
DER EINFLUSS VISUELLER GESTALTUNGSMERKMALE
AUF KOGNITIVE PROZESSE BEI DER BEARBEITUNG VON
AUFGABEN MIT DIAGRAMMEN

Dissertation

zur Erlangung des Doktorgrades der Philosophischen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Dipl.-Psych. Benjamin Strobel

Kiel,

06. Februar 2018

Erstgutachter:

Prof. Dr. Olaf Köller

Zweitgutachterin:

Prof. Dr. Friederike Zimmermann

Tag der mündlichen Prüfung:

14. Mai 2018

Durch den Prodekan für Studium und Lehre,
Prof. Dr. Elmar Eggert, zum Druck genehmigt:

14. Mai 2018

Danksagung

Großer Dank gilt Prof. Dr. Olaf Köller für sein Vertrauen in meine Vorhaben und seine großzügige Unterstützung. Ich bin ausgesprochen dankbar für die Freiheiten und Entwicklungsmöglichkeiten, die ich am IPN genießen durfte. Die hervorragenden finanziellen und strukturellen Möglichkeiten sowie die herausragenden Bildungsgelegenheiten am IPN waren von unschätzbarem Wert für mich und meine Arbeit.

Ich danke meinen Kolleginnen und Kollegen, die mich durch die Promotionszeit am IPN begleitet haben. Besonders danken möchte ich Dr. Steffani Saß für ihre Unterstützung, Christian Schöber und Melike Ömeroğulları für eine gute Stimmung im Büro, Simon Grund für seine kompetente Beratung in methodischen Fragen sowie Gráinne Newcombe und Cornelia Gerigk für ihre hilf- und lehrreichen Sprachkorrekturen. Mein Dank geht auch an Dr. Marc Eckhardt für seine fröhlichen Grüße, die mir stets gute Laune bereitet haben.

Mein größter Dank gilt meiner geschätzten Kollegin, Mentorin und Freundin Dr. Marlit Annalena Lindner, die mich während meiner gesamten Promotionszeit begleitet und unterstützt hat. Marlit, du bist eine herausragende Wissenschaftlerin und alles Wichtige, das ich für diese Arbeit gelernt habe, weiß ich von dir. Ich danke dir für deine Unterstützung, dein Engagement und deine Hilfe zu allen Tages- und Nachtzeiten. Ohne dich wäre ich nicht hier.

Ein herzlicher Dank geht auch an unsere studentischen Hilfskräfte Charlotte Quednau, Milan Frank, Torben Janssen, Julia Barenthien und unserer Praktikantin Nathalie Möhle für ihre unersetzliche Hilfe bei den Datenerhebungen.

Ich danke meinen Freunden, meiner Familie und dir, liebe Merle, für eure Unterstützung in den letzten Jahren. Es war nicht immer leicht und ich schätze es sehr, dass ihr mich in den Höhen und Tiefen meiner Promotionszeit begleitet habt. Ganz besonders danke ich meinem guten Freund Julius, der immer für mich da war. Lieber Julius, dein offenes Ohr wie auch dein Rat waren von unschätzbarem Wert für mich und werden es immer sein.

Inhaltsverzeichnis

Kapitel 1 - Einleitung und theoretischer Rahmen	9
1.1 Was ist ein Diagramm?	11
1.1.1 Klassifikation von Repräsentationen	12
1.1.1.1 Intrinsische und extrinsische Repräsentationen	14
1.1.1.2 Realistische und logische Bilder	16
1.1.2 Arbeitsdefinition für Diagramme	17
1.1.3 Diagrammentypen	18
1.1.3.1 Streudiagramm	18
1.1.3.2 Liniendiagramm	18
1.1.3.3 Balken- und Säulendiagramm	19
1.1.3.4 Gestapeltes Säulendiagramm	19
1.1.3.5 Kreisdiagramm	21
1.1.4 Abgrenzung zu Tabellen und Texten	21
1.2 Modelle des Diagrammverstehens	22
1.2.1 Taxonomien für Diagrammleseaufgaben	23
1.2.2 Allgemeine Prozessmodelle des Diagrammverstehens	26
1.2.2.1 Aufgabenanalyse nach Bertin (1983)	26
1.2.2.2 Pinkers (1990) Theorie des Diagrammverstehens	27
1.2.2.3 Interaktionsmodell des Diagrammverstehens nach Freedman & Shah (2002)	31
1.2.3 Cognitive Load Theory	33
1.2.3.1 Intrinsische kognitive Belastungen bei Diagrammen	34
1.2.3.2 Extrinsische kognitive Belastungen bei Diagrammen	34
1.2.3.3 Lernbezogene kognitive Belastungen bei Diagrammen	35
1.3 Computationale Eigenschaften von Diagrammen	36
1.3.1 Informationale und Computationale Äquivalenz	36
1.3.2 Visuelle Merkmale von Diagrammen	39
1.3.2.1 Dimensionalität	39
1.3.2.2 Farbe	40
1.3.2.3 Legenden und Beschriftung	41
1.3.2.4 Größe und Skalierung	42
1.3.2.5 Datenkomplexität	42
1.3.3 Säulen- und Liniendiagramme im Vergleich	43
1.4 Eyetracking als Untersuchungsmethode	47
1.4.1 Eyetracking zur Untersuchung von Diagrammen	48

1.5 Forschungsfragen	50
1.5.1 Forschungsfrage 1: Sind Personen dazu in der Lage, für gegebene Aufgaben dasjenige Diagramm zur Bearbeitung zu nutzen, das für die jeweilige Aufgabe computationale Vorteile bietet?	51
1.5.2 Forschungsfrage 2: Beeinflusst eine hohe Datenkomplexität den Bearbeitungsprozess und die Leistung, wenn zusätzliche Daten vollständig irrelevant für die Bearbeitung einer Aufgabe sind?.....	51
1.5.3 Forschungsfrage 3: Beeinflussen interessante, aber irrelevante Zusatzinhalte (Seductive Details) in Diagrammen den Bearbeitungsprozess und die Leistung von Personen, die Diagrammleseaufgaben bearbeiten?.....	52
1.6 Literatur	54
Kapitel 2 - Do Graph Readers Prefer the Graph Type Most Suited to a Given Task?	65
Abstract	66
2.1 Graph reading tasks	68
2.2 Computational properties of graphs	68
2.3 Graph reading and learning effects.....	69
2.4 Research questions	70
2.5 Method	71
2.5.1 Sample and study design	71
2.5.2 Material and measures	72
2.5.2.1 Graph tasks	72
2.5.2.2 Paper-pencil questionnaire.....	72
2.5.3 Apparatus.....	73
2.5.5 Analysis	74
2.5.5.1 Eye-movement data pre-processing.....	74
2.5.5.2 Linear mixed-effects models.....	75
2.6 Results	76
2.6.1 Overall graph preference	77
2.6.2 Preferential graph processing	79
2.6.3 Preference development.....	79
2.7 Discussion	80
2.7.1 Overall graph preference	81
2.7.2 Preferential graph processing	81
2.7.3 Preference Development.....	83
2.8 Conclusion.....	83
2.9 References	86

Kapitel 3 - Task-irrelevant Data Impair Processing of Graph Reading Tasks.....	93
Abstract	94
3.1 Data complexity	96
3.2 Cognitive load theory	97
3.3 Using eye tracking to gain insight into processing of graph reading tasks	98
3.4 Research hypotheses.....	99
3.5 Method	99
3.5.1 Sample and study design	99
3.5.2 Material and measures.....	100
3.5.2.1 Graph task.....	100
3.5.2.2 Cognitive load measurement.....	101
3.5.2.3 Graph literacy	101
3.5.3 Apparatus.....	101
3.5.4 Procedure.....	103
3.5.5 Analysis	103
3.5.5.1 Eye-Movement Data Pre-Processing	103
3.5.5.2 Linear mixed-effects models.....	103
3.6 Results	105
3.6.1 Error Rate	106
3.6.2 Total Processing Time	106
3.6.3 Cognitive Load	106
3.6.4 Attention Distribution among AOIs	107
3.6.4.1 Item Stem and Options.....	107
3.6.4.2 Axes and Labels.....	107
3.6.4.3 Relevant Data.....	108
3.6.4.4 Irrelevant Data	108
3.7 Discussion	108
3.7.1 Implications for Instruction	110
3.7.2 Implications for Graph Tasks in Tests.....	111
3.7.3 Limitations & Future Research.....	111
3.8 Conclusion.....	112
3.9 References	113
 Kapitel 4 - Seductive Details in the Context of Graph Comprehension	 119
Abstract	119
4.1 The seductive details effect	122

4.2 Seductive details in the context of graph comprehension	123
4.3 Using eye tracking to gain insight into the graph reading process.....	125
4.4 Research hypotheses.....	126
4.5 Method	127
4.5.1 Sample and study design	127
4.5.2 Material and measures.....	127
4.5.2.1 Graph task.....	127
4.5.2.2 Working memory capacity.....	128
4.5.2.3 Graph literacy	129
4.5.3 Apparatus.....	129
4.5.4 Procedure.....	130
4.5.5 Analysis	130
4.5.5.1 Eye-movement data pre-processing.....	130
4.5.5.2 Linear mixed-effects models.....	130
4.5.5.3 Model comparisons and Bayes factors	131
4.6 Results	132
4.6.1 Error rate.....	132
4.6.2 Processing time.....	133
4.6.3 AoI analysis	135
4.7 Discussion	137
4.7.1 The seductive details effect in the context of graph comprehension.....	137
4.7.2 Distraction and disruption	139
4.7.3 The role of working memory capacity.....	139
4.7.4 Limitations & future research.....	140
4.8 Conclusion.....	141
4.9 References	142
Kapitel 5 - Zusammenfassung und Gesamtdiskussion.....	149
5.1 Zusammenfassung Studie 1.....	150
5.2 Zusammenfassung Studie 2.....	152
5.3 Zusammenfassung Studie 3.....	154
5.4 Diskussion und Forschungsausblick.....	158
5.4.1 Multiple informationale äquivalente Diagramme.....	159
5.4.2 Systeminhärente und systemfremde Gestaltungsmerkmale	161
5.4.3 Der Einfluss von Individualfaktoren	163
5.4.4 Implikationen für Lehren und Testen	165

5.4.4.1 Implikationen für die Gestaltung von Diagrammen in Lehrmaterialien.....	165
5.4.4.2 Implikationen für die Gestaltung von Diagrammen in Leistungstests.....	167
5.4.5 Limitationen.....	168
5.4.6 Conclusio.....	169
5.5 Literatur.....	171
Lebenslauf.....	179

Kapitel 1 - Einleitung und theoretischer Rahmen

Aufgrund der hohen Anforderungen der fortschreitenden technischen Entwicklung, der Digitalisierung und der zunehmenden Menge an Informationen im Alltag steigt die Bedeutung einer naturwissenschaftlichen Grundbildung. Bereits Schülerinnen und Schüler müssen wissen, wie sie ihre Fähigkeiten nutzen, kritisch denken, Wissen auf neue Situationen anwenden, Informationen analysieren, neue Ideen kommunizieren, Probleme lösen und Entscheidungen treffen (Glazer, 2011). Eine naturwissenschaftliche Grundbildung soll Schülerinnen und Schülern dabei helfen, sich auf diese Herausforderungen vorzubereiten (Millar, 2006).

Es werden ständig neue Erkenntnisse gewonnen, während die zunehmende Digitalisierung alte und neue Informationen schnell und leicht zugänglich macht. Das Lesen und Verstehen von Daten und zugehörigen Repräsentationsformen sind somit nicht nur notwendiger Bestandteil für die Ausbildung von Naturwissenschaftlern, sondern auch für die gesellschaftliche Teilhabe eines Individuums allgemein (Bybee & Fuchs, 2006; Roberts, 2007). Entsprechend spielen Diagramme nicht nur eine Rolle beim Problemlösen (Baker, Corbett & Koedinger, 2001), beim Lehren und Lernen von Mathematik (Cucuo & Curcio, 2001) und beim Verstehen wissenschaftlicher Daten (OECD, 2016; Shah & Hoeffner, 2002), sondern in fast allen Bereichen des täglichen Lebens – in der Schule, im Internet, in Büchern und Zeitschriften, dem Fernsehen und vielen weiteren Bereichen (z. B. Glazer, 2011; Mayer, 2009; Pereira-Mendoza, Goh, & Bay, 2004; Purchase, 2014). So findet sich das Darstellen, Lesen und Verstehen von Daten beispielsweise in der Rahmenkonzeption von PISA (Programme for International Student Assessment; OECD, 2016) und den Standards des National Council of Teachers and Mathematics (2000) in den USA.

In der Rahmenkonzeption von PISA ist das Lesen und Verstehen von Diagrammen fest im Konzept der *Scientific Literacy* (Wissenschaftliche Grundbildung; Schiepe-Tiska et al., 2016) verankert. Dort heißt es „data interpretation begins with looking for patterns, constructing simple tables and graphical visualisations, such as pie charts, bar graphs, scatterplots or Venn diagrams [...] Scientists make choices about how to represent the data in graphs [...] Any relationships or patterns must then be read using a knowledge of standard patterns” (OECD, 2016, S. 21). Entsprechend findet sich der Umgang mit Diagrammen gleich in zwei der drei naturwissenschaftlichen Teilkompetenzen wieder: dem *Erklären naturwissenschaftlicher Phänomene* und dem *Interpretieren von Daten und Evidenz* (siehe OECD, 2016; Schiepe-Tiska et al., 2016). Zum Erklären naturwissenschaftlicher Phänomene zählen das Erkennen, Erstellen und Nutzen von Erklärungsmodellen und zugehörigen

Repräsentationen. Das Interpretieren von Daten und Evidenz beschreibt die Fähigkeit, naturwissenschaftliche Daten in unterschiedlichen Repräsentationsformen zu analysieren und zu interpretieren sowie angemessene Schlussfolgerungen daraus abzuleiten. Außerdem zählt die Datentransformation von einer Repräsentation in eine andere zu dieser Fähigkeit.

Der National Council of Teachers and Mathematics (2000) in den USA sieht Repräsentationen als einen der fünf wichtigsten Inhalte an, während Datenanalyse als einer von fünf zentralen Prozessen gilt und als essenziell für Schülerinnen und Schüler unserer Zeit betrachtet wird. Hierbei steht neben dem Entwickeln wissenschaftlicher Fragen das Sammeln, Organisieren und Darstellen von Daten im Vordergrund.

Beliebt ist die Darstellung von Daten in Diagrammen in der Wissenschaft und anderen Bereichen deshalb, weil sie computationale Vorteile gegenüber anderen Darstellungsformen bieten, indem sie eine effizientere kognitive Verarbeitung ermöglichen (z. B. Larkin & Simon, 1987; Pinker, 1990). Sie unterscheiden sich von anderen Repräsentationsformen (z. B. Texten) dadurch, dass sie räumliche Relationen zwischen visuellen Objekten als Analogie für nicht-räumliche Relationen einsetzen (Koerber, 2011; Schnotz, 2002; Winn, 1990). In vielen Situationen bedeuten Inferenzen von nicht-räumlichen Repräsentationen einen hohen kognitiven Aufwand. Beispielsweise ist es schwieriger, Zahlen aus einem Fließtext zu extrahieren und zu vergleichen als auf den räumlichen Dimensionen in einem Diagramm zu operieren (z.B. indem man die visuelle Höhe von Säulen vergleicht). Burke (2007) betont den besonderen Nutzen von Diagrammen im Vergleich zu verbalen Mitteln: „things that are otherwise difficult to see or understand can sometimes become dramatically apparent when we look at the right graph, table or chart” (S.2).

Auch wenn Diagramme einige Aufgaben erleichtern, bieten sie keine universelle kognitive Unterstützung. Mittlerweile existiert eine Vielzahl von Visualisierungen (vgl. Bertin, 1983; Kosslyn, 1989; Lohse, Biolsi, Walker, & Rueler, 1994), aber es hängt gleichermaßen von den Anforderungen einer Aufgabe ab, welche Darstellungsform zur Lösung am besten geeignet ist (Peebles & Cheng, 2003; Simkin & Hastie, 1987; Shah, Mayer, & Hegarty, 1999). Shah und Hoeffner (2002) konkludieren in ihrem Forschungsüberblick beispielsweise, dass Trends leichter in Liniendiagrammen abgelesen werden können, während sich einzelne Datenpunkte besser in Säulendiagrammen vergleichen lassen. Darüber hinaus können zahlreiche Merkmale der Darstellung das Ablesen von Daten aus Diagrammen beeinflussen; darunter Farbe, Position und Größe, der Gebrauch von Legenden, die Anzahl räumlicher Dimensionen, Datenkomplexität und einige mehr (z. B. Cleveland & McGill, 1987; Kim & Lombardino, 2015; Kumar & Benbasat, 2004; Shah & Hoeffner, 2002).

