Bereich von 28% bis 51% (Armenia, Onori & Bertini, 2004; Booth Sweeney & Sterman, 2000; Kapmeier, Happach, Tilebein, 2014; Lungwitz et al., 2018; Ossimitz, 2002; Schwarz, 2016). In der eigenen ersten Studie (Kapitel 2) mit dem extrem schlechten Abschneiden beim Zeichnen des Verlaufs, lag die mittlere prozentuale Leistung ebenfalls bei 14%; in der dritten Studie lag sie in der Baseline-Bedingung bei 27%. Das ist eher ein Zeichen für schlechteres Abschneiden im Vergleich mit bisherigen Ergebnissen. Aus den berichteten Zahlen ist es schwierig abzulesen, ob die Stichprobe

aufgrund der Abbruchzahlen besser abgeschnitten hat. Weitere Studien sollten klären, ob eine Online-Studie tatsächlich entsprechende Ergebnisse erbringt, wie eine Labor-Studie.

Falls die Abbruchquote einen Indikator für die Lesefreundlichkeit der Diagramme darstellt, so ergibt sich eine Tendenz für ikon-oDiff als unangenehmste Darstellung und Pikt-mitDiff als angenehmste. Das könnte ein Ausgangspunkt für weitere Studien sein, um Derartiges zu untersuchen. Insgesamt wäre eine Replikation im Labor das Günstigste, um zu prüfen, ob die unterschiedlichen Abbruchzahlen in den verschiedenen Bedingungen einen Einfluss haben und ob sich durch die online-Erhebung veränderte BF-Leistung ergibt.

5.4.8.3 Fehlende Werte

In der vorliegenden Studie wurden besonders oft in der Bedingung Pikt-mitDiff demographische Angaben ausgelassen. Weiterhin fällt auf, dass ein Großteil dieser Personen 0 Punkte in der Gesamtsumme aufwies. Für den Fall, dass die Versuchspersonen bewusst die Ausprägung von Geschlecht oder Studienfach nicht angaben, stellt sich die Frage nach den Gründen. In den Anmerkungen zum Vorgehen fanden sich keine Hinweise für die Gründe. Günstigerweise ergaben sich keine drastischen Unterschiede zwischen den Regressionsanalysen mit den und ohne die

unvollständig ausgefüllten Variablen. Somit wird an dieser Stelle auf Überlegungen bezüglich möglicher Ursachen verzichtet.

5.4.9 Fazit

Das Ergebnis hier ist nicht mehr so eindeutig wie in der vorhergegangenen Studie (Kapitel 4). Zwar verbesserte sich die Leistung im Gegensatz zur Baseline-Bedingung, jedoch nicht mehr so stark. Mit dem symbolischen Darstellungsformat scheint das BF-Verständnis nur dann stark aktiviert oder hervorgerufen zu werden, wenn die Lösung für jeden einzelnen Zeitpunkt angegeben werden soll. Das legt die Vermutung nahe, dass das Berücksichtigen und Berechnen zu einzelnen Zeitpunkten ein starker Wirkfaktor ist. Andererseits scheint auch die symbolische Darstellung zu wirken und die Leistung zu verbessern.

6 Diskussion der gesamten vorliegenden Arbeit

In dieser Diskussion erfolgt ein Überblick der Ergebnisse: Zunächst geht es um die Auswirkungen der Manipulationen betreffend der BF-Aufgaben. Weiterhin fasse ich Ergebnisse zu den Fehleranalysen zusammen. Es folgen die untersuchten Kontrollvariablen und ihr Zusammenhang mit der BF-Leistung. Im weiteren Verlauf gehe ich auf die Frage ein, ob Intuitionen zu BF-Systemen überhaupt bestehen. Abschließend folgen Implikationen betreffend der Kommunikation von BF-Systemen und ein kurzes Fazit .

6.1 Einfluss externaler Repräsentation auf BF-Leistung – eine Zusammenfassung

Vielfach wurde in Forschung zu BF-Systemen gezeigt, dass diese schwierig zu verstehen sind. Eine modifizierte Darstellung der Flussdiagramme sollte diesen Verständnisproblemen entgegenwirken. Eine erste Studie sollte unter anderem klären, ob dieses Vorgehen sinnvoll sein könnte. Tatsächlich war ein klassisches Flussdiagramm (Liniendiagramm) nicht hilfreich, um BF-Aufgaben korrekt zu lösen. In den weiteren Studien wurden so-

mit verschiedene Formate für das Flussdiagramm näher betrachtet. Es folgen die zusammengefassten Ergebnisse.

6.1.1 Studie 1

In Kapitel 2.3 zeigte sich eine sehr schlechte BF-Leistung: Die Aufgabe mit Dreieckverlauf lösten so wenige Versuchspersonen komplett oder auch nur teilweise, dass dieses Problem für die Analysen nicht genutzt wurde (außer für die Fehleranalyse). Für den Vergleich der Studien liegen also nur die Daten des Rechteckverlaufs vor.

Konfidenzintervalle der β-Werte

Um die Effekte in den Studien zu vergleichen, wurden Konfidenzintervalle der β-Werte für die einzelnen Bedingungen der Studien erstellt. Abbildung 6.1 zeigt die β-Werte für die Bedingungen aus der ersten Studie, wobei die horizontalen Linien jeweils ein 90% Konfidenzintervall darstellen. Diese basieren auf Bootstrap-Verfahren mit 1000 simulierten Stichprobenziehungen, durchgeführt in SPSS. Bei negativen β-Werten hatte sich die BF-Leistung verschlechtert. Es ergaben sich folgende gerundete Mittelwerte: Abbildung .24; Diagramm -.20 und BF-Größe .05.

Für den Rechteckverlauf war eine schematische Abbildung hilfreich, während ein Flussdiagramm zusätzlich zu einem aus-

führlichen Text die BF-Leistung verschlechterte. Auch in anderen Studien führte das Flussdiagramm nicht zu besserem BF-Verständnis (vgl. Kapitel 2.4.2). In Studie 1 zeigte die BF-Größe keinen Einfluss, außer wenn zusätzlich die Position berücksichtigt wird (Kapitel 2.3.6). Die diskrete BF-Größe Murmeln führte an zweiter Position zu besseren Lösungen als an erster Position und zu besserer Leistung als die kontinuierliche Größe Wasser und das unabhängig von der Position.

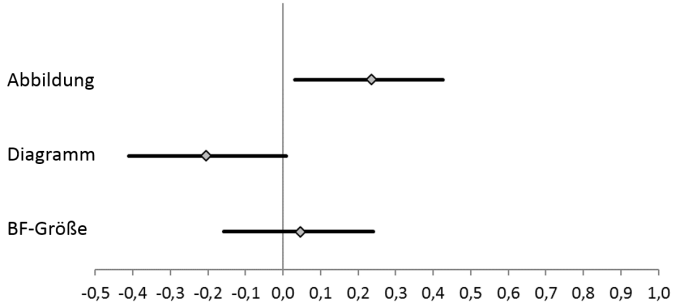


Abbildung 6.1: β-Werte für die unabhängigen Variablen betreffend der Darstellung aus der ersten Studie mit 90%-KI (Bootstrap-Verfahren mit 1000 Ziehungen).
Anmerkung: die unabhängigen Variablen waren Abbildung (ohne schematische Abbildung 0, mit 1); Diagramm (ohne Flussdiagramm 0, mit 1); BF-Größe (Murmel 0, Wasser 1).

Aus dem Ergebnis betreffend des Flussdiagramms wurde gefolgert, dass das Ansetzen an der Diagrammdarstellung sinnvoll sein könnte.