Diese Forschungsarbeit widmet sich dem Einfluss visueller Gestaltungsmerkmale auf kognitive Prozesse bei der Bearbeitung von Diagrammleseaufgaben. Im Kontext von Lehren und Lernen spielen diese Einflüsse eine besondere Rolle, da Lehrende häufig auf vorgefertigtes Material zurückgreifen, bei dem zunächst nicht geklärt ist, ob es für den geplanten Einsatz geeignet ist. Kenntnisse über den Einfluss von Gestaltungsmerkmalen auf die kognitive Verarbeitung von Diagramm-Aufgaben können dabei helfen, für gegebene Aufgaben geeignete Repräsentationsformen zu identifizieren oder die Merkmale von Abbildungen derart anzupassen, dass diese für einen intendierten Zweck auch geeignet sind. Wissen über kognitive Hürden und Erschwernisse kann in der Folge dazu beitragen, für Lernende geeignete Werkzeuge zur Verfügung zu stellen, um sich effizient und erfolgreich mit den Erfordernissen unserer Welt auseinanderzusetzen und sie für die zunehmende Digitalisierung und steigenden Menge an Informationen zu wappnen (z. B. durch Vermeidung überflüssiger grafischer Elemente, Wahl geeigneter Komplexität und Vorgabe effizienter Repräsentationsformen).

1.1 Was ist ein Diagramm?

Es existiert eine Vielzahl von visuellen Repräsentationen. Darunter Graphen, Tabellen, Karten, Symbole, fotorealistische Bilder und einige mehr (siehe Bertin, 1983; Kosslyn, 1989; Lohse et al., 1994). Ein Teil dieser Repräsentationsformen wird im deutschsprachigen Raum unter dem Begriff *Diagramme* zusammengefasst (z. B. Struktur-, Venn-, Kreis-, Linien-, Säulen- und Balkendiagramme). So genannte Strukturdiagramme werden jedoch auch als *Graphen* bezeichnet. Dabei handelt es sich um bildhafte Darstellungen von Zusammenhängen zwischen qualitativen Merkmalen durch so genannte Knoten und gerichtete oder ungerichtete Kanten. Knoten werden in der Regel durch geometrische Formen (z. B. Rechtecke, Ellipsen) dargestellt und repräsentieren Begriffe oder andere Sachverhalte. Ungerichtete Kanten werden durch Linien, gerichtete Kanten durch Pfeile dargestellt und repräsentieren die Relationen zwischen den durch Knoten dargestellten Sachverhalten. Ein anschauliches Beispiel für den Einsatz von Graphen sind Bus- und Bahnnetze (siehe Abb. 1).

In diesem Zusammenhang kommt es zu einer terminologischen Konfusion mit entsprechenden Bezeichnern aus dem angloamerikanischen Sprachraum. Hier wird vornehmlich zwischen *Charts* und *Graphs* unterschieden (vgl. Kosslyn, 1989; Winn, 1987). Unter *Charts* werden in der Regel Darstellungen qualitativer Merkmale verstanden (z. B. Struktur- und Venn-Diagramme), während *Graphs* überwiegend Darstellungen quantitativer Merkmale meinen (z.B. Säulen-, und Liniendiagramme). Mitunter werden die Begriffe

allerdings auch synonym verwendet, sodass für Balkendiagramme sowohl die Bezeichnung *Bar Graphs* als auch *Bar Charts* zu finden ist. Zusätzlich wird der Charts-Begriff auch für Tabellen verwendet. Darüber hinaus existiert die Bezeichnung *Diagrams*, die neben Diagrammen aber auch realistische Schema-Zeichnungen einschließt. Aufgrund der hohen Heterogenität im Gebrauch der Bezeichnungen, und um Verwechslungen vorzubeugen, ist es notwendig, den Diagrammbegriff für die Verwendung in der vorliegenden Arbeit zu präzisieren. Hierzu stelle ich zunächst Klassifikationen von Repräsentationen vor (Abschnitt 1.1.1) und erarbeite auf dieser Grundlage eine Arbeitsdefinition für Diagramme (Abschnitt 1.1.2). Anschließend werden die gängigsten Diagrammtypen kurz vorgestellt und voneinander abgegrenzt (Abschnitt 1.1.3).

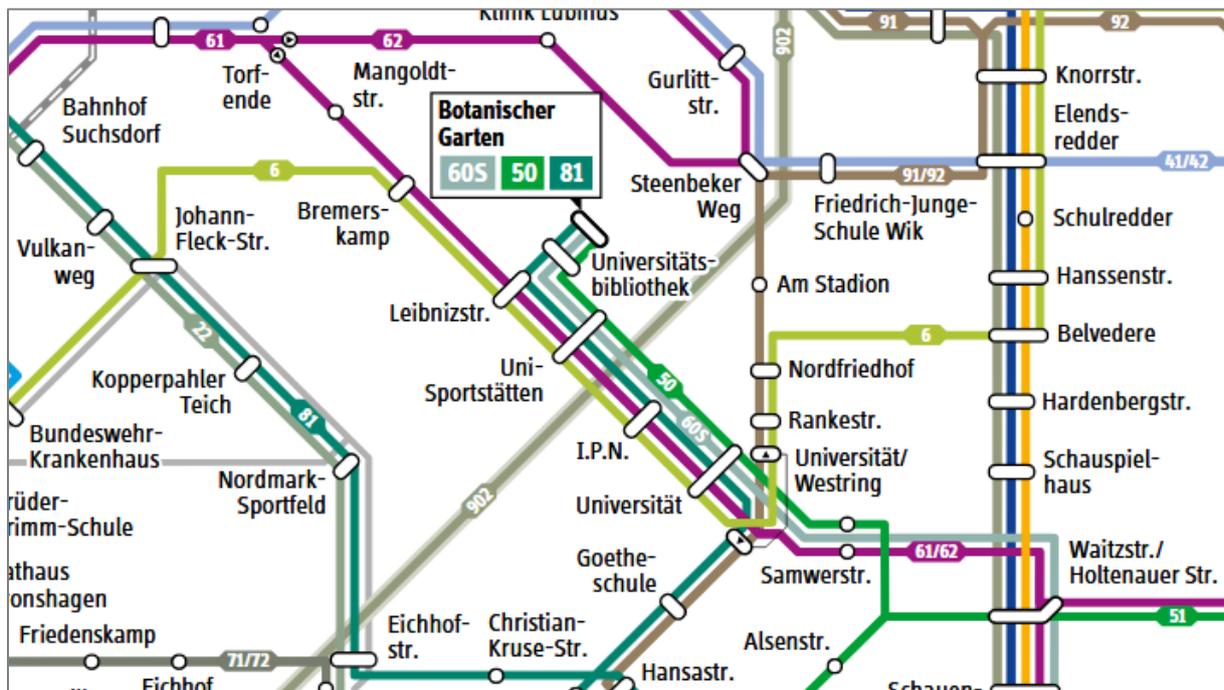


Abbildung 1. Ausschnitt aus dem Kieler Busliniennetz, dargestellt als Graph (Kieler Verkehrsgesellschaft, 2017). Die Haltestellen werden als Knoten repräsentiert (z. B. „I.P.N.“), die Buslinien sind als ungerichtete Kanten dargestellt.

1.1.1 Klassifikation von Repräsentationen

Palmer (1978) hat als einer der ersten Forscher den Repräsentationsbegriff systematisiert und dadurch erst fruchtbar für die psychologische Forschung gemacht hat. Grundsätzlich lassen sich interne Repräsentationen (d.h. mental, kognitiv) von externen Repräsentationen (d.h. physisch, in der Außenwelt) unterscheiden. Um spezifische, externe Repräsentationsformen wie Balken- und Liniendiagramme adäquat zu beschreiben und

voneinander abzugrenzen, werden hier zunächst die definierenden Merkmale von Repräsentationen im Allgemeinen geklärt.

Grundsätzlich lassen sich Repräsentationen als Modell für diejenige Sache verstehen, die sie repräsentieren. Daraus folgt, dass sowohl ein Repräsentiertes (die *repräsentierte Welt*) als auch ein Repräsentierendes (die *repräsentierende Welt*) existieren. Die Aufgabe der repräsentierenden Welt ist es, Aspekte der repräsentierten Welt auf bestimmte Weise wiederzugeben. Dabei müssen nicht alle Elemente der repräsentierten in der repräsentierenden Welt enthalten sein und gleichzeitig modellieren nicht alle Elemente einer repräsentierenden Welt notwendigerweise Aspekte der repräsentierten Welt. Nach Palmer (1978) gibt es fünf Aspekte, die notwendig sind, um eine Repräsentation vollständig zu spezifizieren: (1) was die repräsentierte Welt ist; (2) was die repräsentierende Welt ist; (3) welche Aspekte der repräsentierten Welt modelliert werden; (4) welche Aspekte der repräsentierenden Welt die Modellierung leisten; und (5) welche Übereinstimmungen zwischen den beiden Welten existieren.

Beispiel 1. Man stelle sich eine Gruppe von vier Personen vor (A, B, C, D). Die Personen sind unter anderem verschiedenen Geschlechts, unterschiedlichen Alters und unterscheiden oder gleichen sich auch sonst in zahlreichen physischen und psychischen Merkmalen. Dies sei die repräsentierte Welt (1). Wir wählen ein Säulendiagramm als repräsentierende Welt (2). Der Aspekt der repräsentierten Welt, den wir modellieren wollen, sei die Körpergröße der Personen (3). Hierzu verwenden wir die Höhe der Säulen im Diagramm (4). Dabei steht eine Säule für je eine Person und die Höhe der Säule im Diagramm repräsentiert die zugehörige Körpergröße der Person. Ist eine Person A größer als eine Person B, dann ist auch eine Säule A' im Diagramm länger als eine Säule B' (5). Relationen, die im Diagramm in Bezug auf die Höhe der Säulen gelten, gelten gleichsam für die Körpergröße der Personen. Fragen nach der Körpergröße lassen sich folglich allein durch Inspektion des Modells beantworten, indem man die Höhen der Säulen betrachtet. Die Breite der Säulen modelliert hingegen keinen Aspekt der repräsentierten Welt und zahlreiche Aspekte der repräsentierten Welt (Geschlecht, Alter etc.) werden im Diagramm nicht modelliert. Siehe Abbildung 2 für eine Veranschaulichung des Beispiels.

Beispiel 1 verdeutlicht, in welchem Bezug repräsentierte Welt und repräsentierende Welt in repräsentationalen Systemen zueinander stehen. Eine Vielzahl vergleichbarer oder gar

identischer Säulendiagramme wäre vorstellbar. Nur wenn die repräsentierte Welt, die modellierten Aspekte und die Übereinstimmungen zwischen den beiden Welten bekannt sind, lassen sich die verschiedenen Repräsentationen voneinander unterscheiden. Andersherum lassen sich dieselben Aspekte (Körpergröße der Personen) durch völlig andere Repräsentationsformen darstellen, die sich in denjenigen Elementen unterscheidet, welche die Aspekte der repräsentierten Welt modellieren.

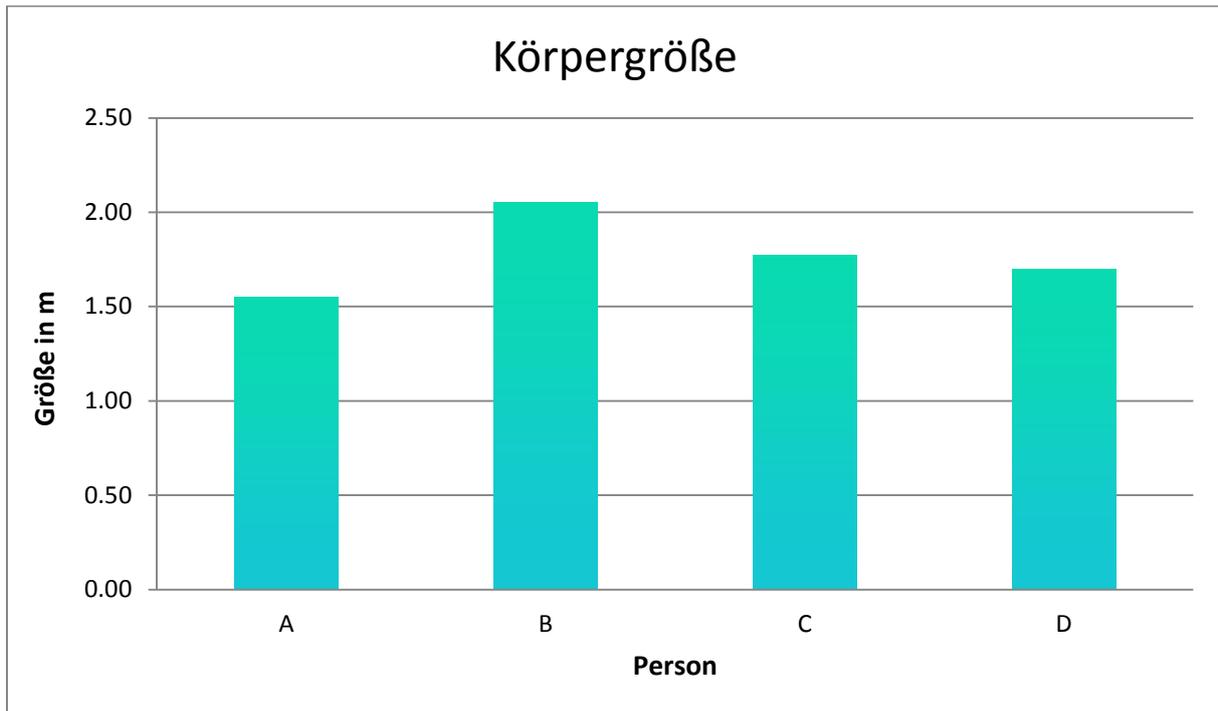


Abbildung 2. Beispieldiagramm: Körpergröße von Personen

1.1.1.1 Intrinsische und extrinsische Repräsentationen

Palmer (1978) bemerkt darüber hinaus, dass nur solche Informationen in einer Repräsentation enthalten sind, für die Operationen definiert sind, um sie zu erhalten. Wie in Beispiel 1 illustriert, kann eine repräsentierende Welt anstelle der repräsentierten Welt genutzt werden, um Informationen zu gewinnen. Ein Säulendiagramm ist aber nur dann sinnvoll interpretierbar, wenn Personen Operationen zur Verfügung stehen, mit denen sie Säulen identifizieren und Längenunterschiede zwischen ihnen feststellen können. Das ist in diesem Beispiel leicht, da wir uns auf Kategorien wie die „größer als“-Relation stützen können, die uns intuitiv zur Verfügung stehen. Es sind demgegenüber aber auch Repräsentationsformen vorstellbar, für deren Darstellungsprinzipien solche intuitiven Kategorien nicht zur Verfügung stehen. In Graphen beispielsweise können die dargestellten Kanten eine Vielzahl von Relationen zwischen den Knoten abbilden, beispielsweise Zugehörigkeit, Rangordnung oder

Reihenfolge (für ein Beispiel siehe Abb. 1). Diese sind nicht intuitiv zugänglich, sondern konventionell geregelt. Ist die Konvention nicht bekannt, steht die entsprechende Operation nicht zur Verfügung, sodass der Graph keinerlei Information bietet.

Palmer (1978) unterscheidet deshalb zwischen *intrinsischen* und *extrinsischen Repräsentationen*. Eine Repräsentation heißt intrinsisch, wenn die Relationen in der repräsentierenden Welt denselben inhärenten Beschränkungen unterliegen wie die Relationen der repräsentierten Welt. Die „größer als“-Relation in Bezug auf Körpergröße wird in Beispiel 1 durch die „länger als“-Relation zwischen Säulen repräsentiert. Beide Relationen sind asymmetrisch (d. h. Wenn x größer als y ist, kann y nicht größer als x sein) und transitiv (d. h. wenn y größer als z ist, muss auch x größer als z sein). Im Gegensatz dazu heißt eine Repräsentation extrinsisch, wenn die inhärente Struktur der repräsentierenden Welt vollständig arbiträr ist und diejenige der repräsentierten Welt es nicht ist. Die Strukturinformationen werden gewissermaßen erst von außen hinzugegeben.

Schnotz (2001) spricht auch von *deskriptionalen* (= extrinsischen) und *depiktionalen* (= intrinsischen) Repräsentationen. In deskriptionalen Repräsentationen werden Sachverhalte mithilfe von arbiträren Symbolen beschrieben, die keine Ähnlichkeit zu dem Bezeichneten haben, wie in Texten und gesprochener Sprache (z. B. hat das Wort Flugzeug keine Ähnlichkeit mit der realen Maschine). Deskriptionale Repräsentationen enthalten zudem explizite Zeichen für Relationen (z. B. Verben und Präpositionen), deren Bedeutung konventionell von außen festgelegt ist. Depiktionale Repräsentationen hingegen enthalten keine expliziten Relationszeichen, sondern nutzen inhärente Struktureigenschaften, die mit bestimmten anderen Struktureigenschaften des dargestellten Sachverhalts übereinstimmen. Gemäß dieser Unterscheidung fallen sowohl Fotos als auch Diagramme in den Bereich der depiktionalen beziehungsweise intrinsischen Repräsentation. Um diese voneinander zu trennen, lässt sich darüber hinaus zwischen *konkreten* und *abstrakten* strukturellen Übereinstimmungen unterscheiden (Schnotz, 2001). Bei dem konkreten Typ stimmen repräsentierte und repräsentierende Attribute miteinander überein. Das ist beispielsweise bei einer maßstabgetreuen Karte der Fall und bei verkleinerten Objektmodellen. Länge wird hier durch Länge und Höhe durch Höhe repräsentiert. Bei einer abstrakten strukturellen Übereinstimmung unterscheiden sich die repräsentierten Attribute von den repräsentierenden. So kann in einem Säulendiagramm beispielsweise der Kraftstoffverbrauch von Fahrzeugen durch die Länge einer Säule abgebildet werden.

1.1.1.2 Realistische und logische Bilder

Schnotz (1994) nennt Abbildungen mit konkreten Übereinstimmungen auch *realistische Bilder* und Abbildungen mit abstrakten Übereinstimmungen *logische Bilder*. Bei so genannten realistischen Bildern handelt es sich um Zeichnungen, Gemälde, Photographien und andere Abbildungen, die derart gestaltet sind, dass sie in der menschlichen Wahrnehmung ein ähnliches Perzept erzeugen wie das Original (d. h. das physicogeometrische Energiemuster der Lichtpartikel einer Abbildung, das auf die Retina trifft, ähnelt demjenigen Energiemuster, das auch von der visuellen Szene erzeugt wird, die abgebildet ist; vgl. Mausfeld, 2010). Beispielsweise erkennen Personen die Abbildung eines Hauses als solches aufgrund der perzeptuellen Ähnlichkeit zu dem tatsächlichen Haus. Das Erkennen des Abgebildeten erfolgt demnach anhand von Schemata der alltäglichen Wahrnehmung und bedarf keiner weiteren Instruktionen.

Demgegenüber sind mit logischen Bildern solche Arten von Abbildungen gemeint, die Zusammenhänge zwischen quantitativen und qualitativen Merkmalen oder ausschließlich quantitativen Merkmalen eines Sachverhaltes darstellen, die nicht direkt wahrnehmbar sein müssen und keine perzeptuelle Ähnlichkeit zu dem Dargestellten besitzen (z. B. die Darstellung des Altersaufbaus einer Bevölkerung in einem Balkendiagramm). Logische Bilder zeichnen sich dadurch aus, dass vorwiegend räumliche Relationen als Analogie für nichträumliche Relationen verwendet werden (z.B. Koerber, 2011; Schnotz, 2002; Winn, 1990). Zwischen der Abbildung und dem Abgebildeten besteht eine abstrakte Strukturähnlichkeit. Sie ermöglichen ein Sichtbarmachen von Sachverhalten, die nicht direkt wahrnehmbar sind und können auf diese Weise komplexe Sachverhalte einfacher darstellen (Schnotz, 1994). Allerdings erfordern logische Bilder zum Teil Vorwissen über die jeweiligen Darstellungskonventionen, die aufschlüsseln, welche nichträumlichen Merkmale und Relationen durch welche räumlichen Merkmale und Relationen repräsentiert werden (Pinker, 1990). Zum Wissen um die Darstellungskonventionen eines Säulendiagramms zählt beispielsweise, dass jede Säule für die Ausprägung eines (kategorialen) Merkmals steht und die Höhe der Säulen die Ausprägung eines zweiten Merkmals abbildet, während die Breite der Säulen und die Größe der Abstände zwischen ihnen keine Bedeutung tragen. Eine Übersicht über die Klassifikation von Repräsentationen nach Schnotz (1994, 2001) und Palmer (1978) findet sich in Abbildung 3.

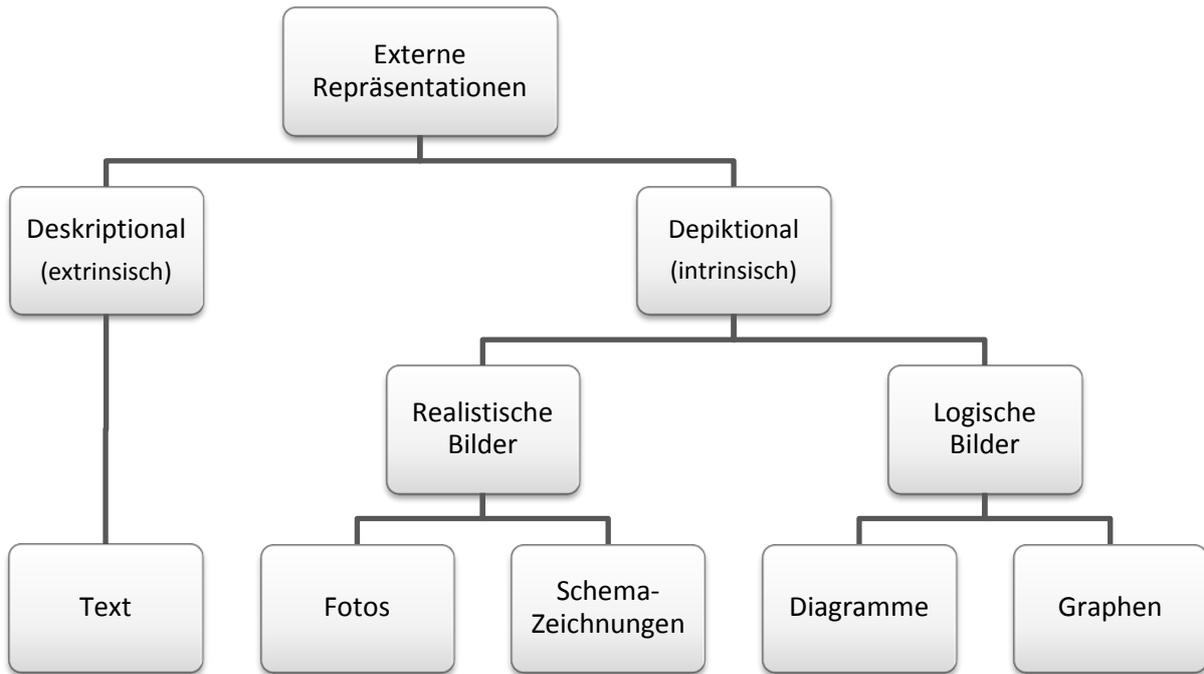


Abbildung 3. Klassifikation von Repräsentationen nach Schnotz (1994; 2001). Bezeichnungen von Palmer (1978) sind in Klammern notiert.

1.1.2 Arbeitsdefinition für Diagramme

Um den Diagrammbegriff einzugrenzen und unbeabsichtigte Verwechslungen mit anderen Auslegungen auszuschließen, wähle ich für die vorliegende Arbeit die folgende Terminologie:

Als *Diagramm* werden depiktionale (intrinsische) Repräsentationen verstanden, die vorwiegend räumliche Relationen nutzen, die in abstrakter struktureller Übereinstimmung mit den repräsentierten Relationen eines Sachverhaltes stehen (vgl. Palmer, 1978; Schnotz, 2001). Diagramme sind der Gruppe logischer Bilder zuzuordnen (Schnotz, 1994) und unterscheiden sich von realistischen Bildern dadurch, dass die repräsentierenden Merkmale (z. B. Höhe einer Säule) mit den repräsentierten Merkmalen *nicht* übereinstimmen (z. B. Kraftstoffverbrauch eines Fahrzeugs). Sie stellen Zusammenhänge zwischen qualitativen und quantitativen oder ausschließlich quantitativen Merkmalen dar (z. B. Kraftstoffverbrauch und Fahrzeugtyp). Dazu zählen insbesondere Kreis-, Säulen-, Balken-, Linien- und Streudiagramme. Nicht gemeint sind im deutschsprachigen Raum mit dem Begriff *Graphen* bezeichnete Abbildungen, die ausschließlich qualitative Merkmale repräsentieren und aufgrund ihres symbolischen Zeichencharakters der Gruppe deskriptionaler (extrinsischer) Repräsentationen zuzuordnen sind.

In den englischsprachigen Teilen dieser Arbeit wird die im angloamerikanischen Sprachraum gebräuchliche Bezeichnung *Graph* anstelle des deutschsprachigen Begriffs Diagramm verwendet.

1.1.3 Diagrammentypen

Die Menge an Repräsentationen, die unter den zuvor erarbeiteten Diagrammbegriff fallen, ist trotz enger Definition noch immer sehr groß. Aufgrund der hohen Zahl an existierenden Diagrammtypen, sollen hier lediglich eine Auswahl der gebräuchlichsten Abbildungen vorgestellt werden. Dazu zählen insbesondere Streudiagramme, Liniendiagramme, Balken- und Säulendiagramme sowie Kreisdiagramme (siehe z. B. Cleveland & McGill, 1987; Pinker, 1990; Schnotz, 1994). Die gemeinsame Basis der meisten Diagrammtypen ist ein ebenes, kartesisches Koordinatensystem mit zwei senkrecht (normal) stehenden Achsen, die eine zweidimensionale Fläche aufspannen. Sie werden daher auch als *Achsendiagramme* bezeichnet (Kreisdiagramm ausgenommen). In weiteren Merkmalen können sich die verschiedenen Diagrammtypen zum Teil deutlich unterscheiden. Eine Übersicht findet sich in Abbildung 4.

1.1.3.1 Streudiagramm

In einem Streu- oder Punktdiagramm (auch *Scatterplot*) werden in der Regel zwei quantitative Merkmale in Form von Wertepaaren als Punkte eingetragen, wobei die Ausprägung des einen Wertes der Position auf der horizontalen Achse (der Abszisse), die Ausprägung des anderen Wertes der Position auf der vertikalen Achse (der Ordinate) entspricht, wodurch die Lage eines Punktes im Raum eindeutig bestimmt werden kann. Streudiagramme eignen sich beispielsweise zur Darstellung stochastischer Zusammenhänge zwischen den jeweiligen Merkmalen (Schnotz, 1994).

1.1.3.2 Liniendiagramm

Liniendiagramme besitzen einen ähnlichen Aufbau wie Streudiagramme, mit dem Unterschied, dass die Wertepaare geordnet dargestellt werden (typischerweise über die Abszisse) und mit einer Linie verbunden sind. Liniendiagramme eignen sich insbesondere für die Darstellung von Verläufen und Trends (Shah & Hoeffner, 2002). Ist die Fläche zwischen Linien oder Achsen und Linien ausgefüllt, spricht man auch von einem Flächendiagramm. Die Flächen zwischen Achsen und den Linien werden typischerweise mit unterschiedlichen

Farben oder Mustern hervorgehoben. Flächendiagramme eignen sich besonders zur Darstellung und zum Vergleich von Mengen.

1.1.3.3 Balken- und Säulendiagramm

In Balken- und Säulendiagrammen werden quantitative Merkmale in Kombination mit qualitativen Merkmalen (Kategorien) dargestellt, wobei jede Kategorie durch eine rechteckige Fläche dargestellt wird, deren Länge die Ausprägung des quantitativen Merkmals in der entsprechenden Kategorie repräsentiert. Die Breite der Fläche hat hingegen keinerlei Bedeutung. Stehen die rechteckigen Flächen senkrecht auf der Abszisse, heißen sie Säulen. Wenn sie waagrecht an der Ordinate anliegen, spricht man von Balken. Sie werden dazu verwendet, um absolute oder relative Häufigkeiten verschiedener Kategorien oder um Rangfolgen von Merkmalen darzustellen (Büchter & Henn, 2006). Um ein weiteres qualitatives Merkmal in Säulendiagrammen zu repräsentieren, können Säulen durch Farben oder Muster zu Gruppen zusammengefasst werden. Diese Darstellung wird dann als Gruppirtes Säulendiagramm bezeichnet.

Das Skalenniveau des zweiten Merkmals ist typischerweise nominal (z. B. Parteien bei einer Wahl), sodass sich die Säulen gemäß dem ersten Merkmal der Höhe nach ordnen lassen (Büchter & Henn, 2006). Manchmal hat das zweite Merkmal jedoch Ordinalniveau. Dann empfiehlt sich auch eine entsprechende Reihung gemäß dem zweiten Merkmal (z. B. aufsteigend nach Jahren).

1.1.3.4 Gestapeltes Säulendiagramm

Gestapelte Säulendiagramme heißen auch *Stapelendiagramme*. In dieser Spezialform des Säulendiagramms werden absolute oder relative Häufigkeiten in einer rechteckigen Säule dargestellt. Hierbei werden mehrere Säulensegmente übereinander angeordnet, wobei jedes Segment einem Wert eines kategorialen Merkmals entspricht (z. B. der Name einer Partei im Bundestag) und die Höhe eines Säulensegments, wie im klassischen Säulendiagramm, dem Wert eines quantitativen Merkmals (z. B. dem Stimmanteil bei der Bundestagswahl) entspricht. Über die Position auf der Achse werden die Werte einer dritten Variablen abgebildet (z. B. die Wahljahre).

Stapelendiagramme, die relative Häufigkeiten abbilden, werden häufig zum Vergleich von Anteilen und deren Veränderung, beispielsweise über die Zeit, genutzt (Büchter & Henn, 2006). Die Abbildung absoluter Häufigkeiten wird hingegen für den Vergleich von Gesamtwerten verwendet.

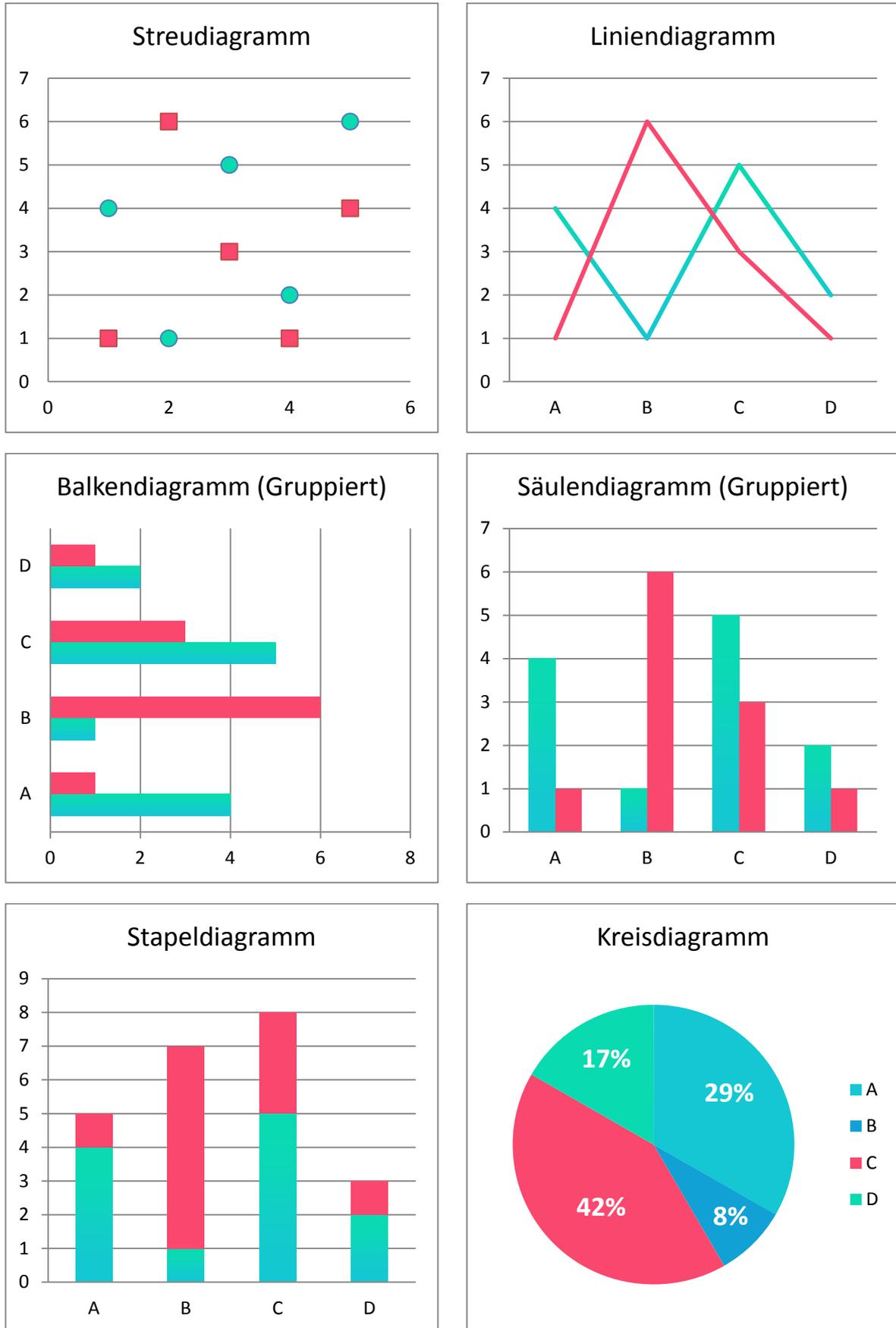


Abbildung 4. Beispiele gebräuchlicher Diagrammtypen

1.1.3.5 Kreisdiagramm

Handelt es sich bei quantitativen Werten um Teile eines Ganzen, die unterschiedlichen Kategorien zugeordnet werden können, lassen sich diese in Form von Kreissektoren in einem Kreisdiagramm darstellen. Bei Kreissektoren (umgangssprachlich auch als Kuchenstück bezeichnet) entsprechen die Verhältnisse von Flächeninhalten, Bogenlängen und Winkeln einander (Büchter & Henn, 2006). Diese visuellen Indikatoren stehen für die Ausprägung des repräsentierten Merkmals in der jeweiligen Kategorie. Ähnlich wie Stapeldiagramme werden Kreisdiagramme zum Vergleich von Anteilen eingesetzt.