6.1.2 Studien 2 bis 4

In Kapitel 3 zeigte die Baseline-Bedingung das beste Ergebnis. Die Modifikationen „Nettofluss-gegeben“, „Nettofluss-ermitteln“, und „geteiltes-Koordinatensystem“ für die Flussgrößen verschlechterten das BF-Verständnis. Auch in einer weiteren Studie führte das zusätzliche Geben des Nettoflusses zu keiner Verbesserung (Schwarz, 2016). In der folgenden Studie wurden die Flussgrößen wie Bestandsgrößen dargestellt. Dadurch ergab sich eine Darstellung, die nicht mehr im Bereich der konventionalisierten Diagramme lag. In Kapitel 4 wurde eine symbolische Darstellung gewählt, welche zu einer starken Verbesserung der BF-Leistung führte. Die Darstellung über einzelne Piktogramm-Menschen zeigte in etwa gleich gute Lösungsraten wie die Darstellung über Wasserbehälter, trotz der vermutlich ungünstigen Gestaltung der Lösungsdiagramme über ein Dreier-System. In der folgenden Studie (Kapitel 5) wurden Fragen nach dem Bestand zu einzelnen Zeitpunkten gestellt. Somit sollte eher der Gesamtüberblick im Fokus liegen. In dieser Studie schnitten die Bedingungen mit piktogrammartiger oder ikonischer Darstellung

wiederum besser ab als die Baseline-Bedingung. Zusätzliches Betonen des Nettoflusses innerhalb der Abbildung war nicht hilfreich. Die beste Leistung ergab sich in der Bedingung mit ikonischer Darstellung ohne zusätzliches Betonen des Nettoflusses. Insgesamt jedoch hat sich die Leistung in den Bedingungen mit piktogrammartiger oder ikonischer Darstellung nicht so stark verbessert wie in der vorhergegangenen Studie.

Konfidenzintervalle der β-Werte

Auch für diese drei Studien folgen Konfidenzintervalle der β-Werte. Die Regressionsanalysen enthielten das vollständige Modell der jeweiligen Studien. Es wurden also alle zusätzlichen Prädiktoren wie in Kapitel 3 bis Kapitel 5 genutzt, mit einer Ausnahme: In allen Fällen wurde die jeweilige Bedingung nur gegen die Baseline getestet. Die Dummy-Variablen der anderen Bedingungen waren also nicht mehr im Modell enthalten. Abbildung 6.2 zeigt Konfidenzintervalle der β-Werte (Vorgehen wie in Abbildung 6.1)[11]. Es ergaben sich folgende gerundete Mittelwerte: für KS-geteilt -.12; für NF-gegeben -.13; für NF-ermitteln -.28;

11 Vereinzelt wurden Bootstrap-Stichproben aus der Analyse gelöscht. Das galt bspw. für „vorherige Teilnahme“ (4. Studie). Hier waren nur 8 Personen, die für diese Variable „ja“ angegeben hatten, während der Rest „nein“ angab. Wenn nur Personen mit der Angabe „nein“ gezogen wurden, stellte die Variable eine Konstante dar. Dadurch ergaben sich für die Analysen 944 bis 1000 Stichproben.

für die Wasser-Bedingung .60; für die Bus-Bedingung .54; für ikon-oDiff .30; für ikon-mitDiff .21; für pikt-oDiff .17 und für pikt-mitDiff ebenfalls .17.

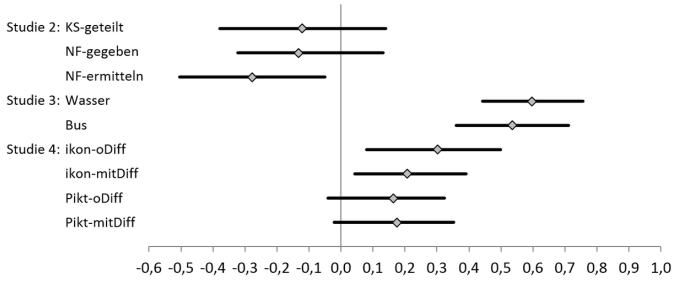


Abbildung 6.2: β-Werte für die unabhängigen Variablen betreffend der Darstellung aus den Studien 2 bis 4 mit 90%-KI (Bootstrap-Verfahren mit 1000 Ziehungen), jeweils die Bedingungen versus die Baseline-Bedingung.
Anmerkung: Studie 2: KS-geteilt: Flussgrößen zusammen in einem Quadranten 0, Flussgrößen im getrennten Koordinatensystem 1; NF-gegeben: nein 0, ja 1; NF-ermitteln: nein 0, ja 1;
Studie 3: Wasser: Wasser-Szenario nein 0, ja 1; Bus: Bus-Szenario nein 0, ja 1;
Studie 4: ikon-oDiff: ikonisch-ohne-Differenz; ikon-mitDiff: ikonisch-mit-Differenz; Pikt-oDiff: Piktogramm-ohne Differenz; Pikt-mitDiff: Piktogramm-mit-Differenz; jeweils: nein 0, ja 1.

Die vorliegenden Studien wiesen darauf hin, dass eine Kombination aus symbolischer Darstellung und schrittweisem Vorgehen die besten Ergebnisse erzielte. Zudem schien eine Darstellung mit ikonischen Zeichen zu leicht besserer Leistung zu führen als mit piktogrammartiger Darstellung. In Kapitel 5.3.2 ergab sich kein besseres Abschneiden der ikonischen Bedingungen im Vergleich mit den piktogrammartigen. Auch hier wird ersichtlich, dass die Unterschiede nicht sehr groß waren. Jedoch führten die Vergleiche von nur der jeweiligen Bedingung mit Baseline dazu, dass die Konfidenzintervalle der Piktogramm-Bedingungen den Wert 0 überdecken, während das für die ikonischen Bedingungen nicht der Fall ist.

6.2 Fehleranalyse

Das durchgängig häufigste Fehlermuster war die Korrelations-Heuristik. In den vorliegenden Studien wurden die falschen Lösungen dahingehend untersucht, wann die Versuchspersonen die Korrelations-Heuristik nutzten. In Studie 1 (Kapitel 2) wurde die Korrelations-Heuristik relativ selten angewandt. Es handelte sich um die Stichprobe mit vermutlich geringstem BF-Verständnis (wegen des hohen Anteils an Frauen aus nicht-naturwissenschaftlichen Fächern). Die Probleme schienen für die Versuchspersonen sehr schwierig zu sein. Unter Umständen wird die

Korrelations-Heuristik bei zu hoher Schwierigkeit seltener angewandt – das müssten weitere Studien klären.

In Kapitel 4.3.3 zeigte sich, dass bei den leicht zu lösenden Aufgaben in symbolischem Darstellungsformat die Korrelations-Heuristik so gut wie gar nicht auftrat. Möglicherweise gehen falsche Lösungen auch bei anderen leichten BF-Aufgaben eher auf Flüchtigkeitsfehler zurück. Unter Umständen wenden die Versuchspersonen die Korrelations-Heuristik am ehesten im mittelschwierigen Bereich an. In der Literatur vergleichbar könnten die Ergebnisse von Cronin et al. (2009) angesehen werden. Zwar ergab sich dort bei mittelschweren und schwierigen Aufgaben etwa der gleiche Anteil an Korrelations-Heuristik in den falschen Lösungen. Jedoch handelte es sich um einen Vergleich innerhalb der Studie, und eventuell wurde dort als schwieriger Bereich angesehen, was insgesamt betrachtet eher im mittelschweren Bereich liegt.

Weiterhin orientierten sich die meisten Versuchspersonen beim Anwenden der Korrelations-Heuristik am Nettofluss. In Kapitel 3.3.4 ergab sich ein erster Hinweis innerhalb dieser Arbeit dafür, dass eher der Nettofluss als die anderen Flussgrößen als Orientierung für die Korrelations-Heuristik genutzt wurde. Das zeigte sich in der Bedingung KS-geteilt: In dieser Bedingung war es möglich, bei falschen Lösungsgraphen zu erkennen, welche

Flussgröße als Orientierung für die Korrelations-Heuristik genutzt wurde. Alle Versuchspersonen orientierten sich in diesem Fall am Nettofluss. In Kapitel 5.3.3 ließen sich die meisten Fehler systematisieren. Die Versuchspersonen wandten oft die Korrelations-Heuristik an, insbesondere orientierten sie sich dann am Nettofluss.

Daraus wurde die Vermutung abgeleitet, dass ein Großteil der Personen die Flussgrößen verrechnet und die Akkumulation vergisst (ausführlicher beschrieben in Kaptiel 3.1.2). Ursprünglich war meine Annahme zur Ursache des Problems, dass es in den meisten gängigen Liniendiagrammen um Vergleiche geht. Das heißt, wenn beispielsweise in den Medien Liniendiagramme gezeigt werden, so sollen häufig die Werte verglichen werden. Daraus könnte sich ein Schema ergeben, die beiden gegebenen Größen zu verrechnen. In Kapitel 1.2.2 war beschrieben, dass teilweise Hinweisreize aus einer Problemstellung zu einem bestimmten, fehlerhaften Vorgehen führen. In diesem Zusammenhang auch interessant sind so genannte Operatoren-Schemata (operator schemata) und irreführende Korrelationen. Bei den Operatoren-Schemata (Lewis & Anderson, 1985) handelt es sich um eine Gruppe von Schemata, die das Aufspüren von Korrelationen zwischen Problemmerkmalen und Operator oder Algorithmus zum Lösen beinhalten. Die Idee dahinter war, dass Schüle-

rInnen lernen (explizit oder implizit), einen Hinweisreiz aus dem Problem mit bestimmter Problemlösestrategie zu assoziieren. Das passt zusammen mit dem Anwenden der Korrelations-Heuristik orientiert am Nettofluss als Folge davon, dass Liniendiagramme im Alltag eher für Vergleiche genutzt werden. Dadurch könnte „Differenzbilden“ ein sehr automatisierter Prozess sein und der Hinweisreiz eines Liniendiagramms mit zwei Linien führt sehr stark darauf hin zu. Dagegen spricht, dass auch in den symbolischen Bedingungen die Korrelations-Heuristik oft angewandt wird und dort ebenfalls häufig orientiert am Nettofluss (Kapitel 5.3.3). Diese Form eines Schemas, so vorhanden, scheint also grundlegender vorzuliegen als dass es ausschließlich durch die Darstellung in Liniendiagrammen getriggert wird.

Eine weitere häufig auftretende Fehlerkategorie zeigte sich in Kapitel 4: In den Bedingungen mit symbolischer Darstellung waren häufig Rechenfehler die Ursache für die falschen Lösungen. Die Korrelations-Heuristik tritt dort fast gar nicht auf, während sie in der Baseline-Bedingung vorliegt. In dieser gibt es zudem „sonstige Fehler“ (im Sinne von nicht systematisierbar). Rechenfehler treten in der Baseline-Bedingung nur beim diskontinuierlichen Flussgrößen-Verlauf auf. Das kann als Hinweis betrachtet werden, dass der „Zwang“ zum Rechnen einen Wirkfaktor darstellt. Dieser Zwang scheint in der Darstellung mit

Liniendiagramm nur zu wirken, wenn einerseits der Verlauf gezeichnet werden muss und andererseits ein diskontinuierliches Muster vorliegt (wie in der Studie aus Kapitel 4). Die Leistung war dann immer noch schlechter als bei den symbolischen Darstellungen. Jedoch war die Leistung im Vergleich der Aufgaben innerhalb der Baseline-Bedingung auf recht gutem Niveau, wenn Versuchspersonen mit vereinzelten Rechenfehlern dazugenommen werden, unter der Annahme, diese hatten das Prinzip verstanden.

6.3 Personenmerkmale

Es folgt ein kurzer Überblick über Ergebnisse aus den vorliegenden Studien zu den Variablen Geschlecht, mathematisches Verständnis und Geschlechterrolle.

6.3.1 Geschlecht

Männer schneiden besser ab – das zieht sich durch die gesamte BF-Literatur (vgl. Kapitel 2.4). In den gegenwärtigen Studien zeigten sich entweder Tendenzen (Kapitel 3.3.5) oder eindeutig bessere Leistung (Kapitel 4.3.4.1) für die Männer in der Baseline-Bedingung (klassische Darstellung). In den Bedingungen mit symbolischer Darstellung hob sich der Geschlechtereffekt auf. In der letzten Studie (Kapitel 5) befanden sich ungünsti-

gerweise in der Baseline-Bedingung nur sehr wenige Männer, weshalb dort die Ergebnisse zum Geschlechtereffekt weder unterstützt noch ihnen widersprochen werden konnten.

6.3.2 Mathematisches Verständnis

In der zweiten Studie schnitten Studierende jener Fächer besser ab, in deren Curricula mehr mathematische Inhalte gelehrt werden (Kapitel 3.3.5.2). Das galt teilweise auch in der folgenden Studie mit symbolischer Darstellung (Kapitel 4). Dort beeinflusste das Studienfach, definiert über „naturwissenschaftlich nah“ versus „naturwissenschaftlich fern“, insbesondere in der Baseline-Bedingung die BF-Leistung. Wohingegen in den Bedingungen mit symbolischer Darstellung dieser Einfluss reduziert war (Kapitel 4.3.4). In der letzten Studie zeigte das Studienfach keinerlei Einfluss auf die BF-Leistung, was auf den Vergleich von ausschließlich Psychologie und SeKo zurückgehen könnte. Die Mathematiknote erwies sich jedoch als der beste Prädiktor in der Regressionsanalyse (Kapitel 5.3.4).

6.3.3 Geschlecht, Geschlechterrolle und mathematisches Verständnis

Um den Einfluss der Geschlechterrolle zu prüfen, wurde in der letzten Studie der BSRI-R (s. Kapitel 5.2.4) zusätzlich gegeben. Tatsächlich stellte die Femininitäts-Skala einen guten Prädiktor für die BF-Leistung dar: je höher ihre Werte, desto schlechter die Leistung.

Die BF-Szenarien können im Bereich Problemlösen eingeordnet werden. Eine Pisa-Studie verglich die Leistung von Jungen und Mädchen im Bezug auf Problemlösen, wobei die Probleme in ihrer Konzeption dem mathematischen Bereich zuzuordnen waren (Budde, 2009). Hier erbrachten Mädchen gute Leistung im Bereich Problemlösen. Die Mädchen zeigten einerseits bessere Leistungen als sie es im Bereich Mathematik zeigten. Andererseits schnitten Mädchen besser ab als Jungen im Teilbereich Problemlösen. Budde sah die Ursachen für das schlechtere Abschneiden von Mädchen in Mathematik eher im Unterricht: bei ihrem Selbstvertrauen oder ihren Erwartungen an die eigenen Fähigkeiten. Somit könnte im Aufgabenbereich Problemlösen hilfreich gewesen sein, dass die Problemstellungen weniger „mathematische Symbolik“ als im Mathematikunterricht aufwiesen und dass sie stärker in einen Kontext eingebettet waren. Beim Problemlö-

sen ergab sich trotz mathematischen Anforderungsprofils ein anderes Bild betreffend der Geschlechter, und zwar zum Vorteil der Mädchen. Bei BF-Aufgaben mit klassischer Darstellung könnte die oben genannte „mathematische Symbolik“ über das Liniendiagramm zum Tragen kommen. Wenn diese einen Einfluss darauf hat, wie stark ein Leistungsunterschied zwischen Männern und Frauen ausfällt, so sollte das auch bei BF-Aufgaben Auswirkungen haben. Weitere Studien könnten in diesem Bereich einige offene Fragen klären.

6.3.4 Fazit

Personenmerkmale (bezüglich mathematischer Vorkenntnisse und Geschlecht) waren in den vorliegenden Studien von Bedeutung, wobei sich ihr Zusammenhang mit BF-Leistung bei der symbolischen Darstellung abschwächte bis verschwand. Nur die Mathematiknote bleibt auch in der symbolischen Darstellung bester Prädiktor, wobei diese Variable in der dritten Studie (Kapitel 4) nicht erhoben wurde. Bei mathematischen Kenntnissen und Geschlechterrolle (Femininitäts-Skala) scheint es sich um zwei unabhängige Faktoren zu handeln. Sie waren gemeinsam in der Regressionsanalyse (Kapitel 5.3.2) enthalten und zeigten beide einen Effekt.

6.4 Intuition

Das Ziel der vorliegenden Arbeit war es, durch Ändern der Darstellung der Flussgrößen über Aktivierung von intuitivem Wissen das BF-Verständnis zu erhöhen. Hierbei wurde vorausgesetzt, dass intuitives Wissen für BF-Systeme vorliegt, auch wenn dies in der Literatur teilweise den Menschen abgesprochen wurde (vgl. Kapitel 1.2.1).