Die Darstellung von Rangordnungen ist in Kreisdiagrammen nicht komplikationsfrei möglich. Zwar lassen sich die Kreissektoren im Uhrzeigersinn der Größe nach ordnen, doch der Endpunkt grenzt schließlich wieder an den Startpunkt. Daher bieten sich Kreisdiagramme vor allem für nominal skalierte Daten an (Büchter & Henn, 2006).

1.1.4 Abgrenzung zu Tabellen und Texten

Diagramme sind weiterhin von Texten und Tabellen abzugrenzen. Texte werden ebenfalls im zweidimensionalen Raum dargestellt, allerdings verwenden sie arbiträre Symbolsysteme, deren räumliche Anordnung in keiner Ähnlichkeitsrelation zu dem Bezeichneten steht. Die inhärenten Struktureigenschaften von Texten stimmen nicht mit denjenigen des Bezeichneten überein. Somit sind sie den deskriptionalen beziehungsweise extrinsischen Repräsentationen zuzuordnen (vgl. Palmer, 1978; Schnotz, 2001). Tabellen nutzen, genau wie Diagramme, räumliche Relationen für Merkmalszuordnungen, allerdings können in dieser Repräsentationsform, ähnlich wie in Graphen, nur qualitative Relationen abgebildet werden, etwa die Zuordnung von Werten zu verschiedenen Kategorien. Nicht möglich ist hingegen die Abbildung quantitativer Relationen wie beispielsweise in Säulendiagrammen, in denen die Höhe einer Säule in direkter Relation zu einem nichträumlichen, quantitativen Merkmal steht.

Diese inhärenten Struktureigenschaften von Tabellen schlagen sich entsprechend in ihren computationalen Eigenschaften nieder (siehe auch Abschnitt 1.3). So können Personen aus Tabellen besonders schnell und zuverlässig einzelne Datenpunkte extrahieren (Guthrie, Weber, & Kimmerly, 1993; Meyer, 2000). Wenn Informationen integriert werden müssen, beispielsweise um mehrere Datenpunkte zu vergleichen oder einen Trend zu identifizieren, erfordern Tabellen wesentlich stärker als Diagramme eine serielle Verarbeitung der einzelnen Datenpunkte, denn es gibt keine visuellen Elemente, die quantitative Relationen kodieren. Personen können diese Aufgaben schneller und zuverlässiger mit Diagrammen lösen (Carey

& White, 1991; Greaney & MacRae, 1997; Legge, Gu & Luebker, 1989; Tuttle & Kershaw, 1998).

Zusammengenommen bedeutet das, dass Texte und Tabellen konzeptuell anderen Repräsentationsformen zuzuordnen sind (d.h. den deskriptionalen; Schnotz, 2001), während sich gleichzeitig auch ihre computationalen Eigenschaften systematisch von denen der Diagramme unterscheiden. Im folgenden Abschnitt werden zunächst psychologische Modelle des Diagrammverstehens vorgestellt. Computationale Eigenschaften von Diagrammen, die aus einzelnen Gestaltungsmerkmalen resultieren, werden in Abschnitt 1.3 näher besprochen.

1.2 Modelle des Diagrammverstehens

Nachdem ich zunächst begrifflich den Gegenstand dieser Forschungsarbeit beschrieben habe, stehen die psychologischen Grundlagen des Lesens und Verstehens von Diagrammen im Fokus dieses Abschnitts. Ziel der Diagrammforschung ist es, dem Lesen und Verarbeiten von Diagrammen zugrundeliegende kognitive und perzeptuelle Prozesse zu verstehen, diese zu beschreiben und Vorhersagen daraus abzuleiten, wie verschiedene Faktoren den Prozess des Ablesens, das Verstehen von Diagrammen und das Behalten von Information beeinflussen.

Im Wesentlichen existieren zwei Modellklassen für das Verstehen von Diagrammen (Shah, Freedman, & Vekiri, 2005). Zum einen sind das computationale Modelle, die isoliert auf simple Ableseprozesse fokussieren (z. B. das Extrahieren einzelner Datenpunkte aus einer Datenreihe). Sie liefern detaillierte Beschreibungen dieser Vorgänge und machen konkrete Vorhersagen über Bearbeitungszeit und Lösungsrichtigkeit (z. B. Meyer, 2000; Hollands & Dyre, 2000; Simkin & Hastie, 1986). Diese Modelle haben in der Regel nur einen kleinen Geltungsbereich, der auf konkrete Ableseaufgaben beschränkt ist und lassen weitere Variablen unberücksichtigt, wie etwa den Einfluss von Personeneigenschaften (z. B. Vorwissen, Arbeitsgedächtniskapazität) und den dargestellten Inhalten in den Diagrammen (Shah et al., 2005). Demgegenüber fokussiert eine zweite Modellklasse auf allgemeine Prozessanalysen des Diagrammverstehens, die einen breiteren Geltungsbereich anstreben, dafür jedoch weniger genaue Vorhersagen für konkrete Aufgaben liefern können (Bertin, 1983; Freedman & Shah, 2002; Pinker, 1990). Darüber hinaus sind zahlreiche Bemühungen unternommen worden, um Ableseprozesse und typische Aufgaben zu klassifizieren (z. B. Bertin, 1983; Curcio, 1987; Schnotz, 1994; Wainer, 1992).

In den folgenden Abschnitten werden zunächst Taxonomien für typische Diagrammleseaufgaben beschrieben (Abschnitt 1.2.1) und anschließend Modelle des Diagrammverstehens vorgestellt (Abschnitt 1.2.2).

1.2.1 Taxonomien für Diagrammleseaufgaben

In diesem Abschnitt widme ich mich der Frage, was Aufgaben auszeichnet, die typischerweise mit Diagrammen bearbeitet werden. Wie eine Reihe empirischer Studien dokumentiert, hängen Schwierigkeit, Richtigkeit und Bearbeitungszeit bei der Bearbeitung von Aufgaben mit Diagrammen von einer Interaktion zwischen gewählter Darstellungsform und der Art der Aufgabe ab (z. B. Peebles & Cheng, 2003; Simkin & Hastie, 1987; Shah et al., 1999). Daher ist neben einem Verständnis für die Merkmale von Diagrammen auch ein Wissen über typische Diagrammleseaufgaben von großer Bedeutung. Taxonomien, die zu diesem Zwecke entwickelt wurden, weisen eine ausgesprochen hohe Konvergenz auf (siehe Bertin, 1983; Curcio, 1987; Schnotz, 1994; Wainer, 1992). Allen Taxonomien ist dabei gemeinsam, dass sie sich auf verschiedene *Ableseprozesse* beziehen. Insofern können das Ablesen von Daten und die Anwendung diverser Operationen auf eine Datenbasis (wie z. B. Vergleiche) als Spezifikum von Aufgaben betrachtet werden, die mit Diagrammen bearbeitet werden. Bertin (1983) nennt diese Ableseprozesse *Stufen des Erfassens*. Sowohl er als auch Wainer (1992) bezeichnen sie mitunter auch als *Fragen*, die man ein Diagramm richten kann. Schnotz (1994) nennt sie hingegen *Stufen der Informationsentnahme*. Ich verwende in dieser Arbeit die Bezeichnung *Diagrammleseaufgaben* synonym zu diesen Begriffen. Im Folgenden stelle ich unterschiedliche Ansätze der Klassifikation vor und weise dabei auf die überwiegenden Gemeinsamkeiten und wenigen Unterschiede hin. Siehe Abbildung 5 für eine grafische Übersicht.

Curcio (1987)	Bertin (1983)	Wainer (1992)	Schnotz (1994)
Lesen der Daten	Externe Identifizierung		
	Interne Identifizierung		
	Elementare Stufe	Elementare Stufe	Ablesen 1. Ordnung
Lesen zwischen den Daten	Mittlere Stufe	Mittlere Stufe	Ablesen 2. Ordnung
	Obere Stufe	Obere Stufe	Ablesen 3. Ordnung
Lesen über die Daten hinaus	-	-	-

Abbildung 5. Grafischer Vergleich verschiedener Taxonomien für Diagrammleseaufgaben. Bei Bezeichnungen, die in derselben Zeile stehen, stehen für inhaltlich verwandte Konzepte.

Eine frühe Klassifikation von Diagrammleseaufgaben findet sich bei Bertin (1983). Er unterteilt das Lesen von Diagrammen in drei hierarchisch gegliederte Stufen des Erfassens: Auf der *Elementaren Ebene* nehmen Aufgaben lediglich Bezug auf ein einziges Element der Darstellung. Ein typisches Beispiel ist das Ablesen eines einzelnen Datenpunktes. Aufgaben der *Mittleren Stufe* nehmen Bezug auf eine Gruppe von Elementen, beispielsweise beim Ablesen eines Trends. Eine Aufgabe der mittleren Stufe muss nicht zwingend alle Datenpunkte einer Datenreihe berücksichtigen, sondern kann sich auf Untergruppen von Datenpunkten beziehen. Die *Obere Stufe* erfordert schließlich die Gesamterfassung einer dargestellten Variablen, also die globale Betrachtung aller Datenpunkte eines Diagramms.

Wainer (1992) präziserte rund zwanzig Jahre später die drei Stufen und formulierte sie neu. Fragen der *Elementaren Stufe* beziehen sich bei ihm ebenfalls auf die Extraktion eines einzelnen Datums (z. B. „Wie hoch war der Petroleumverbrauch im Jahr 1980?“). Bei Fragen der *Mittleren Stufe* müssen Trends in Teilen der Daten identifiziert werden (z. B. „Wie hat sich der Petroleumverbrauch zwischen 1970 und 1985 verändert?“). Fragen der *Oben Stufen* zielen wie bei Bertin (1983) auf das Verständnis von der Datenstruktur in ihrer Gesamtheit ab und beinhalten beispielsweise den Vergleich von Trends oder das Erkennen von Gruppierungen (z. B. „Welcher der Brennstoffe zeigt den größten Anstieg im Verbrauch?“).

Zeitgleich entwickelte Curcio (1987) eine eigene Taxonomie des Diagrammlesens, in der sich vergleichbare Kategorien wiederfinden. Die unterste Stufe bezeichnet er als das *Lesen der Daten*. Dies beinhaltet das Ablesen eines einzelnen Datums, Titels oder Achsen-Labels. Damit ist die Stufe vergleichbar mit der Elementaren Ebene bei Bertin (1983) und Wainer (1992). Die zweite Stufe ist mit *Lesen zwischen den Daten* bezeichnet und umfasst sowohl Vergleiche zwischen verschiedenen Datenpunkten als auch die Integration verschiedener Informationen, beispielsweise zu einem Trend. Damit umfasst Curcios zweite Stufe sowohl die zweite als auch dritte Stufe von Bertin und Wainer und bietet in diesem Sinne eine geringere Differenzierung. Seine dritte und höchste Stufe heißt *Lesen über die Daten hinaus*. Auf dieser Stufe sind Lesende gefordert, auf einer dargestellten Datengrundlage Vorhersagen zu treffen oder Daten zu extrapolieren. Diese Stufe geht über die anderen Ansätze hinaus und hat keine Entsprechung in den Taxonomien von Bertin und Wainer.

Schnotz (1994) formulierte auf Basis seiner Vorläufer schließlich eine deutlich umfangreichere Klassifikation, die er als *Hierarchien der Informationsentnahme* bezeichnet. Der streng hierarchische Gedanke drückt sich im Besonderen in den Bezeichnungen der drei

Stufen bei Schnotz aus: das Ablesen 1. Ordnung, das Ablesen 2. Ordnung und das Ablesen 3. Ordnung.

Als *Ablesen 1. Ordnung* wird das Ablesen von Einzelwerten aus einem Diagramm verstanden. Das abgefragte Datum kann dabei unterschiedlichen Variablen angehören. Ist die Körpergröße mehrerer Personen nach Geschlecht getrennt abgetragen, kann beispielsweise das Geschlecht einer Person ebenso bestimmt werden wie die Körpergröße.

Das *Ablesen 2. Ordnung* bezieht sich auf Relationen zwischen einzelnen Merkmalsausprägungen oder von Variablenzusammenhängen, beispielsweise der Vergleich von Datenpunkten oder das Ablesen von Trends. Beim Ablesen 2. Ordnung lässt sich zwischen Querschnittvergleichen und Längsschnittvergleichen unterscheiden. Vergleicht man Datenpunkte an derselben Stelle in unterschiedlichen Datenreihen, spricht man von einem Querschnittvergleich. Werden Werte innerhalb einer Datenreihe verglichen, handelt es sich um einen Längsschnittvergleich. Hierbei lassen sich auch Trends und Zusammenhänge zwischen Variablen feststellen.

Beim *Ablesen 3. Ordnung* handelt es sich um die komplexeste Prozedur, nämlich das Ablesen von Relationen zwischen Relationen zwischen einzelnen Werten. Das beinhaltet den Vergleich von Trends wie auch den Vergleich von Unterschieden zwischen je zwei Werten. Auch hier lassen sich Längsschnitt- und Querschnittvergleiche unterscheiden. Im Längsschnitt werden Trends oder Unterschiede innerhalb einer Datenreihe verglichen, während im Querschnitt die Trends oder Unterschiede zwischen verschiedenen Datenreihen verglichen werden.

Schnotz (1994) weist zudem darauf hin, dass die aufgeführten Ableseprozesse auf unterschiedlichen Skalenniveaus stattfinden können. Häufig findet das Ablesen, wie auch in vielen Beispielen dieser Arbeit, auf Ordinalniveau statt (z. B. Ist ein Wert größer oder kleiner als ein anderer? Ist ein Trend steigend oder fallend?). Ableseprozesse sind jedoch ebenso auf Intervallskalenniveau (z. B. Um wieviel ist ein Wert größer als ein bestimmter anderer?) und Verhältnisskalenniveau (z. B. Um das wieviel fache größer ist ein Wert als ein bestimmter anderer?) vorstellbar.

In der Zusammenschau der vorgestellten Klassifikationsansätze wird deutlich, dass Diagrammleseaufgaben mit hoher Konfidenz hierarchisch geordnet werden können. Das heißt, einige Aufgabentypen sind per definitionem komplexer als andere. Während dies bei allen Ansätzen zu beobachten ist, wird es bei der Betrachtung der Taxonomie von Schnotz (1994) besonders deutlich. Diese ist derart organisiert, dass auf jeder Stufe der

Informationsentnahme eine weitere Relation berücksichtigt werden muss (zunächst nur ein Datum, dann Relationen zwischen mehreren Daten und schließlich Relationen zwischen diesen Relationen). Das bedeutet auch, dass die zu erwartende Komplexität einer Aufgabe gut durch die Zuordnung zu einer der Stufen vorhergesagt werden kann und dass es leicht möglich ist, Aufgaben hinsichtlich ihrer Komplexität einzustufen und welche von vergleichbarer Komplexität herzustellen.