Die vorliegenden Studien ergaben Folgendes: Klassische Liniendiagramme führten zu stark fehlerhaftem BF-Verhalten. Bei vielen Versuchspersonen ergab sich immer wieder als fehlerhafte Lösung das Anwenden der Korrelations-Heuristik: Der Bestand ändert sich entsprechend einer der Flussgrößen. Eine Veränderung der klassischen Diagramm-Darstellung in ein symbolisches Darstellungsformat führte zu extremer Verbesserung in der Leistung, wenn die Versuchspersonen die Lösung für die einzelnen Zeitpunkte berechneten. BF-Aufgaben in symbolischen Darstellungen waren nicht mehr so gut zu lösen, wenn es eher um den Gesamtüberblick ging.

Die letzten beiden Punkte führen zu der Frage, ob für das Bearbeiten von BF-Aufgaben Intuitionen vorliegen. Einerseits spricht die Tatsache, dass Versuchspersonen über das Rechnen zur korrekten Lösung gelangen, für ein eher „analytisches“ Vor-

gehen. Andererseits könnten die Versuchspersonen auch in Studien wie bei Booth Sweeney und Sterman (2000), in denen die exakten Werte verlangt sind, ihre Lösung berechnen. Tatsächlich existierten Hinweise, dass die Versuchspersonen teilweise rechneten. Zum Beispiel waren bei den Antwortdiagrammen Rechnungen zu finden, oder in den Fragen zum Vorgehen schrieben Versuchspersonen, dass sie rechneten (z.B. Booth Sweeney & Sterman, 2000; Studie 1; vereinzelt in Studie 2; in Studie 3 und Studie 4 in der Baseline-Bedingung). Jedoch führte das Rechnen oft zur falschen Lösung. Somit schien die symbolische Darstellung den zusätzlichen Effekt zu haben, dass diese simple Form des Rechnens tatsächlich angewandt wurde. Es könnte zusätzlich wirken, was eingangs unter „sekundärer Intuition“ abgehandelt wurde. Zwar ist es unwahrscheinlich, dass das Vorgehen bei BF-Aufgaben im Mathematikunterricht erlernt wurde. Aber es könnte sein, dass die Darstellung hilft, den korrekten Lösungsweg zu finden – also eine Kombination aus intuitivem Triggern und analytischem Vorgehen, um zum korrekten Rechenweg zu kommen. Dafür spricht auch, dass analytisches Vorgehen stark mit „sequential processing“ zusammenhängt (Plessner et al., 2008), was in Kapitel 4 durch das Zeichnen des Bestands zu jedem einzelnen Zeitpunkt stark gefördert wird.

Nun zu der allgemeineren Frage: Bestehen überhaupt Intuitionen zu BF-Systemen? Nach ursprünglichen Annahmen von Booth Sweeney und Sterman (2000) sollten wir aus dem Alltag heraus Intuitionen für BF-Systeme entwickeln, da wir immer wieder mit ihnen konfrontiert sind und mit ihnen umgehen können. Beispielsweise füllen wir eine Badewanne mit Wasser, und wir managen unser Konto. Dennoch scheinen diesbezüglich die wenigsten Versuchspersonen über intuitives Wissen zu verfügen. Andererseits stellt sich bei vielen dieser Beispiele die Situation im Alltag anders dar. Wenn wir eine Badewanne füllen, regulieren wir den Zufluss und nicht parallel dazu den Abfluss; dieser ist im Normalfall verschlossen. Wir können mit unserem Konto umgehen, aber auch hier besteht auf den Kontoauszügen keine Parallelität von Zu- und Abfluss. Außerdem ist der finale Kontostand zusätzlich angegeben. Die einfachsten Beispiele sollten den meisten Personen einleuchten: Wir füllen unsere Kaffeetasse, und je mehr wir auf einmal einschenken, desto schneller ist sie gefüllt, und je größere Schlucke wir entnehmen, desto schneller ist sie leer. Auch hier findet keine Parallelität statt.

Wenn es um Erlernen von intuitivem Wissen geht, müssen wir einerseits Situationen ausgesetzt sein, und andererseits müssen wir diesen Situationen in freundlichen Lernumgebungen begegnen (Hogarth, 2001). Situationen mit einer Flussgröße und

Bestand sind wir im Prinzip andauernd ausgesetzt. Hierzu könnte sehr wohl intuitives Verständnis vorliegen. Bis zum Zeitpunkt dieser Studien war mir nur ein Experiment bekannt, in dem nur eine Flussgröße gegeben war, aufgrund derer die Versuchspersonen den Bestand beurteilen sollten. Dafür ergab sich hier ein Anteil von 64% an Korrektlösenden (Ossimitz, 2002). Die Beschreibung des Zuflusses lag als Liniendiagramm vor. Das Liniendiagramm schien nicht für alle Versuchspersonen leicht verständlich gewesen zu sein, was sich in ihren Erklärungen zum Verlauf zeigte. Damit ist zumindest nicht widerlegbar, dass intuitives Wissen zu einem BF-System mit nur einer Flussgröße existiert. Weitere Studien könnten untersuchen, wie gut tatsächlich das Verständnis für ein BF-System mit nur einer Flussgröße ist und welche BF-Beschreibungsdimensionen auf das Verständnis wirken. In diesem Rahmen wäre es auch möglich zu untersuchen, ob BF-Systeme mit zwei Flussgrößen, jedoch ohne Parallelität in der Darstellung, verständlich sind (wie auf einem Kontoauszug).

Die nächste Frage, ob intuitives Wissen zu BF-Systemen mit zwei Flussgrößen existiert, scheint mir hingegen unklarer. Einerseits existieren auch im Alltag Situationen, in denen zwei Flussgrößen auf einen Bestand wirken, wobei alles zu beobachten ist. Das sind Voraussetzungen, die gegeben sein müssen, um diesbezüglich Intuition zu erlernen (Hogarth, 2001). Diese Situationen

kommen vermutlich eher seltener vor oder sind nicht leicht nachzuvollziehen. Ein einfaches Beispiel könnte sein, dass jemand mit dem Bus fährt und an jeder Haltestelle beobachtet, wie viele Personen aus- und einsteigen. In Abhängigkeit davon ändert sich die Anzahl der Personen im Bus. Es besteht Parallelität wie in den bisherigen Studien zu BF-Systemen, und alle Größen sind beobachtbar, wodurch Intuitionen erlernt werden könnten. Die Frage ist, ob dafür etwas Derartiges oft genug passiert, und dann tatsächlich beobachtet wird. Weitere Studien könnten also untersuchen, ob diese Intuitionen überhaupt allgemein vorhanden sind. Formal erlerntes derartiges Wissen scheint eher nicht vorzuliegen (Booth Sweeney & Sterman, 2000). Sinnvollerweise sollten Schulen und Universitäten BF-Systeme und ihre Eigenschaften behandeln, um BF-Verständnis zu fördern. Dann könnten auch Intuitionen im Sinne von mathematischen Intuitionen entstehen (vgl. Kapitel 1.2.2).

6.5 Implikation für die Kommunikation von BF-Systemen

Kosslyn (1994) bezeichnete einen Graphen als „visual display“, das eine oder mehrere Beziehungen zwischen Zahlen zeigt. Seiner Meinung nach ist er erfolgreich gestaltet, wenn das relevante Muster, der Trend oder Vergleich sofort ersichtlich ist.

Bei BF-Systemen ist es zusätzlich erforderlich, von den Flussgrößen auf den Bestand zu schließen. Das erfolgt äußerst selten auf korrekte Art, wenn Liniendiagramme verwendet werden. Diese sollten also im Zusammenhang mit BF-Systemen vermieden werden (außer im Falle von Training, wenn das Ziel ist, dass diese verständlich werden).

Die vorliegende Arbeit zeigt, dass eine Änderung des Flussdiagramms helfen kann, wobei es auf die Details ankommt. Die symbolische Darstellung half in Kombination mit Berechnen des Bestandes zu den einzelnen Zeitpunkten. Sie half weniger stark für einen Gesamtüberblick. Sollen sich Personen selber den Bestand über einen Zeitraum erschließen, ist die symbolische Darstellung kombiniert mit schrittweisem Vorgehen hilfreich. Sie zu bearbeiten, ist jedoch sehr zeitaufwändig. Für eine Darstellung in Medien, beispielsweise in Zeitungen oder ähnlichen Formaten, ist eine derartige symbolische Darstellung wohl ungünstig. Es ist unwahrscheinlich, dass die LeserInnen die einzelnen Zeitpunkte berechnen wie in Kapitel 4. Weitere Studien könnten untersuchen, ob es auf andere Art möglich ist, die symbolische Darstellung wieder zurück in Richtung der konventionalisierten Darstellungen zu entwickeln und zu prüfen, ob das verbesserte Verständnis dann erhalten bleibt.