1.2.2 Allgemeine Prozessmodelle des Diagrammverstehens

Allgemeine Prozessmodelle zeichnen sich dadurch aus, dass sie einen möglichst breiten Geltungsbereich für eine Vielzahl von Situationen, Aufgaben und Darstellungsformen anstreben. Dazu zählt beispielsweise auch die Berücksichtigung von individuellen Faktoren und den Inhalten der Darstellung. Im Folgenden stelle ich zunächst die Aufgabenanalyse nach Bertin (1983) und anschließend die umfassenden Prozessmodelle von Pinker (1990) sowie Freedman und Shah (2002) vor.

1.2.2.1 Aufgabenanalyse nach Bertin (1983)

Das Lesen von Diagrammen (und anderen depiktionalen Repräsentationsformen) ist nach Bertin (1983) in drei Schritte unterteilt: (1) Die externe Identifizierung, (2) die interne Identifizierung und (3) die Wahrnehmung der ursprünglichen Beziehungen.

Bei der externen Identifizierung müssen Personen zunächst die Invarianten (= invariable Kennzeichnung der Inhalte) und die Komponenten (= Variablen) erkennen. Diese werden in der Regel durch Titel und andere Beschriftungen in Diagrammen kenntlich gemacht. Stellt man in einem Säulendiagramm beispielsweise die Tordifferenzen einer Fußballmannschaft in den letzten zehn Spielen dar, so stellt die Fußballmannschaft eine Invariante dar, während die Spiele und die Höhe der Tordifferenz als Komponenten zu verstehen sind.

Sind Invarianten und Komponenten erkannt und gedanklich konkretisiert, müssen Personen erkennen, welche visuellen Objekte die Komponenten repräsentieren. Diesen Vorgang bezeichnet Bertin (1983) als interne Identifizierung. Bei dem Beispiel der Fußballmannschaft wird jedes Spiel durch eine Säule repräsentiert und die Größe der Tordifferenzen durch die Höhe der Säulen.

Die Wahrnehmung der ursprünglichen Beziehungen ist schließlich der letzte Schritt des Diagrammlesens in Bertins Analyse. Hierbei handelt es sich um den Ableseprozess einer Information, für den notwendigerweise eine Frage bzw. Aufgabe entweder implizit und

explizit formuliert werden muss (z. B. Wie hoch war die Tordifferenz im dritten Spiel?). Diese Typen von Fragen unterteilt Bertin (1983) in drei Stufen ein: Fragen zu einem einzigen Element einer Komponente, Fragen zu einer Gruppe von Elementen und Fragen zur Gesamtheit der Elemente einer Komponente (siehe Abschnitt 1.2.1).

1.2.2.2 Pinkers (1990) Theorie des Diagrammverstehens

Pinkers Theorie des Diagrammverstehens (*Theory of graph comprehension*; 1990) bietet ein deutlich umfassenderes Modell zum Lesen von Diagrammen, da es eine Reihe weiterer Faktoren einbezieht wie beispielsweise Arbeitsgedächtniskapazität und Vorwissen. Eine Visualisierung von Pinkers Modell des Diagrammverstehens findet sich in Abbildung 6.

Im ersten Schritt des Wahrnehmungsprozesses trifft ein physicogeometrisches Energiemuster auf die Retina. Das daraus resultierende Helligkeitsmuster bezeichnet Pinker als visuelles Feld (*visual array*). Dieser Begriff steht in Übereinstimmungen mit Konzepten wie dem *2½ dimensional sketch* (Marr & Nishihara, 1978) und dem *ikonischen Gedächtnis* (Phillips, 1974). Dieses Rohformat bietet allerdings nur basale Informationen (z. B. Helligkeit, Größe, Form etc.), die unbrauchbar sind, um ein Diagramm zu verstehen. Es muss zunächst in eine visuelle Beschreibung (*visual description*) transformiert werden (Bottom-Up-Prozess). Dabei handelt es sich um eine symbolische, strukturelle Repräsentation, welche die konstituierenden Elemente einer visuellen Szene und ihre Attribute (z. B. Größe, Form und Farbe) enthält sowie die Relationen zwischen diesen Elementen. Höhere kognitive Funktionen sind dann in der Lage, anstelle des Rohformats, auf diese symbolische Repräsentation zuzugreifen.

Die visuelle Beschreibung nutzt Aussagen (*propositions*), deren Variablen für wahrgenommene Objekte stehen und in denen *Prädikate* die Attribute von und Relationen zwischen visuellen Objekten spezifizieren. Pinker (1990) beschreibt einstellige Prädikate (z. B. konvex, flach), zweistellige Prädikate (z. B. größer, unter) und mehrstellige Prädikate (z. B. zwischen). Auch diese Beschreibung einer visuellen Szene ist allerdings nicht spezifisch genug, um Diagramme zu verstehen. Tatsächlich unterliegt sie notwendigerweise weiteren Beschränkungen (*constraints*). Dazu zählen Einschränkungen des visuellen Systems wie die Regeln perzeptueller Gruppierung, die von den Gestalt-Psychologen beschrieben wurden (z. B. Wertheimer, 1923), Beschränkungen des kognitiven Systems wie die Arbeitsgedächtniskapazität (Baddeley, 1994; Baddeley & Hitch, 1972) und Diagrammschemata (d. h. das Vorwissen einer Person über Diagramme, beispielsweise dass gerade Linien für lineare Zusammenhänge stehen).

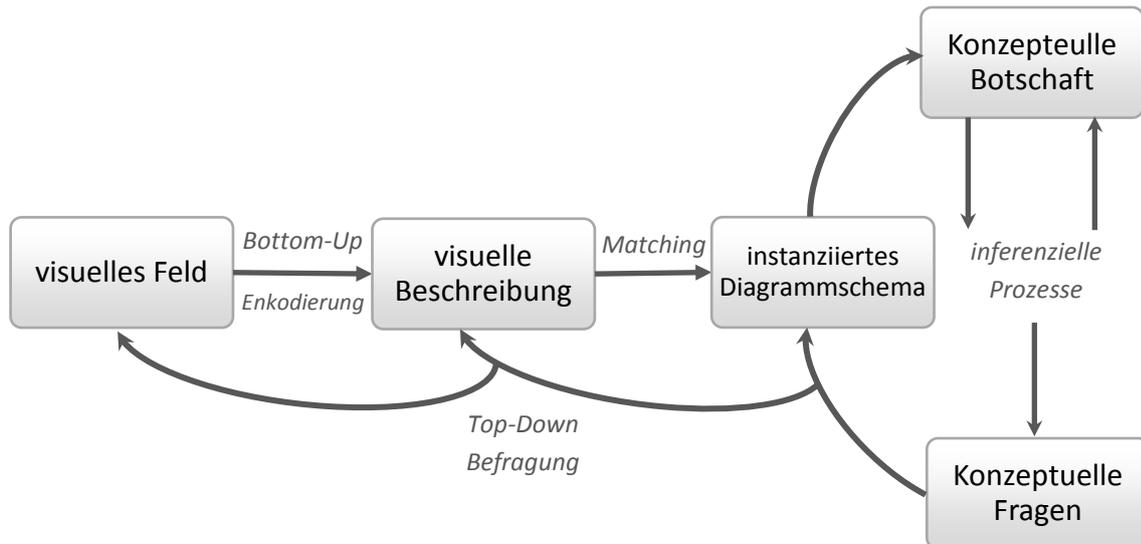


Abbildung 6. Modell des Diagrammverstehens nach Pinker (1990).

Im nächsten Schritt wird die visuelle Beschreibung parallel mit allen Schemata für visuelle Szenen abgeglichen. Bei diesem Matching-Prozess wird das auf die Szene am besten passende Schema ausgewählt und instanziiert. Bei der Betrachtung eines Diagramms muss außerdem das korrekte Subschema für den jeweiligen Diagrammtyp aktiviert werden (z. B. Balkendiagramm). Ein Schema enthält sowohl Elemente, die notwendigerweise vorhanden sein müssen (für ein Balkendiagramm beispielsweise Achsen und ein Koordinatensystem) als auch Leerstellen, die von der visuellen Beschreibung aufgefüllt werden können (z. B. die konkrete Bezeichnung der Achsen, die Anzahl der Balken und ihre jeweiligen Längen). Durch das Instanziiieren eines Schemas werden die entsprechenden Leerstellen aufgefüllt. Dies ermöglicht schließlich, dass eine *konzeptuelle Botschaft* (*conceptual message*) spezifiziert werden kann, die konkrete Informationen über das wahrgenommene Diagramm kodiert. Die konzeptuelle Botschaft kann man sich als eine Liste von Variablenpaaren (oder n -Tupeln bei komplexeren Diagrammen) vorstellen. Jedes Paar enthält die Ausprägung einer unabhängigen Variablen und den Wert einer korrespondierenden abhängigen Variablen. Gleichzeitig können Personen *konzeptuelle Fragen* formulieren, die man sich erneut als Variablenpaare (bzw. n -Tupel) vorstellen kann, bei denen jedoch der Wert einer Variablen leer ist. Das ist derjenige Wert, den eine Person durch Inspektion des Diagramms in Erfahrung bringen möchte. Der Prozess des Diagrammverstehens nach Pinker (1990) wird im Folgenden an einem Beispiel illustriert.

Beispiel 2. Gegeben sei ein Liniendiagramm mit Niederschlagsmengen im Jahr 2016 für die Monate Januar bis Dezember, getrennt für Hamburg und München (Abb. 7).

Kapitel 1 – Einleitung und theoretischer Rahmen

Bei der Betrachtung des Diagramms treffen zunächst die vom Papier (oder Bildschirm) reflektierten Lichtpartikel auf die Retina. Aus dem resultierenden Luminanzmuster (visuelles Feld) wird eine visuelle Beschreibung geformt, in der beispielsweise die Achsen bereits als senkrecht aufeinander stehende Linien repräsentiert werden. Aufgrund ihrer farblichen Codierung und der Verbindung durch Linien werden die Datenpunkte zu zwei perzeptuellen Einheiten gruppiert (türkise Linie und rote Linie). Während die Zahl von 24 Datenpunkten die Arbeitsgedächtniskapazität bei Weitem übersteigt, kann der ungefähre Verlauf von zwei Linien vergleichsweise leicht im Arbeitsgedächtnis gehalten werden.

Die unter diesen Beschränkungen geformte visuelle Beschreibung der Szene wird daraufhin mit den verfügbaren Schemata abgeglichen. Aufgrund der senkrecht aufeinander stehenden Achsen lässt sich das Bild schnell mit einem allgemeinen Schema für Achsendiagramme abgleichen. Das Erkennen der Linien im Achsendiagramm führt im Folgenden dazu, dass ein Subschema für Liniendiagramme gefunden und instanziiert wird. Ist die Betrachterin beziehungsweise der Betrachter mit Liniendiagrammen jedoch nicht vertraut, so kann der Leseprozess an diesem Punkt scheitern, weil kein Diagrammschema instanziiert werden kann.

Daraufhin können die Leerstellen des Schemas mit den konkreten Werten des vorliegenden Diagramms gefüllt werden. Beispielsweise, dass auf der Abszisse die Monate von Januar bis Dezember und auf der Ordinate die Niederschlagsmenge in Litern pro Quadratmeter abgetragen sind und dass die türkise Linie für Hamburg steht und die rote für München. Die daraus resultierende konzeptuelle Botschaft kann nun durch Variablen-Tripel spezifiziert werden, zum Beispiel:

- 1: $V_1 = \text{Stadt}$, $V_2 = \text{Monat}$, $V_3 = \text{Niederschlagsmenge}$
- 2: V_1 absoluter Wert = Hamburg, V_2 absoluter Wert = Juni, V_3 Level = hoch
- 3: V_1 absoluter Wert = Hamburg, V_2 Bereich = August – Oktober, V_3 Trend = fallend

Andersherum lassen sich von Personen entsprechende konzeptuelle Fragen formulieren, die sich auf ähnliche Weise repräsentieren lassen. Beispielsweise könnte sich eine Betrachterin oder ein Betrachter die Frage stellen, wie hoch der Niederschlag in München im Juli gewesen ist:

- 4: V_1 absoluter Wert = Hamburg, V_2 absoluter Wert = Juli, V_3 absoluter Wert = ?

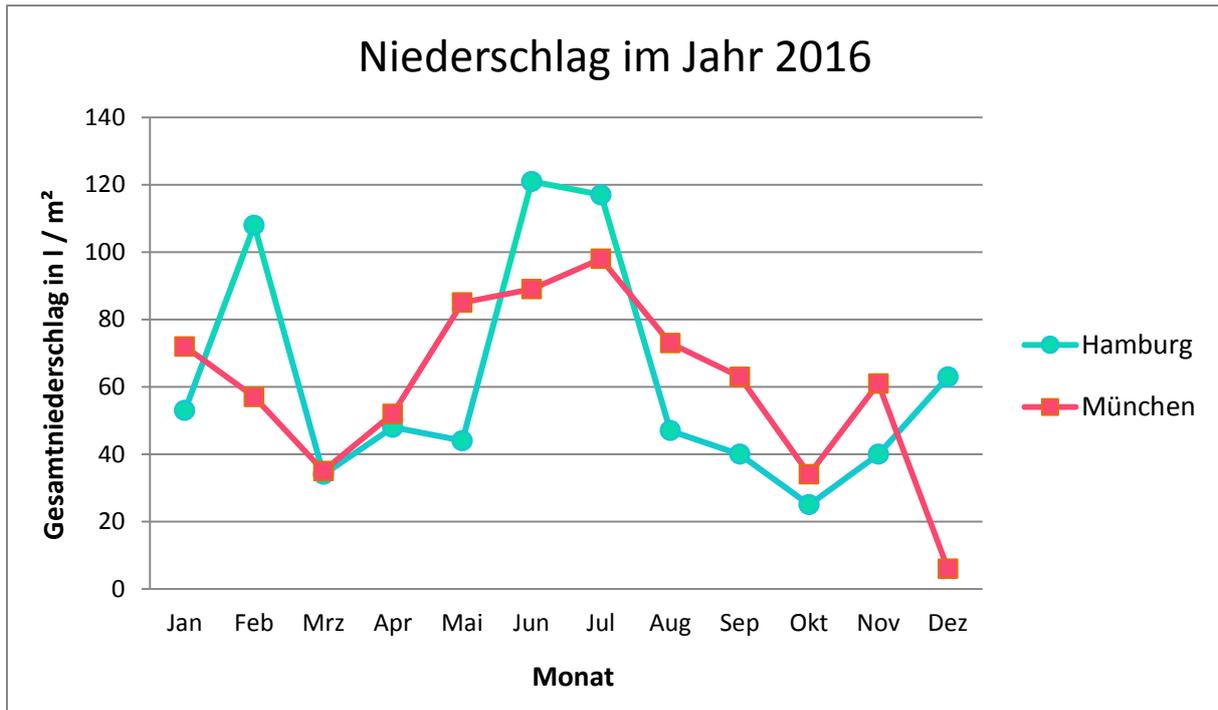


Abbildung 7. Beispieldiagramm: Niederschlag im Jahr 2016 in Hamburg und München

Wie bereits im Beispiel angedeutet, führen unter anderem Limitationen im Arbeitsgedächtnis dazu, dass zu jedem Zeitpunkt nur ein Teil der visuellen Beschreibung enkodiert wird. Die Elemente, die schließlich in der konzeptuellen Botschaft enthalten sind, sind in der Regel also diejenigen, die aus einem bestimmten Diagramm am einfachsten extrahiert werden können. Befindet sich eine Information nicht bereits in der konzeptuellen Botschaft, muss sie entweder durch Top-Down-Befragungsprozesse der konzeptuellen Botschaft hinzugefügt werden oder durch inferenzielle Prozesse aus existierenden Einträgen berechnet werden (vgl. Abb. 6). Beide Prozesse beinhalten eine Reihe von Computationen unter der Restriktion von limitierten kognitiven Ressourcen. Daher sind diese Prozesse zeitaufwändiger und fehleranfälliger (und werden schwieriger empfunden) als der simple Abruf von Informationen, die bereits in der konzeptuellen Botschaft enthalten sind. Diese Erkenntnis formuliert Pinker (1990) als das *Graph Difficulty Principle*:

„A particular type of information will be harder to extract from a given graph to the extent that inferential processes and top-down encoding processes, as opposed to conceptual message look-up, must be used” (S. 108).

Die Bedeutung limitierter kognitiver Ressourcen beim Lesen und Verstehen von Diagrammen wird in Abschnitt 1.2.3 vor dem Hintergrund der *Cognitive Load Theory* weiter vertieft.

1.2.2.3 Interaktionsmodell des Diagrammverstehens nach Freedman & Shah (2002)

In ihrem Interaktionsmodell betonen Freedman und Shah (2002) die Rolle des Vorwissens beim Lösen von Aufgaben mit Diagrammen. In diesem Modell steht, noch mehr als bei Pinker (1990), die Interaktion von Top-Down-Prozessen (beeinflusst durch Merkmale der Darstellung) und Bottom-Up-Prozessen (beeinflusst durch Vorwissen) im Vordergrund.