Nach aktuellem Stand ist es am günstigsten, Flussgrößen und die Bestandsgröße anzugeben, wenn die RezipientInnen den Verlauf all dieser Größen wissen sollten. Eine Studie zeigte, dass Versuchspersonen in derartigem Fall sowohl Flussgrößen als auch die Bestandsgröße korrekt einschätzen (Allenstein, 2013). Auch Meadows et al. (1992) konzipierten Diagramme zu Bestands- und Flussgrößen derart, dass alle Größen explizit gezeigt wurden.

6.6 Fazit

Wiederum bestätigte sich das Bild, dass BF-Größen schwierig zu verstehen sind. Es scheint, dass intuitives Verständnis für BF-Systeme mit mindestens zwei Flussgrößen kaum vorhanden oder bisher nicht zu triggern ist. Eine Ausnahme stellte die symbolische Darstellung in Kombination mit schrittweisem Vorgehen dar. Da BF-Verständnis für unseren Alltag relevant und zusätzlich sehr wichtig ist, um komplexe Systeme wie das Klima zu verstehen, sollte BF-Verständnis in Schule und anderen Bildungseinrichtungen gefördert werden – neben weiteren Facetten, die für das Verständnis komplexer dynamischer Systeme relevant sind. Hierfür könnte die symbolische Darstellung hilfreich sein. Somit eröffnet sich hier ein großes Feld an weiteren und neuen Fragen, beispielsweise: Wie kann das BF-Verständnis gefördert und ge-

schult werden? Bis es hierzu Antworten gibt lautet die sich aus den vorliegenden Studien ergebende Botschaft: In Medien ist es sinnvoll Flussgrößen *und* Bestand zu zeigen. Für bestmögliches, selbständiges Lösen von Aufgaben zu BF-Systemen ist momentan die symbolische Darstellung die günstigste.

Referenzen

Allenstein (2013). *Stock-Flow Aufgaben verstehen. Der Einfluss von zusätzlicher Information, Geschlecht und Geschlechterrollen.* (Bachelorarbeit) Technische Universität Chemnitz, Chemnitz.

[Homepage des MIT]. (o.D.). http://mitsloan.mit.edu/faculty-and-research/academic-groups/system-dynamics/ (letztmalig abgerufen 15.09.2017)

[Download von Bilddokumenten]. (o.D.). https://pixabay.com (letztmalig abgerufen 15.09.2017)

Armenia, S., Onori, R., & Bertini, A. (2004). *Bathtub Dynamics At The "Tor Vergata" University In Rome Italy.* Paper presented at the Proceedings of the 22nd International System Dynamics Conference, Oxford, England, UK.

Bar-Hillel, M. (1973). On the Subjective Probability of Compound Events. *Organizational Behavior & Human Performance, 9*(3), 396-406.

Beermann, L., Heller, K., & Menacher, P. (1992). Mathe: nichts für Mädchen? Begabung und Geschlecht am Beispiel von Mathematik, Naturwissenschaften und Technik: Bern: Huber.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology, 42*(2), 155-162.

Ben-Zeev, T. (1996). When erroneous mathematical thinking is just as "correct": The oxymoron of rational errors. *The nature of mathematical thinking*, 55-79.

Ben-Zeev, T., & Star, J. (2001). Intuitive Mathematics: Theoretical and Educational Implications. In B. Torff & R. Sternberg (Eds.), *Understanding and Teaching the Intuitive Mind* (pp. 29-56). Mahwah: LEA Books.

Booth Sweeney, L., & Sterman, J. D. (2000). Bathtub dynamics: initial results of a systems thinking inventory. *System Dynamics Review, 16*(4), 249-286.

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica, 81*(3), 211-241.

Broadbent, D. E. (1977). Levels, hierarchies, and the locus of control. *The Quarterly Journal of Experimental Psychology, 29*(2), 181-201.

Brunstein, A., Gonzalez, C., & Kanter, S. (2010). Effects of domain experience in the stock-flow failure. *System Dynamics Review (Wiley), 26*(4), 347-354.

Buchner, A. (1995). Basic topics and approaches to the study of complex problem solving. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 27-63). Hillsdale, NJ: Erlbaum.

Budde, J. (2009). Mathematikunterricht und Geschlecht. Empirische Ergebnisse und pädagogische Ansätze. In B. f. B. u. Forschung (Eds.), Bildungsforschung Band 3. https://www.ph-freiburg.de/fileadmin/dateien/sonstige/gleichstellung/Budde_Mathematik_Geschlecht.pdf (abgerufen 29.08.2016).

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31*(3), 187-276.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58*(1), 1-73.

Cronin, M. A., & Gonzalez, C. (2007). Understanding the building blocks of dynamic systems. *System Dynamics Review, 23*(1), 1-17.

Cronin, M. A., Gonzalez, C., & Sterman, J. D. (2009). Why Don't Well-Educated Adults Understand Accumulation? A Challenge to Researchers, Educators, and Citizens. *Organizational Behavior and Human Decision Processes, 108*(1), 116-130.

Domagk, S. (2008). Bilder. In Niegemann, H., Domagk, S., Hessel, S., Hein, A., Hupfer, M. & Zobel, A. *Kompendium multimediales Lernen.* (pp. 207- 238). Berlin: Springer.

Dörner, D. (1981). Über die Schwierigkeiten menschlichen Umgangs mit Komplexität. *Psychologische Rundschau, 32*, 163-179.

Dörner, D. (2008). *Die Logik des Misslingens: Strategisches Denken in komplexen Situationen*: Rowohlt Verlag GmbH.

Dörner, D.; Strohschneider, Stefan (1999). Komplexes Problemlösen – Königsweg der Theoretischen Psychologie? *Psychologische Rundschau, 50*(4), 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, prospects and limits* (pp. 218–239). Cambridge University Press.

Dwyer, F. M. (1975). On Visualized Instruction Effect of Students' Entering Behavior. *The Journal of Experimental Education, 43*(3), 78-83.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods, 39*(2), 175-191.

Fischbein, E. (1994). The interaction between the formal, the algorithmic and the intuitive components in a mathematical activity. *Didactics of mathematics as a scientific discipline*, 231-245.

Fischer, H., Degen, C., & Funke, J. (2015). Improving stock-flow reasoning with verbal formats. *Simulation & Gaming, 46*(3-4), 255-269.

Forrester, J. W. (1968). Industrial dynamics-after the first decade. *Management Science, 14*(7), 398-415.

Forrester, J. W. (1989). *The beginning of system dynamics.* Paper presented at the Banquet Talk at the international meeting of the System Dynamics Society, Stuttgart, Germany.

Forrester, J. W. (2003). Dynamic models of economic systems and industrial organizations. *System Dynamics Review, 19*(4), 329-345.

Funke, J. (2003). *Problemlösendes Denken* (1. Aufl. ed.). Stuttgart: Kohlhammer.

Gick, M. L. (1986). Problem-solving strategies. *Educational psychologist, 21*(1-2), 99-120.

Gigerenzer, G., Hertwig, R., Hoffrage, U., & Sedlmeier, P. (2008). Cognitive illusions reconsidered. *Handbook of experimental economics results, 1*, 1018-1034.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological review, 102*(4), 684.

Gonzalez, C., Vanyukov, P., & Martin, M. K. (2005). The use of microworlds to study dynamic decision making. *Computers in Human Behavior, 21*(2), 273-286.

Goodman, S. (2008). *A dirty dozen: twelve p-value misconceptions.* Paper presented at the Seminars in hematology.

Greeno, J. G. (1979). *Some examples of cognitive task analysis with instructional implications*: DTIC Document. In R. E. Snow, P-A Federico, & W. E. Montague (Eds.), *Aptitude, learning, and instruction: Vol. 2. Cognitive process analyses of learning and problem solving* (pp. 1-21). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hammond, K. R. (1988). *Judgement and decision making in dynamic tasks*: Information and Decision Technologies, 3–14.

Haroz, S., Kosara, R., & Franconeri, S. L. (2015). *Isotype visualization: Working memory, performance, and engagement with pictographs.* Paper presented at the Proceedings of the 33rd annual ACM conference on human factors in computing systems.

Hogarth, R. M. (2001). *Educating intuition*: University of Chicago Press.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*(3), 430-454.

Kainz, D., & Ossimitz, G. (2002). *Can students learn stock-flow-thinking? An empirical investigation.* Paper presented at the Proceedings of the 20th international conference of the system dynamics society, Palermo, Italy.

Kalyuga, S. (2005). Prior knowledge principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 325-337). Cambridge University Press.

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology, 92*(1), 126.

Kaplan, C. A., & Simon, H. A. (1990). In search of insight. *Cognitive Psychology, 22*(3), 374-419.

Kapmeier, F. (2004). *Findings from four years of bathtub dynamics at higher management education institutions in Stuttgart.* Paper presented at the 22nd International Conference of the System Dynamics Society, Oxford, England, UK.

Kapmeier, F., Happach, R. M., & Tilebein, M. (2014). *Bathtub dynamics revisited: does educational background matter?* Paper presented at the Proceedings of the 32nd International Conference of the System Dynamics Society, Delft, Netherlands.

Kasperidus, H. D., Langfelder, H., & Biber, P. (2006, 23.-27.7.2006). *Comparing systems thinking inventory task*

performance in German classrooms at high school and university level. Paper presented at the 24th International conference of the System Dynamics Society, Nijmegen, Netherlands.

Kimura, D. (1999). *Sex and cognition*. Cambridge MA: MIT Press.

Kosslyn, S. M. (1994). *Elements of graph design*. New York NY: Freeman.

Levie, W. H. L., Richard (1982). Effects of Text Illustrations: A Review of Research. *Educational Communication and Technology: A Journal of Theory, Research, and Development, 30 (4)*, 195-232.

Levin, J. R., Anglin, G. J., & Carney, R. N. (1987). On empirically validating functions of pictures in prose. *The psychology of illustration, 1*, 51-85.

Lewis, M. W., & Anderson, J. R. (1985). Discrimination of operator schemata in problem solving: Learning from examples. *Cognitive Psychology, 17*(1), 26-65.

Lungwitz, V., Sedlmeier, P., & Schwarz, M. (2018). Can gender priming eliminate the effects of stereotype threat? The case of simple dynamic systems. *Acta psychologica, 188*, 65-73.

MacDonald-Ross, M. (1977). How Numbers Are Shown: A Review of Research on the Presentation of Quantitative Data in Texts. *Audiovisual Communication Review, 25*(4), 359-409.

Mayer, R. E. (2009). *Multimedia learning* (2. ed.). Cambridge University Press.

Meadows, D. H., Meadows, D. L., Zahn, E., & Milling, P. (1972). *Die Grenzen des Wachstums. Bericht des Club of Rome zur Lage der Menschheit.* Stuttgart: Deutsche Verlags-Anstalt.

Meadows, D. H., Meadows, D. L., & Randers, J. (1992). *Die neuen Grenzen des Wachstums: die Lage der Menschheit: Bedrohung und Zukunftschancen*. (6. Auflage) Stuttgart: Deutsche Verlags-Anstalt.

Nadolski, R. J., Kirschner, P. A., & Merriënboer, J. J. (2005). Optimizing the number of steps in learning tasks for complex skills. *British Journal of Educational Psychology, 75*(2), 223-237.

Nadolski, R. J., Kirschner, P. A., van Merriënboer, J. J., & Hummel, H. G. (2001). A model for optimizing step size of learning tasks in competency-based multimedia practicals. *Educational Technology Research and Development, 49*(3), 87-101.

Neurath (1936). International Picture Language. https://monoskop.org/images/e/ec/Neurath_Otto_International_Picture_Language.pdf (abgerufen 26.6.2017).

Niegemann, H. M. (2008). Bilder. In Niegemann, H. M., Domagk, S., Hessel, S., Hein, A., Hupfer, M., & Zobel, A. (Eds.), *Kompendium multimediales Lernen* (pp. 207-238). Springer-Verlag.

Ossimitz, G. (2001). Bestandsgrößen und Flussgrößen. Manuskript, Universität Klagenfurt. http://www.uni-klu.ac.at/~gossimit/pap/bestflus.pdf (abgerufen 18.09.2008).

Ossimitz, G. (2002). *Stock-Flow-Thinking and Reading stock-flow-related Graphs: An Empirical Investigation in Dynamic Thinking Abilities*. Paper presented at the 20th International Conference of the System Dynamics Society, Palermo, Italy.

Ossimitz, G., Kotzent-Pietschnig, U., Kreisler, B., Waiguny, M., & Zoltan, M. (2001). Unterscheidung von Bestands- und Flussgrößen. https://campus.aau.at/cris/publication/0f4dc44845ae29440145b2049e7e7d37;jsessionid=FF4D82F3A2131F01B72F297070677092.app-campus1 (abgerufen 18.9.2008).

Pala, Ö., & Vennix, J. A. M. (2005). Effect of system dynamics education on systems thinking inventory task performance. *System Dynamics Review, 21*(2), 147-172.

Phuah, T. (2010, July 25 – 29, 2010). *Can people learn behaviours of stock and flow using their ability to calculate running total? An experimental study.* Paper presented at the 28th International Conference of the System Dynamics Society, Seoul, Korea.

Pinker, S. (1990). A theory of graph comprehension. *Artificial intelligence and the future of testing*, 73-126.

Plessner, H. E., Betsch, C. E., & Betsch, T. E. (Eds.) (2008). *Intuition in judgment and decision making*. Psychology Press.

Reed, S. K. (1999). *Word problems: Research and curriculum reform*: Lawrence Erlbaum Associates Mahwah, NJ.

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology, 38*(2), 194-201.

Schneider-Düker, M., & Kohler, A. (1988). Die Erfassung von Geschlechtsrollen: Ergebnisse zur deutschen Neukonstruktion des Bem Sex-Role Inventory. *Diagnostica 34,* 256-270.

Schnotz, W. (2005). Integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*. Cambridge University Press.

Schwarz, M. A. (2016). *Vom stockenden Verständnis fließender Zusammenhänge. Darstellungs- und personenbezogene Einflussfaktoren auf das basale Verständnis einfacher dynamischer Systeme.* (Dissertation) http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-203692 (abgerufen 10.2016).

Sedlmeier, P. (2007). Statistical reasoning: valid intuitions put to use. In M. Lovett. & P. Shah (Eds.), *Thinking with data.* (pp. 389-419). New York: Lawrence Erlbaum Asssociates.

Sedlmeier, P., & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General, 130*(3), 380-400.

Sedlmeier, P., & Hilton, D. J. (2011). Improving judgment and decision making through communication and representation. In M. K. Dhami, A. Schlottmann & M. R. Waldmann (Eds.), *Judgment and Decision Making as a*

Skill: Learning, Development and Evolution (pp. 229-258). Cambridge University Press.

Sedlmeier, P., & Renkewitz, F. (2013). *Forschungsmethoden und Statistik: Ein Lehrbuch für Psychologen und Sozialwissenschaftler (2. überarbeitete und erweiterte Auflage).* München, Deutschland: Pearson.

Selkow, P. (1985). Male/female differences in mathematical ability: A function of biological sex or perceived gender role? *Psychological Reports, 57*(2), 551-557.

Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational psychology review, 14*(1), 47-69.

Shah, P., Mayer, R. E., & Hegarty, M. (1999). Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. *Journal of Educational Psychology, 91*(4), 690.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype Susceptibility: Identity Salience and Shifts in Quantitative Performance. *Psychological Science, 10*(1), 80-83.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype Threat and Women's Math Performance. *Journal of Experimental Social Psychology, 35*(1), 4-28.

Sterman, J. D. (1989). Misperceptions of feedback in dynamic decision making. *Organizational Behavior and Human Decision Processes, 43*(3), 301-335.

Sterman, J. D. (1994). Learning in and about complex systems. *System Dynamics Review, 10*(2-3), 291-330.