In Anlehnung an das *Construction-Integration-Modell* (CI-Modell) von Text- und Redeverständnis (Kintsch, 1988) definieren Freedman und Shah zwei Phasen des Diagrammverstehens: Eine *Konstruktionsphase* (*construction phase*) und eine Integrationsphase (*integration phase*). In der Konstruktionsphase wird zunächst die visuelle Information verarbeitet. Bestimmte visuelle Merkmale stoßen dabei die initialen Verarbeitungsprozesse an. Besonders relevant sind hierbei solche Merkmale, die beeinflussen, wie Informationen perzeptuell, und mental, gruppiert werden (z. B. Farbe und räumliche Nähe, siehe auch Abschnitt 1.3.2). Eine solche Informationseinheit nennen Freedman und Shah (2002) *Visual Chunk*. Wenn eine relevante, quantitative Information direkt aus einem einzelnen Visual Chunk abgelesen werden kann, genügen perzeptuelle und assoziative Prozesse, um die Information schnell und richtig aus dem Diagramm zu entnehmen.

In der Integrationsphase werden Informationen aus unterschiedlichen Quellen zu einer kohärenten mentalen Repräsentation integriert. Dieser Prozess gelingt mühelos, wenn relevante Informationen bereits auf perzeptueller Ebene identifiziert und mit Vorwissen verknüpft werden können. Fehlt Personen relevantes Vorwissen oder die nötigen Informationen lassen sich nicht direkt aus den visuellen Elementen des Diagramms ablesen, so ist der Leseprozess mühevoller und potenziell fehlerbehafteter. Diese Schlussfolgerung steht in Übereinstimmung mit Pinkers (1990) Graph Difficulty Principle, betont zusätzlich jedoch die Rolle des Vorwissens.

Konstruktion und Integration laufen, ähnlich wie beim Textverstehen, in einem alternierenden Zyklus ab, bei dem Personen zunächst eine Repräsentation von einem Teil des Diagramms (z. B. einer einzelnen Relation) konstruieren und diese dann mit dem Inhalt des Diagramms in Verbindung setzen, bevor sie den nächsten Teil verarbeiten (Freedman & Shah, 2002). Diese Annahme wird bereits durch empirische Befunde aus einer Eyetracking-Studie gestützt: Carpenter und Shah (1998) untersuchten das Blickbewegungsmuster von Personen

beim Lösen von Aufgaben mit Liniendiagrammen. Die Betrachterinnen und Betrachter identifizierten zunächst einzelne quantitative Datenpunkte und Relationen in der Abbildung und setzten diese anschließend mit dem Bezeichneten in Verbindung. Ihr iteratives Modell konnte in diesem Fall vorhersagen, wie die Fixationen sich über verschiedene Teile des Diagramms verteilen und wie viele Fixationen nötig sind, um bestimmte Aufgaben mit den Diagrammen zu lösen.

Vorwissen kann sich sowohl auf die Konstruktions- als auch auf die Integrationsphase auswirken. Unter Vorwissen werden hier drei Bereiche verstanden: domänenspezifisches Wissen (*domain knowledge*), Fähigkeiten im Diagrammlesen (*graph literacy skills*) und Fähigkeiten im Erklären (*explanatory skills*).

Domänenspezifisches Wissen kann beispielsweise die Wahrscheinlichkeit erhöhen, bestimmte Relationen zu interpretieren und Schlussfolgerungen daraus abzuleiten (Gattis & Holyoak, 1996). Es wird angenommen, dass die Wahrscheinlichkeit, mit der domänenspezifisches Wissen aktiviert wird, davon abhängt, wie gut erreichbar das Wissen ist. Das ist vor allem dann der Fall, wenn es häufig abgerufen wird. So geht man davon aus, dass domänenspezifisches Wissen bei Experten sogar automatisch aktiviert werden kann (Freedman & Shah, 2002). Vorwissen kann aber auch zu einem situationsabhängig unerwünschten Bias bei der Interpretation von Diagrammen führen. Beispielsweise neigen Personen bei der Darstellung von bekannten Zusammenhängen (z. B. Anzahl von Autounfällen und Anzahl alkoholisierter Fahrer) dazu, die erwarteten Trends zu berichten, auch wenn diese gar nicht aus den Daten abgelesen werden konnten. Bei Themen, mit denen die Personen weniger vertraut sind, scheint dieses Phänomen deutlich seltener aufzutreten (Freedman & Shah, 2002; Shah & Hoeffner, 2002). Dieser Effekt kann wiederum reduziert werden, wenn man Personen darauf hinweist, dass ihr Vorwissen ihre Einschätzungen beeinflussen kann (Sá, West, & Stanovich, 1999).

Fähigkeiten im Diagrammlesen in diesem Modell entsprechen Pinkers (1990) Konzept von Diagrammschemata. Diese Schemata enthalten das Vorwissen von Personen über Diagramme allgemein sowie über spezifische Diagrammtypen. Das Wissen über Diagramme wird gemäß dem CI-Modell früh in der Konstruktionsphase aktiviert und erleichtert die perzeptuelle Zusammenfassung von einzelnen Elementen zu Visual Chunks. Ähnlich wie das domänenspezifische Wissen kann auch Vorwissen über Diagramme zu einem Bias bei der Interpretation von Daten führen (z. B. Schiano & Tversky, 1992; Tversky & Schiano, 1989). Schülerinnen und Schüler im Grundschulalter neigen beispielsweise zur Verwechslung von

Steigung und Höhe, weil ihnen entsprechendes Vorwissen über Diagramme fehlt. (Bell & Janvier, 1981).

Zuletzt ist die Erklärung zugrundeliegender Mechanismen ein häufiges Ziel beim Lesen von Diagrammen (Freedman & Shah, 2002). So nehmen Oestermeier und Hesse (2000) beispielsweise an, dass die Fähigkeit, Erklärungen für Daten zu generieren auch von abstraktem Wissen über das Konzept von Kausalität abhängt. In einer Experten-Studie konnten Freedman & Shah (2001) zeigen, dass Experten häufiger in der Lage waren, spontane Erklärungen für ein Diagramm zu generieren als Novizen. Die Autoren gehen davon aus, dass Fähigkeiten im Erklären als Teil wissenschaftlichen Schlussfolgerns die Interpretation von Diagrammen auf ähnliche Weise beeinflussen können wie Vorwissen. Dies kann auch darauf zurückzuführen sein, dass entsprechendes Vorwissen bei Experten automatisch während der Konstruktionsphase aktiviert wird.

Freedman und Shah (2002) weisen schließlich auch auf die Notwendigkeit hin, zugrundeliegende Wirkmechanismen zu identifizieren: “the major challenge for knowledge-based approaches to graph comprehension is the need to specify the precise mechanisms through which prior knowledge and perceptual features interact with each other” (S. 28). Dennoch ist die Kenntnis um den Einfluss von Vorwissen beim Diagrammlernen für sich genommen bereits ein wertvoller Beitrag zur Untersuchung von Diagrammen.

1.2.3 Cognitive Load Theory

Diagramme enthalten per Konstruktion zahlreiche grafische Elemente (z. B. Achsen, Beschriftungen, Legenden und ein Koordinatensystem) und können darüber hinaus eine Vielzahl von Datenpunkten abbilden (siehe auch Abschnitt 1.3.2.5), die auf mehreren Ebenen der Informationsentnahme abgelesen und interpretiert werden können (siehe Abschnitt 1.2.1). Diese Menge an Elementen und Relationen kann eine hohe Belastung für das Arbeitsgedächtnis darstellen (Baddeley, 1994; Baddeley & Hitch, 1974).

Die *Cognitive Load Theory* (CLT) von Sweller (1988) beschreibt, wie die kognitive Verarbeitung von Informationen erschwert oder unterstützt werden kann, wenn man die stark begrenzte Arbeitsgedächtniskapazität des kognitiven Systems berücksichtigt. Dabei wird zwischen drei Typen kognitiver Belastung unterschieden (De Jong, 2010; Sweller, 1994): Unter *intrinsischer Belastung* (*intrinsic cognitive load*) versteht man die Art kognitive Belastung, die aus der inhärenten Komplexität einer Information resultiert, insbesondere die Anzahl der Elemente, deren Relationen zueinander verstanden werden müssen. *Extrinsische Belastung* (*extraneous cognitive load*) entsteht hingegen aus der Art und Weise, auf die

Informationen dargestellt und aufbereitet werden, beispielsweise durch die unnötige Verarbeitung von irrelevanten Informationen oder solchen Inhalten, die in keinem Bezug zum eigentlichen Thema stehen. Schließlich gibt es noch *lernbezogene Belastungen* (*germane cognitive load*). Damit sind solche Belastungen gemeint, die für einen Lernprozess notwendig sind, also der Aufwand, den Lernende betreiben müssen, um Inhalte zu verstehen. Dazu zählt beispielsweise das Interpretieren und Organisieren von Informationen.

1.2.3.1 Intrinsische kognitive Belastungen bei Diagrammen

Intrinsische kognitive Belastungen resultieren aus der inhärenten Komplexität des gegebenen Materials. In Diagrammen handelt es sich dabei in erster Linie um die zugrundeliegende Datenstruktur, die in der Abbildung repräsentiert wird, insbesondere die Anzahl der Datenpunkte und Variablen sowie deren Relationen (siehe auch Abschnitt 1.3.2.5). Eine hohe Datenkomplexität kann vor allem dann zu einer Belastung führen, wenn sie nicht für die Aufgabe geeignet ist, die mithilfe des Diagramms gelöst werden soll (Demetriadis & Cadoz, 2005; Huang, Eades, & Hong, 2009; Sedig & Parsons, 2013). Huang et al. (2009) beispielsweise variierten in ihrer Studie die Datenkomplexität und baten ihre Probanden, den mentalen Aufwand einzuschätzen, den sie zum Lösen der Aufgaben aufbringen mussten. Die Ergebnisse zeigen, dass mit dem Level der Datenkomplexität auch die subjektive Einschätzung der kognitiven Belastung anstieg. Ein Befund, der die Bedeutung von kognitiver Belastung beim Arbeiten mit Diagrammen unterstreicht.

An dieser Stelle ist jedoch wichtig zu bemerken, dass Datenkomplexität als Quelle kognitiver Belastung nur dann der intrinsischen Kategorie zuzuordnen ist, wenn entsprechenden Daten auch relevant für die zu bewältigende Aufgabe sind. Andernfalls ist sie eher den extrinsischen Belastungen zuzuordnen.

1.2.3.2 Extrinsische kognitive Belastungen bei Diagrammen

Extrinsische Belastungen resultieren aus Merkmalen der Gestaltung eines Materials. Damit steht diese Klasse kognitiver Belastungen in direktem Zusammenhang zur Wahl des Diagrammtyps und dem Einsatz visueller Merkmale wie Farbe und Beschriftungen (siehe Abschnitt 1.3.2). Beispielsweise kann der geeignete Einsatz von Labeln und farblicher Codierung die perzeptuelle Gruppierung von Elementen im Diagramm unterstützen (siehe z. B. Freedman & Shah, 2002; Pinker, 1990) und extrinsische Belastungen somit reduzieren. Andersherum können überflüssige Elemente zur Erhöhung der extrinsischen Belastung

beitragen, beispielsweise wenn Datenreihen oder Variablen für eine gegebene Aufgabe völlig irrelevant sind.

Ein weiteres Beispiel findet sich in der Multimedia-Forschung bei der Untersuchung von *Seductive Details*. Dabei handelt es sich um interessante, aber überflüssige Zusatzinformationen, die benutzt werden, um anderes Material aufzupeppen (Garner, Brown, Sanders, & Menke, 1992; Garner, Gillingham, & White, 1989). Seductive Details können das Interesse von Lernenden unter Umständen erhöhen (Harp & Mayer, 1997), typischerweise führen sie aber zu schlechteren Behaltensleistungen beim Lernen (z. B. Harp & Mayer, 1997; Lehman, Schraw, McCrudden, & Hartley, 2007; Rowland-Bryant et al., 2009). So finden sich in Zeitschriften, Büchern und im Internet häufig Diagramme, die um Texte und Abbildungen ergänzt werden, um zusätzliche Informationen unterzubringen oder das Material grafisch ansprechend zu gestalten (für Beispiele siehe Abb. 8). Da Diagramme räumliche Relationen als Analogie für nicht-räumliche Relationen verwenden, erscheinen diese visuellen Systeme besonders anfällig für externes Material, das zusätzlich verarbeitet werden muss. Bisher wurde der Einfluss von Seductive Details jedoch nicht im Kontext von Diagrammen untersucht.

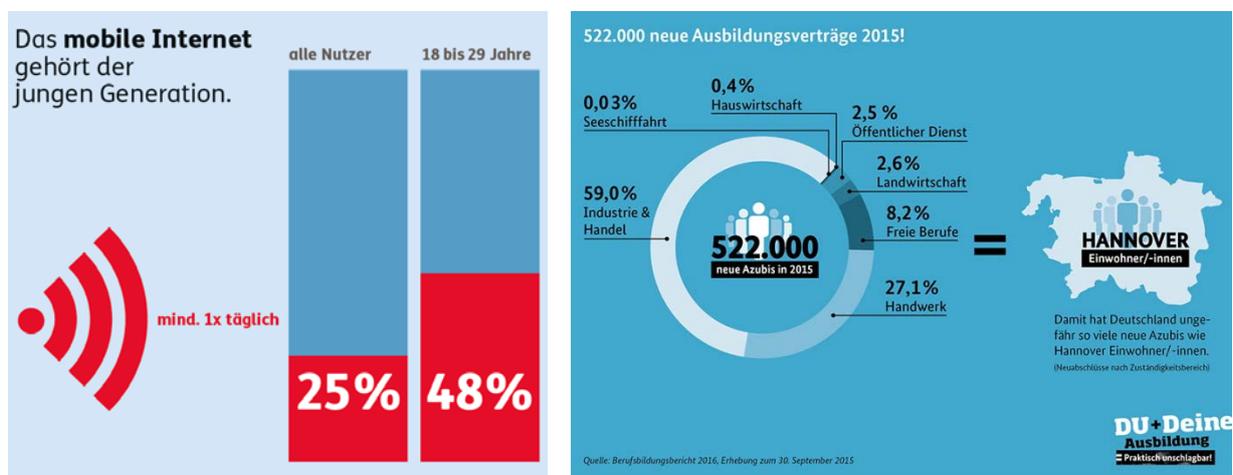


Abbildung 8. Beispiele für Diagramme, die zusätzliche Texte und Abbildungen enthalten. Links eine Abbildung aus einer Online-Umfrage der ERGO Direkt (2012), rechts eine Grafik aus dem Bildungsbericht 2016 des Bundesministeriums für Bildung und Forschung (2016).

1.2.3.3 Lernbezogene kognitive Belastungen bei Diagrammen

Lernbezogene Belastungen resultieren aus solchen Prozessen, die notwendig sind, um einen Sachverhalt zu verstehen. Ob und inwiefern eine kognitive Belastung beim Lesen von Diagrammen dieser Kategorie zuzuordnen ist, hängt davon ab, ob und welches Lernziel mit einer Aufgabe verbunden ist. Beispielsweise kann das mentale Ordnen von Daten mit hoher Komplexität als lernbezogene Belastung verstanden werden, wenn es ein Lernziel ist,

relevante von irrelevanten Informationen zu trennen. Steht demgegenüber eine inhaltliche Schlussfolgerung auf Basis einer konkreten Datenreihe im Vordergrund, wären Belastungen, die aus dem Filtern irrelevanter Informationen resultieren, nicht den lernbezogenen Belastungen zuzuordnen, sondern den extrinsischen.

1.3 Computationale Eigenschaften von Diagrammen

Die Beschäftigung mit unterschiedlichen Diagrammtypen und ihren visuellen Merkmalen ist deshalb von besonderer Relevanz, weil die Gestaltung einer Repräsentation ihre *computationalen Eigenschaften* mitbestimmt. Als *Computation* werden an dieser Stelle alle internen Informationsverarbeitungsprozesse der menschlichen Wahrnehmung und Kognition verstanden (vgl. Mausfeld, 2005). Die Arten und Weisen, auf welche visuelle Merkmale einer Abbildung ihre Verarbeitung im perzeptuellen und kognitiven System beeinflussen, werden als computationale Eigenschaften bezeichnet. Einfacher ausgedrückt beeinflussen die computationalen Eigenschaften einer Repräsentation wie einfach und schnell Personen sie verarbeiten und Inferenzen aus ihr ableiten können. Zahlreiche Studien in der psychologischen Diagrammforschung aus den letzten Jahrzehnten liefern empirische Evidenz dafür, dass computationale Eigenschaften von Diagrammen die Leistung in typischen Ableser-Aufgaben beeinflussen (e.g., Kim & Lombardino, 2015; Peebles & Cheng, 2003; Simkin & Hastie, 1987; Shah et al., 1999). Um beispielsweise für den Einsatz im Unterricht oder für einen Test zu entscheiden, ob eine Repräsentation eine geeignete Wahl ist oder welche von mehreren Repräsentationen für eine Aufgabe die beste Eignung hat, bieten computationale Eigenschaften deshalb ein wichtiges Entscheidungskriterium.