Sterman, J. D. (2002). All Models are Wrong: Reflections on Becoming a Systems Scientist. *System Dynamics Review, 18*, 501-531.

Sterman, J. D. (2010). Does formal system dynamics training improve people's understanding of accumulation? *System Dynamics Review (Wiley), 26*(4), 316-334.

Sterman, J. D., & Booth Sweeney, L. (2002). Cloudy skies: assessing public understanding of global warming. *System Dynamics Review (Wiley), 18*(2), 207-240.

Strohhecker, J. (2009, July 26 – 30, 2009). *Does a Better Understanding of Accumulation Indeed Predict a Higher Performance in Stock and Flow Management?* Paper presented at the 27th International Conference of the System Dynamics Society, Albuquerque, New Mexico, USA.

Tarmizi, R. A., & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology, 80*(4), 424.

Troche, S., & Rammsayer, T. H. (2011). Eine Revision des deutschsprachigen Bem Sex-Role Inventory. *Klinische Diagnostik und Evaluation, 4*, 262-283.

Van Merrienboer, J. J., & Ayres, P. (2005). Research on cognitive load theory and its design implications for e-learning. *Educational Technology Research and Development, 53*(3), 5-13.

Wagenaar, W. A., & Timmers, H. (1979). The pond-and-duckweed problem: Three experiments on the misperception of exponential growth. *Acta Psychologica, 43*(3), 239-251.

Wason, P. C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *The Quarterly Journal of Experimental Psychology, 23*(1), 63-71.

Weidenmann, B. (1994). *Lernen mit Bildmedien: Psychologische und didaktische Grundlagen* (2., neu ausgestattete Auflage). Weinheim: Beltz Verlag.

Weidenmann, B. (2006). Lernen mit Medien. In A. Krapp & B. Weidenmann (Eds.), *Pädagogische Psychologie: Ein Lehrbuch* (5., vollständig überarbeitete Auflage, pp. 424-476). Weinheim: Beltz Verlag.

Zelazny, G. (1989). *Wie aus Zahlen Bilder werden: Wirtschaftsdaten überzeugend präsentiert* (2. Auflage). Wiesbaden: Gabler.

Anhang

Anhang A

Kriterien für die Aufgaben mit Rechteck- und Dreieckverlauf genutzt von Booth Sweeney und Sterman (2000)

Rechteckverlauf	***Dreieckverlauf***
Wenn Zufluss über Abfluss liegt, nimmt der Bestand zu *[1]*	
Wenn Abfluss über Zufluss liegt, nimmt der Bestand ab *[2]*	
Die Höchst- und Tiefstpunkte des Bestandes treten auf, wenn der Nettofluss gleich 0 ist:	
(t= 4, 8, 12, 16) *[3]*	(t = 2,6,10,14) *[3]*
Der Graph für den Bestand sollte keine diskontinuierlichen Sprünge enthalten (ist kontinuierlich) *[4]*	
In jedem Zeitabschnitt ist der Nettofluss konstant, der Bestand muss linear fallen oder steigen *[5]*	In jedem Zeitabschnitt ist Steigung des Bestandes gleich der Nettorate also a) Wenn Nettofluss positiv und fallend, nimmt Bestand mit abnehmender Geschwindigkeit/Rate zu (0<t<= 2; 8 < t =< 10) b) Wenn Nettofluss negativ und fallend, nimmt Bestand mit zunehmender Rate ab(2<t<=4; 10<t=<12) c) Wenn Nettofluss negativ und zunehmend, nimmt Bestand mit

	abnehmender Rate ab (4< t <= 6; 12 < t <=14) d) Wenn Nettofluss positiv und zunehmend, nimmt Bestand mit zunehmender Rate zu (6 < t =< 8; 14<= 16) *[5]*
Steigung von Bestand während jeden Zeitabschnittes ist Nettorate (d.h. +/- 25 Einheiten pro Minute) *[6]*	Steigung von Bestand wenn Nettorate am höchsten, ist 50 Einheiten pro Zeiteinheit (min) (bei t = 0, 8, 16) *[6]*
	Steigung von Bestand wenn Nettorate am niedrigsten, ist -50 Einheiten pro Zeiteinheit (min) (bei t = 4, 12) *[7]*
Die Menge, die während jedes Zeitabschnittes zum Bestand dazu kommt bzw. weggeht, ist die Fläche, die von der Nettorate eingeschlossen wird (z.B. 25 Einheiten pro Zeiteinheit * 4 Zeiteinheiten = 100 Einheiten), also ist der Höchststand des Bestandes bei 200 Einheiten erreicht und fällt bis zu einem Minimum von 100 Einheiten *[7]*	Die Menge, die während jedes Zeitabschnittes von 2 Zeiteinheiten zum Bestand dazu kommt bzw. weggeht, ist gleich der Fläche, die von der Nettorate eingeschlossen wird (d.h. ein Dreieck mit Fläche (1/2)*50 Einheiten pro Minute * 2 Minuten = +/- 50 Einheiten), also ist der Höchststand vom Bestand bei 150 Einheiten erreicht und fällt bis zu einem Minimum von 50 Einheiten *[8]*

Anhang B

Beispielablauf einer Aufgabe in Studie 1: Rechteckverlauf mit Flussdiagramm und mit schematischer Abbildung für die Wasser-Bedingung.

Stellen Sie sich ein Becken mit Wasser vor, in das von oben weiterhin Wasser hinein fließt. Unten an dem Becken gibt es einen Wasserhahn, aus dem Wasser gleichzeitig wieder heraus fließt (vgl. folgende Abbildung). Sie beobachten dieses Becken 16 Minuten lang.

Das Becken enthält 1000 Liter Wasser bevor die Zeitspanne, die Sie beobachten, beginnt.

Die Menge Wasser, die pro Minute unten heraus fließt, ist immer dieselbe. Über den gesamten Zeitraum fließen konstant 500 Liter pro Minute aus dem Becken.

Die Menge Wasser, die von oben hinein fließt, ändert sich über die Zeit hinweg (siehe auch folgendes Diagramm). Die ersten vier Minuten fließen kontinuierlich 750 Liter Wasser pro Minute hinzu. Daraufhin fließen die nächsten vier Minuten jeweils 250 Liter Wasser pro Minute hinzu. Dieses Muster wiederholt sich noch einmal: Vier Minuten fließen wieder 750 Liter Wasser pro Minute dazu und die letzten vier Minuten wiederum 250 Liter Wasser pro Minute.

Überlegen Sie bitte, wie sich das Verhalten des hinein- und heraus fließenden Wassers auf den Inhalt des Beckens auswirkt. Bitte zeichnen Sie nun in das leere Diagramm ein, wie sich der Inhalt des Beckens über den gesamten Zeitraum hinweg verändert. Beschriften Sie hierbei auch die Y-Achse mit den entsprechenden Werten.

Für die Lösung sind die exakten Werte jedoch nicht so wichtig, es kommt vor allem auf das Muster an.

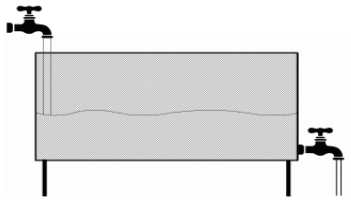


Abbildung 1: Becken (schematische Darstellung)

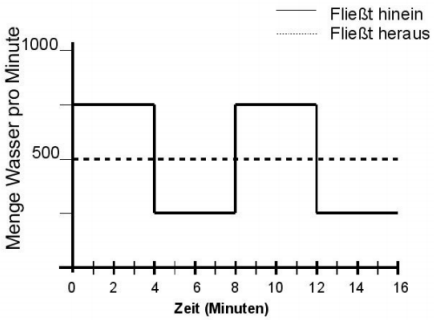


Abbildung 2: Diagramm

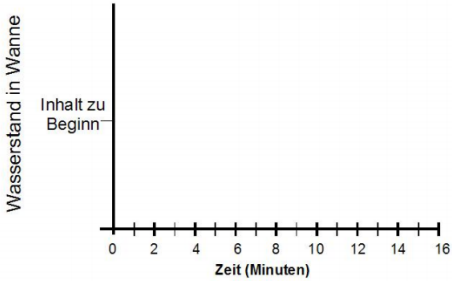


Abbildung 3: Lösung

Anhang C

Die 16 Versionen des Aufgabenmaterials aus Studie 1, denen die Versuchspersonen per Blockrandomisierung zugeordnet wurden.

Version	BF-Größe an erster Position	Verlauf an erster Position	Schematische Abbildung	Fluss-diagramm
1	Wasser	Rechteck	ohne	ohne
2	Wasser	Rechteck	ohne	mit
3	Wasser	Rechteck	mit	ohne
4	Wasser	Rechteck	mit	mit
5	Wasser	Dreieck	ohne	ohne
6	Wasser	Dreieck	ohne	mit
7	Wasser	Dreieck	mit	ohne
8	Wasser	Dreieck	mit	mit
9	Murmeln	Rechteck	ohne	ohne
10	Murmeln	Rechteck	ohne	mit
11	Murmeln	Rechteck	mit	ohne
12	Murmeln	Rechteck	mit	mit
13	Murmeln	Dreieck	ohne	ohne
14	Murmeln	Dreieck	ohne	mit
15	Murmeln	Dreieck	mit	ohne
16	Murmeln	Dreieck	mit	mit

Anmerkung: Jede Versuchsperson bearbeitete zwei Aufgaben, eine mit Rechteck-, eine mit Dreieckverlauf, das jeweils kombiniert mit den beiden verschiedenen BF-Größen. In der Tabelle sind für die BF-Größe und den Verlauf jeweils die Variante an erster Position berichtet. Beispielsweise erhält eine Versuchsperson in der Version 14 zunächst eine Aufgabe mit Dreieckverlauf und Murmeln als Fluss- und Bestandsgrößen, die zweite Aufgabe ist für sie der Rechteckverlauf mit der BF-Größe Wasser; in beiden Fällen beinhaltet ihr Aufgabenmaterial keine schematische Abbildung aber ein Flussdiagramm liegt vor.

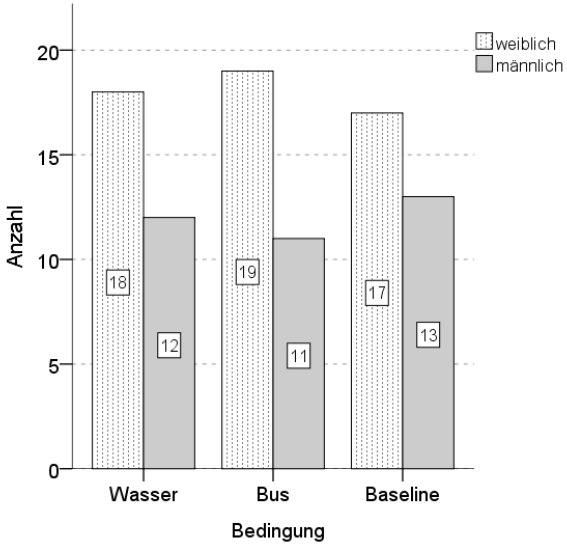
Anhang D

Die Tabelle zeigt die Fächer, die die Versuchspersonen aus Studie 3 studierten, die Einordnung zu naturwissenschaftlich nah oder fern und wie viele Personen das jeweilige Fach studierten.

Studienfach	Naturwissenschaftlich	Anzahl
Automobilproduktion	nah	1
Elektrotechnik	fern	3
Europäische Geschichte	nah	1
Europastudien	fern	2
Finanzmathematik	nah	2
Fremdsprachen in Erwachsenenbildung	fern	2
Germanistik	fern	5
Informatik	nah	1
Informationstechnik	nah	1
Mathematik	nah	1
Media Production	nah	1
Medienkommunikation	fern	1
Mikrotechnik/ Mechatronik	nah	6
Pädagogik	fern	3
Physik	nah	4
Politikwissenschaft	fern	5
Psychologie	fern	23
Sensorik und Kognitive Psychologie	nah	2
Soziologie	fern	6
Sport / BWL	fern	1
Sports-Engineering	fern	1
Sportwissenschaft	fern	1
Wirtschaftsinformatik	nah	1
Wirtschaftsingenieurwesen	nah	7
Wirtschaftswissenschaften	fern	9

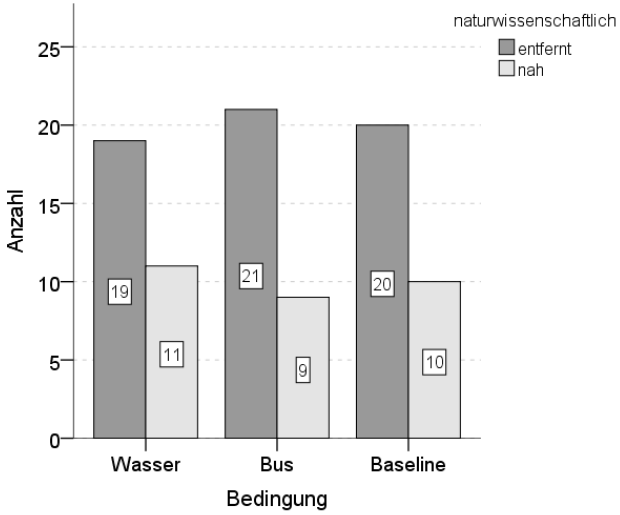
Anhang E

Verteilung der Variable Geschlecht in den verschiedenen Bedingungen der Studie 3.



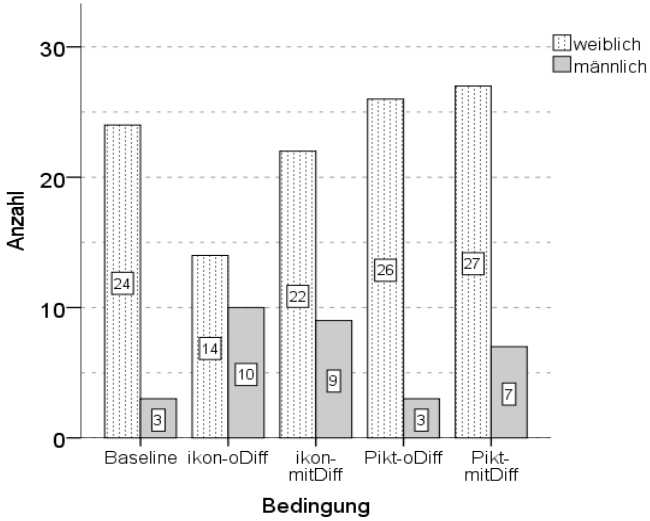
Anhang F

Verteilung der Variable Studienfach in den verschiedenen Bedingungen der Studie 3.



Anhang G

Verteilung von Männern und Frauen je nach Bedingung in Studie 4



Anmerkung: ikon-oDiff: ikonisch-ohne-Differenz; ikon-mitDiff: ikonisch-mit-Differenz; Pikt-oDiff: Piktogramm-ohne-Differenz; Pikt-mitDiff Piktogramm-mit-Differenz

Anhang H

Regressionsanalyse für das Modell ohne die Variablen Geschlecht und Studienfach mit Gesamtsumme als abhängige Variable in Studie 4.

Prädiktor	β	*t*	*p*	*pr*
Baseline	-.21	-2.209	.02	-.19
ikonisch	.04	.406	.35	.04
Differenz	-.10	-1.065	.15	-.09
Mathem.-Note	.27	3.229	<.00	.27
Femininitäts-Skala	-.21	-2.481	.01	-.21
Maskulinitäts-Skala	-.04	-.544	.59	-.05
Interaktion BSRI-R	-.04	-.450	.65	-.04
Teilnahme	.11	1.381	.09	.11

Anmerkungen: Die Analyse ohne die Werte der Ausreißer-Versuchsperson ergab $R^2 = .19$; mit $F(8,130) = 3.854$; $p < .00$. Dummy-kodierte Variablen sind folgendermaßen kodiert:

Baseline 1, symbolische Bedingungen 0; ikonische Zeichen 1, alle anderen Bedingungen 0; Differenz: mit 1, ohne 0; Teilnahme:vorherige Teilnahme 1, bisher noch nicht teilgenommen 0. Alle p-werte wurden halbiert, mit Ausnahme der Werte für die Interaktion und für die BSRI-R-Skalen.