1.3.1 Informationale und Computationale Äquivalenz

Die computationalen Eigenschaften einer Repräsentation werden nicht immer isoliert betrachtet. Stattdessen steht der Vergleich mit anderen Repräsentationen häufig im Vordergrund einer Analyse (beispielsweise, um zu entscheiden, welche von mehreren Repräsentationsformen für die Kommunikation einer Information gewählt werden soll). Vergleicht man zwei Repräsentationen miteinander, so heißen sie *informationally equivalent*, wenn sie hinsichtlich der repräsentierten Informationen ununterscheidbar sind (Palmer, 1978). Ferner sind zwei Repräsentationen *computationally equivalent*, wenn sie informationally äquivalent sind und darüber hinaus jede Schlussfolgerung, die man auf Basis der einen Repräsentation ziehen kann, genauso einfach und schnell bei Betrachtung der anderen Repräsentation gezogen werden kann (Larkin & Simon, 1987).

Beispiel 3. Gegeben sei ein Datensatz zum Verlauf des Ölpreises in den Jahren 1990 bis 1999. Für jedes Jahr existiert ein Datum mit dem zugehörigen Ölpreis in USD pro Barrel. Auf Basis dieses Datensatzes werden zwei Repräsentationen erstellt, ein Fließtext, in dem zu jedem Jahr der zugehörige Ölpreis berichtet wird (z. B. „Im Jahr 1990 lag der Ölpreis bei 21 USD pro Barrel. Im nächsten Jahr, 1991, erreichte der Ölpreis 41 USD pro Barrel [...]“) und ein Säulendiagramm, in dem eine Säule für ein Jahr steht und die Höhe jeder Säule für den mittleren Ölpreis des Jahres (siehe Abb. 9). Gegeben sei die Aufgabe, den günstigsten Ölpreis in den 1990er Jahren zu identifizieren. Erhält eine Person den Fließtext, so muss sie diesen aufmerksam lesen, die Datenpunkte bis zum Jahr 1999 im Kopf behalten und mental miteinander vergleichen, um den niedrigsten Stand des Jahrzehnts zu identifizieren. Dieses Vorgehen stellt eine hohe Belastung des Arbeitsgedächtnisses dar, da zu jeder Zeit mehrere Datenpunkte kognitiv repräsentiert und mentale Operationen darauf ausgeführt werden müssen. In der Folge ist das Vorgehen möglicherweise fehleranfällig, da Datenpunkte vergessen werden oder eine mentale Operation das falsche Ergebnis liefern kann.

Erhält eine Person stattdessen das Säulendiagramm, lässt sich die gegebene Aufgabe ungleich leichter lösen. Es ist nicht nötig, die Höhe der Säulen in numerische Ölpreise zu überführen und diese mental zu vergleichen. Stattdessen kann man unter den ersten zehn Säulen perzeptuell die niedrigste identifizieren und den zugehörigen Preis ablesen, um die Aufgabe zu lösen. Da Diagramme intrinsische Repräsentationen sind, unterliegen die Relationen der visuellen Elemente im Bild denselben inhärenten Beschränkungen wie die Relationen des Bezeichneten („größer als“ im Diagramm ist ebenso wie „mehr als“ bei den Ölpreisen asymmetrisch und transitiv; siehe auch Abschnitt 1.1.1.1). Daher ist es möglich, allein durch Inspektion des Modells (d. h. des Säulendiagramms) zum richtigen Ergebnis zu gelangen.

Gemäß Pinkers (1990) Theorie des Diagrammverstehens ist anzunehmen, dass das Variablenpaar für die Höchste Säule und das dazugehörige Jahr (V_1 absoluter Wert = 2011, V_2 Level = Maximum) bereits in der konzeptuellen Botschaft enthalten ist, da die höchste Säule perzeptuell leicht identifiziert werden kann. Daher sind aufwändige inferenzielle Prozesse, wie etwa beim Durcharbeiten des Textes, zum Lösen der Aufgabe nicht notwendig (Graph Difficulty Principle; siehe auch Abschnitt 1.2.2.2).

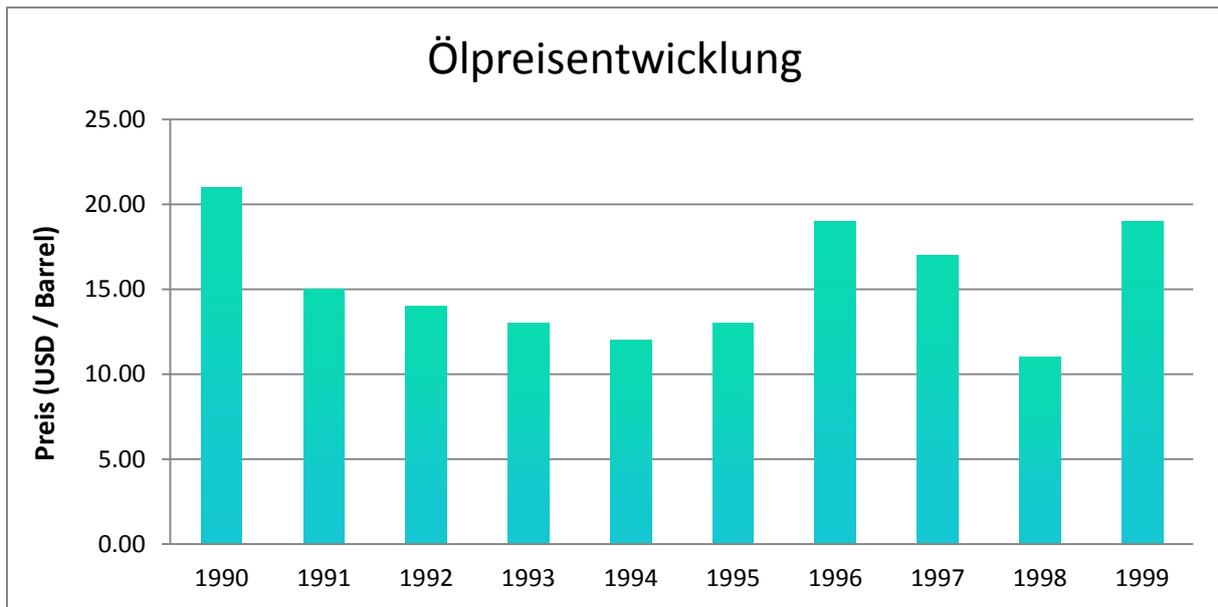


Abbildung 9. Beispieldiagramm: Ölpreisentwicklung in den 1990er Jahren

Wie in Beispiel 3 illustriert, können Personen beim Lesen von Diagrammen auf den räumlichen Dimensionen der Darstellung operieren, um Inferenzen auf die nicht-räumlichen Sachverhalte zu machen. Im Säulendiagramm lässt sich das Maximum einer Datenreihe ohne kognitiven Aufwand bestimmen, indem man die höchste Säule perzeptuell identifiziert. Liegen die Daten in einem Fließtext oder in einer Tabelle vor, lässt sich die Aufgabe nicht allein auf perzeptueller Ebene lösen. Der Lösungsprozess erfordert daher mehr kognitive Ressourcen, kann länger dauern und ist potenziell fehleranfälliger (siehe z. B. Pinker, 1990). Das heißt, für eine Klasse von Aufgaben, die sich auf perzeptueller Ebene ganz oder teilweise lösen lassen, können Diagramme einen computationalen Vorteil gegenüber anderen Repräsentationsformen bieten.

Diagramme unterscheiden sich computational nicht nur von anderen Repräsentationsformen wie Texten und Tabellen, sondern verschiedene Diagrammtypen unterscheiden sich auch untereinander (siehe Abschnitt 1.1.3). Simkin und Hastie (1987) verglichen beispielsweise drei informational äquivalente Diagrammtypen und fanden heraus, dass Bearbeitungszeit und Lösungsrichtigkeit in Ablese-Aufgaben von der Kombination aus Diagrammtyp und Anforderungen der Aufgabe abhingen. In einer neueren Studie konnten Peebles und Cheng (2003) zeigen, dass die computationalen Vorteile eines unbekanntes Diagrammtyps sogar die fehlende Vertrautheit damit kompensieren konnten. Die Probanden zeigten eine signifikante Verbesserung im Testverlauf und wiesen schließlich bessere Ergebnisse auf als mit einem vertrauten Diagrammtyp.

In den folgenden Abschnitten werden zunächst einzelne visuelle Merkmale von Diagrammen vor dem Hintergrund empirischer Befunde betrachtet (Abschnitt 1.3.2). Anschließend werden exemplarisch zwei der gebräuchlichsten Diagrammformate verglichen, nämlich Säulen- und Liniendiagramme (Abschnitt 1.3.3). Im Vordergrund dieser Betrachtung steht die Interaktion zwischen visuellen Gestaltungsmerkmalen und verschiedenen Aufgabentypen unter Berücksichtigung theoretischer Modelle des Diagrammverstehens (siehe Abschnitt 1.2).

1.3.2 Visuelle Merkmale von Diagrammen

Zusätzlich zum Diagrammtyp, der den groben Aufbau der Darstellung festlegt, gibt es eine Zahl von visuellen Merkmalen, die variiert werden können, um ein Diagramm weiter anzupassen. Dazu zählen beispielsweise die Anzahl der dargestellten Dimensionen, der Gebrauch von Farbe, Achsenbeschriftungen und Legenden, Dichte und Streckungsverhältnis, sowie Datenkomplexität (siehe z. B. Cleveland & McGill, 1987; Kim & Lombardino, 2015; Kumar & Benbasat, 2004; Shah et al., 2005; Shah & Hoeffner, 2002). Im Folgenden werden die wichtigsten visuellen Merkmale von Diagrammen vorgestellt und empirische Befunde zu ihrer Wirkung diskutiert.

1.3.2.1 Dimensionalität

Einige Diagrammtypen lassen sich sowohl zweidimensional als auch dreidimensional darstellen, darunter Streudiagramme, Liniendiagramme, Balken- und Säulendiagramme. Ein dreidimensionales Liniendiagramm kann man sich hierbei als *Gitterdrahtmodell* vorstellen (siehe Shah & Hoeffner, 2002; Shah, 2002). In diesen Darstellungen wird anstelle einer zweidimensionalen Fläche ein dreidimensionaler Raum genutzt, um Daten zu repräsentieren. Entsprechend werden solche Koordinatensysteme auch durch drei Achsen beschrieben. Wie im zweidimensionalen Fall kann auch hier jede der drei räumlichen Dimensionen zur Repräsentation einer Variablen genutzt werden.

Eine höhere Dimensionalität lässt sich in vielen Diagrammtypen auch auf zweidimensionaler Fläche erreichen, beispielsweise um drei oder mehr Variablen gleichzeitig darzustellen. Dies kann zum Beispiel durch räumliche Gruppierung von Säulen in Säulendiagrammen oder durch farbliche beziehungsweise texturierte Kennzeichnung von Linien, Säulen oder Balken und einer zugehörigen Legende in entsprechenden Diagrammtypen oder durch Label (siehe Abschnitt 1.3.2.3) erreicht werden.

Empirische Befunde zur Darstellung von Dimensionalität attestieren 3D-Diagrammen einen Vorteil gegenüber zweidimensionalen Repräsentationen (z. B. Kumar & Benbasat, 2004; Shah, 2002; Wickens, Merwin, & Lin, 1994). Kumar und Benbasat (2004) zeigten Probanden entweder zweidimensionale Liniendiagramme oder dreidimensionale Gitterdrahtmodelle und verglichen die Bearbeitungszeiten bei der Identifikation von Trends in den Daten. Personen waren in dieser Studie schneller in der Lage, die Aufgaben zu lösen, wenn ihnen ein 3D-Diagramm präsentiert wurde. Zudem war der Einfluss von hoher Datenkomplexität auf die Bearbeitungszeit signifikant geringer, wenn Personen die Aufgaben an Gitterdrahtmodellen bearbeitet haben (siehe auch Abschnitt 1.3.2.5). Ein Nachteil von dreidimensionalen Darstellungen ist jedoch oft, dass es zu Okklusionen kommen kann, sodass nicht jeder einzelne Datenpunkt erkennbar ist (Merwin, Vincow, & Wickens, 1994; Shah, 2002).

1.3.2.2 Farbe

Farbe kann in Diagrammen genutzt werden, um quantitative Informationen zu repräsentieren (z. B. je näher an rot, desto höher eine Temperatur oder je gesättigter, desto größer ein Wert) oder um kategoriale Daten zu unterscheiden, indem Elemente im Diagramm farblich gruppiert werden (z. B. unterschiedliche Farben für Männer und Frauen, deren Körpergrößen in einem Balkendiagramm abgetragen sind; Brockmann, 1991).

Während Cleveland und McGill (1985) zeigen konnten, dass Farbe quantitative Informationen nicht hinreichend gut wiedergeben und sogar irreführend sein kann, können farbliche Markierungen beim Gruppieren von Daten hilfreich sein (z. B. Kosslyn, 1994; Lewandowsky & Spence, 1989). Diese Art der Gruppierung geht auf Prinzipien perzeptueller Organisation zurück (Wertheimer, 1923) und ist ohne die Zutat von Aufmerksamkeit möglich (Moore & Egeth, 1997; Russell & Driver, 2005; Shomstein, Kimchi, Hammer, & Behrmann, 2010).

Eine Spezialform dieser Vorgehensweise stellen semantisch verknüpfte Farben dar (z. B. rot für wärmere Temperaturen, blau für kältere Temperaturen). Für gewöhnlich müssen Personen zunächst Legenden oder Label inspizieren, um Farben mit den bezeichneten Inhalten zu verknüpfen. Shah und Hoeffner (2002) weisen darauf hin, dass die Verwendung von semantisch verknüpften Farben das Memorieren erleichtern und das Arbeitsgedächtnis dadurch entlasten kann. Dies funktioniert allerdings nur unter der Annahme, dass Personen, die das Diagramm betrachten, dieselben Annahmen über die Bedeutung der farblichen Kodierung teilen wie die Person, die das Diagramm erstellt hat. Die Bedeutung von Farben ist

jedoch nicht universell, sondern hängt unter anderem von Profession und kulturellen Konventionen ab. Beispielsweise kann die Farbe Grün im Finanzwesen für „profitabel“ stehen, während sie im Gesundheitssektor typischerweise die Bedeutung „infiziert“ trägt (Brockmann, 1991). Aufgrund dieser engen Geltungsbereiche erscheint der Einsatz semantisch verknüpfter Farben in Diagrammen als eher unzuverlässiges Mittel, um die Verarbeitung von Diagrammen effizienter zu gestalten.

1.3.2.3 Legenden und Beschriftung

Obwohl es sich bei Diagrammen grundsätzlich um depiktionale Repräsentationen handelt (siehe Sektion 1.1.1.1), enthalten sie auch deskriptionale Elemente in Form von Text. Textelemente werden üblicherweise für Überschriften, Achsenbeschriftungen und Legenden verwendet. Während sich Relationen zwischen den Variablen intrinsisch durch räumliche Relationen repräsentieren lassen, lässt sich die durch eine Säule repräsentierte Kategorie oder die Menge möglicher Kategorien nicht aus Form, Höhe oder Breite der Säule verstehen. Die Bedeutung einer Säule ist zunächst arbiträr. Dies zeigt sich auch darin, dass in einem anderen Diagramm eine Säule für eine völlig andere Kategorie stehen kann. Diese deskriptionalen Elemente bezeichnet Bertin (1983) als *Invarianten*. Das heißt, sie können nicht mit anderen vertauscht werden, ohne dass sich gleichzeitig der Sinn des Diagramms ändert. Demgegenüber können die Werte einer *Komponente* (Variable) variieren ohne dass sich das Thema der Darstellung ändert.

Die Verknüpfung des Bezeichneten (z. B. ein bestimmtes Jahr) mit dem Bezeichnenden (z. B. eine Säule) wird wiederum durch die Struktureigenschaften des Diagramms ermöglicht. So spricht man beispielsweise von *Labels*, wenn der Text mit dem Bezeichneten in räumlicher Nähe zum Bezeichneten (der Säule) steht und auf diese Weise perzeptuell assoziiert wird (Gesetz der Nähe; siehe Wertheimer, 1923). Bei Legenden wird die Verknüpfung hingegen durch das Prinzip der Ähnlichkeit gewährleistet, insbesondere durch Farbe (bei Linien, Säulen und Balken; siehe auch Abschnitt 1.1.3), Texturen (bei Säulen und Balken) oder Form (bei Punkten). Obwohl Legenden häufig eingesetzt werden, können sie das Arbeitsgedächtnis stark belasten, da Personen die Verknüpfung von Farbe und Bedeutung memorieren müssen (Kosslyn, 1994; Shah & Hoeffner, 2002). Konkret bedeutet dies beispielsweise, wie Carpenter und Shah (1998) in einer Eyetracking-Studie zeigen konnten, dass Personen häufig auf die Legende zurückschauen müssen, um ihr Gedächtnis aufzufrischen. Personen können Informationen schneller aus Diagrammen ablesen, wenn Label anstatt Legenden verwendet werden (Shah et al., 2005). Gibt es allerdings viele

Datenpunkte oder zahlreiche Datenreihen, kann es durch Label zu Okklusionen kommen, die das Ablesen der Daten erschweren (Kosslyn, 1994). Demnach ist im Einzelfall abzuwägen, welche Form für ein konkretes Diagramm günstig ist.

1.3.2.4 Größe und Skalierung

Visuelle Merkmale wie die Größe eines Diagramms und die daraus resultierende Dichte der Datenpunkte (insbesondere in Streudiagrammen) können einen Einfluss darauf haben, wie ein Datensatz interpretiert wird. Beispielsweise überschätzen Personen die Stärke von Korrelationen in Streudiagrammen, wenn die perzeptuelle Dichte der Datenpunkte besonders hoch ist (Cleveland, Diaconis, & McGill, 1982; Lauer & Post, 1989). Eine höhere wahrgenommene Dichte der Datenpunkte kann entweder dadurch zustande kommen, dass Datenpunkte hinzugefügt werden, oder indem das Diagramm verkleinert wird (Shah et al., 2005).

1.3.2.5 Datenkomplexität

Unter Datenkomplexität wird hauptsächlich die Anzahl dargestellter Datenpunkte und Variablen in einem Diagramm verstanden. Meyer, Shinar und Leiser (1997) unterscheiden zwischen drei Faktoren, die Datenkomplexität beeinflussen: (a) Die Anzahl von Datenpunkten in einem Diagramm, (b) die Konfiguration der Datenpunkte (d. h. die Organisation der Datenpunkte zu Datenreihen) und (c) die Regelmäßigkeit (*regularity*) der Daten (d. h. ob die Daten geordnet sind und klare Trends zeigen oder ob sie kein erkennbares Muster ergeben). Die Datenkomplexität stand nur selten im Fokus der psychologischen Erforschung von Diagrammen, wurde aber vielfach als Hintergrundvariable untersucht (z. B. Casali & Gaylin, 1988; Meyer et al., 1997; Schutz, 1961a, 1961b; Spence & Lewandowsky, 1991). Jüngere Studien haben allerdings gezeigt, dass eine hohe Datenkomplexität in Diagrammen die Leistung bei typischen Ablese-Aufgaben negativ beeinflusst (z. B. Kim & Lombardino, 2015; Kumar & Benbasat, 2004).

Daten aus Diagrammen abzulesen kann für Personen besonders dann herausfordernd sein, wenn die Datenkomplexität nicht angemessen für die gegebene Aufgabe ist (Huang et al., 2009; Demetriadis & Cadoz, 2005; Sedig & Parsons, 2013). Da der menschlichen Informationsverarbeitung nur begrenzte Kapazitäten zur Verfügung stehen (Baddeley & Hitch, 1974), kann es durch die Verarbeitung zahlreicher Elemente und ihrer Relationen zu einer kognitiven Überbelastung kommen, die sich in schlechterer Leistung niederschlägt (Sweller, 1994).

Schutz (1961a, 1961b) war einer der ersten Forscher, der den Einfluss von der Datenpunktzahl in Diagrammen auf die Ableseleistung untersucht hat. Er betrachtete die Leistung von Personen beim Identifizieren von Trends und fand heraus, dass die Bearbeitungszeit mit der Anzahl von Datenpunkten stieg, während die Lösungsrichtigkeit sich nicht signifikant änderte. Eine Untersuchung von Kumar und Benbasat (2004) fand ebenfalls erhöhte Bearbeitungszeiten, wenn zusätzliche Datenpunkte in Diagramme integriert wurden. Darüber hinaus untersuchten Spence und Lewandowsky (1991) Ablese-Aufgaben, bei denen mehrere Datenpunkte miteinander verglichen werden mussten. Im Gegensatz zu Schutz fanden sie auch eine signifikante Verschlechterung der Lösungsrichtigkeit, wenn mehr Datenpunkte zu sehen waren.

In Bezug auf die Konfiguration von Datenpunkten fand Schutz (1961b) heraus, dass die Bearbeitungszeit mit der Anzahl von Datenreihen stieg. Dies galt jedoch nur, wenn die Datenreihen in mehreren Diagrammen präsentiert wurden (d. h. eine Datenreihe pro Diagramm), nicht jedoch wenn alle Datenreihen im gleichen Diagramm abgebildet waren. In einer Eyetracking-Studie variierten Kim und Lombardino (2015) die Anzahl von Datenreihen innerhalb eines Diagramms sowie den Aufgabentyp (d. h. Ablesen einzelner Datenpunkte vs. Vergleich mehrerer Datenpunkte). Sie fanden heraus, dass die Bearbeitungszeit unabhängig vom Aufgabentyp größer war, wenn eine zusätzliche Datenreihe ins Diagramm integriert wurde. Dies zeigte sich auch in längeren Fixationszeiten auf den Diagrammen.

Befunde zur Regelmäßigkeit von Daten als Quelle von Datenkomplexität sind weniger eindeutig. Meyer et al. (1997) zeigten ihren Probanden Datensätze mit je drei Datenreihen, die entweder den gleichen Trend (= hohe Regelmäßigkeit) zeigten oder unterschiedliche Trends (= geringe Regelmäßigkeit). Sie fanden, dass die Bearbeitungszeit leicht erhöht war, wenn Trends identifiziert werden sollten, nicht jedoch, wenn einzelne Datenpunkte miteinander verglichen werden mussten. Die Lösungsrichtigkeit war jedoch für beide Aufgabentypen geringer, wenn die Daten unregelmäßig waren. Im Vergleich dazu fanden Kumar und Benbasat (2004) keine Effekte von Datenkomplexität, wenn diese durch die Regelmäßigkeit der Daten operationalisiert wurde.

1.3.3 Säulen- und Liniendiagramme im Vergleich

Auch wenn Diagramme einige Aufgaben erleichtern können, bieten sie keine universelle kognitive Unterstützung. Ob ein Diagramm in einer konkreten Situation computationale Vorteile bietet, hängt sowohl von der Darstellungsform (Diagrammtyp und visuelle Merkmale) als auch von den Anforderungen der Aufgabe ab (siehe z. B. Peebles &

Cheng, 2003; Simkin & Hastie, 1987; Shah et al., 1999). Shah und Hoeffner (2002) resümieren in ihrem Literaturüberblick beispielsweise, dass Säulendiagramme besser geeignet sind, um einzelne Datenpunkte zu vergleichen, während Liniendiagramme hilfreicher sind, wenn es darum geht, Trends in Daten zu identifizieren. In seiner Theorie des Diagrammverstehens (*Theory of Graph Comprehension*) liefert Pinker (1990) eine differenzierte Analyse der computationalen Unterschiede zwischen Säulen- und Liniendiagrammen. Er führt an, dass Liniendiagramme die Identifikation von Trends befördern, da in diesem Diagrammtyp ein Trend aus einem einzelnen visuellen Attribut (d. h. an der Steigung einer Linie) einer einzelnen perzeptuellen Einheit (Linie) aus der Darstellung abgelesen werden kann (siehe Abbildung 10b). Gemäß Pinkers (1990) Theorie des Diagrammverstehens ist anzunehmen, dass Trends bereits in der konzeptuellen Botschaft enthalten sind, da der Verlauf einzelner Linien perzeptuell leicht identifiziert werden kann (vgl. Abschnitt 1.2.2.2). Demgegenüber lassen sich Trends in Säulendiagrammen nicht anhand einer einzigen Einheit ablesen. Stattdessen müssen mehrere Attribute (Höhe der Säulen) von mehreren perzeptuellen Einheiten (Säulen) berücksichtigt werden, um einen Trend zu konstituieren, was die Identifikation von Trends in Säulendiagrammen erschwert (siehe Abbildung 10a). Es ist daher weniger wahrscheinlich, dass Trends bereits in der konzeptuellen Botschaft enthalten sind. Es sind stattdessen aufwändige inferenzielle Prozesse, notwendig (Graph Difficulty Principle; siehe auch Abschnitt 1.2.2.2).

Geht es stattdessen darum, Datenpunkte miteinander zu vergleichen, bieten Säulendiagramme einen Vorteil, da einzelne Datenpunkte auch einzelnen perzeptuellen Einheiten entsprechen (eine Säule pro Datum). Beispielsweise könnte man die Investitionen in den Klimaschutz für zwei Jahre vergleichen um festzustellen, für welches Jahr die Investitionen höher waren. Im Säulendiagramm entspricht die horizontale Position der Säulen einem Wert auf der einen Skala (hier: Jahre) und die Höhe der Säulen entspricht einem Wert auf einer anderen Skala (hier: Investition in Mio. Euro). Personen können ausgewählte Jahre leicht identifizieren, da sie durch distinkte perzeptuelle Einheiten (Säulen) repräsentiert werden (siehe Abbildung 11a). Dieselbe Aufgabe ist indessen schwieriger mit Liniendiagrammen zu lösen, da die perzeptuelle Organisation in Linien dazu führt, dass nicht einzelne Datenpunkte als perzeptuelle Einheiten kodiert werden, sondern die Zusammenfassung mehrerer Punkte zu einer Linie (Pinker, 1990; Wertheimer, 1923). Personen müssen einzelne Datenpunkte mental aus den Linien extrahieren, um diese miteinander zu vergleichen (siehe Abbildung 11b).

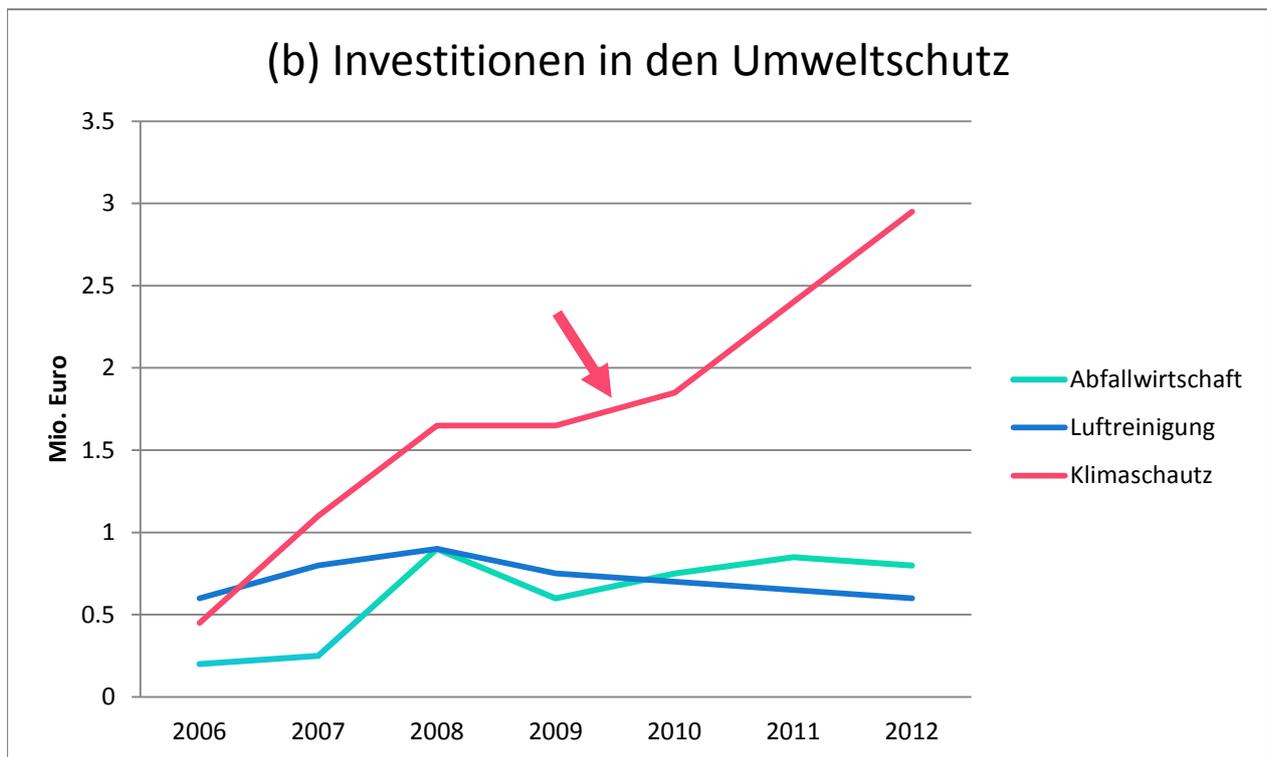
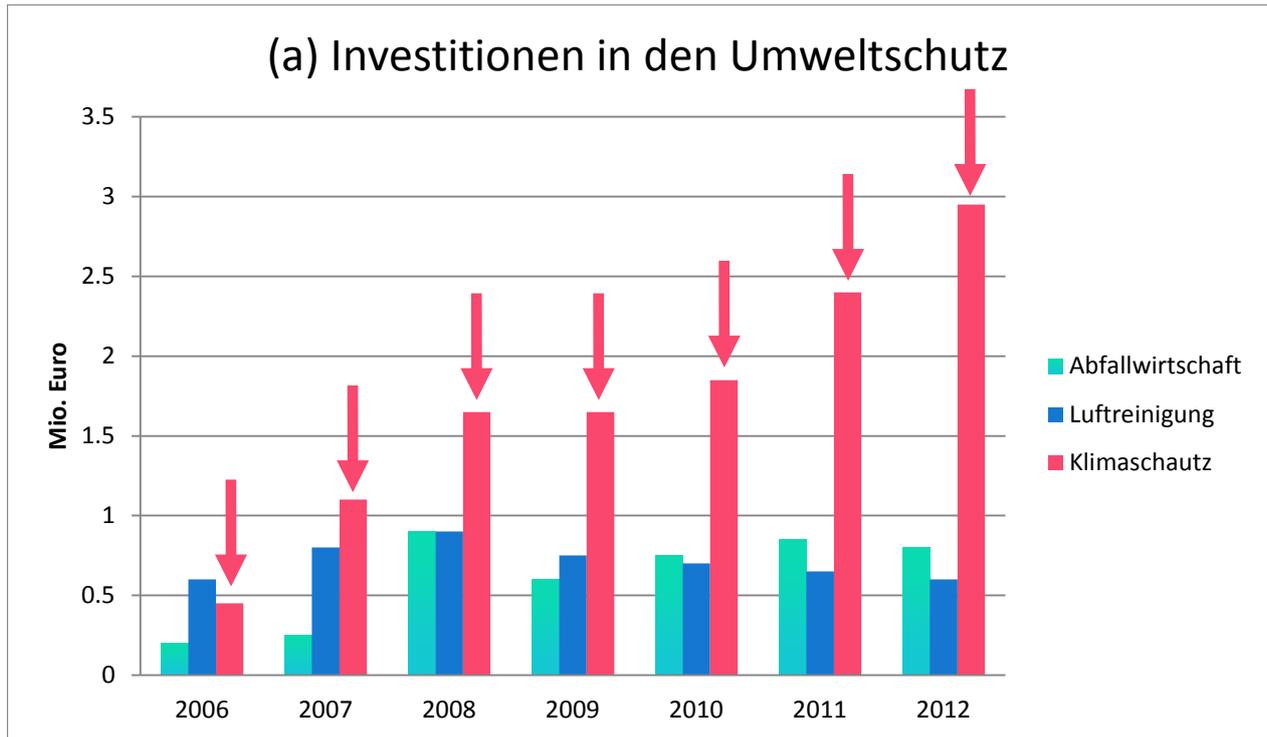


Abbildung 10. Perzeptuelle Einheiten (mit Pfeil markiert), die beim Ablesen von Trends in (a) Säulendiagrammen und in (b) Liniendiagrammen berücksichtigt werden müssen

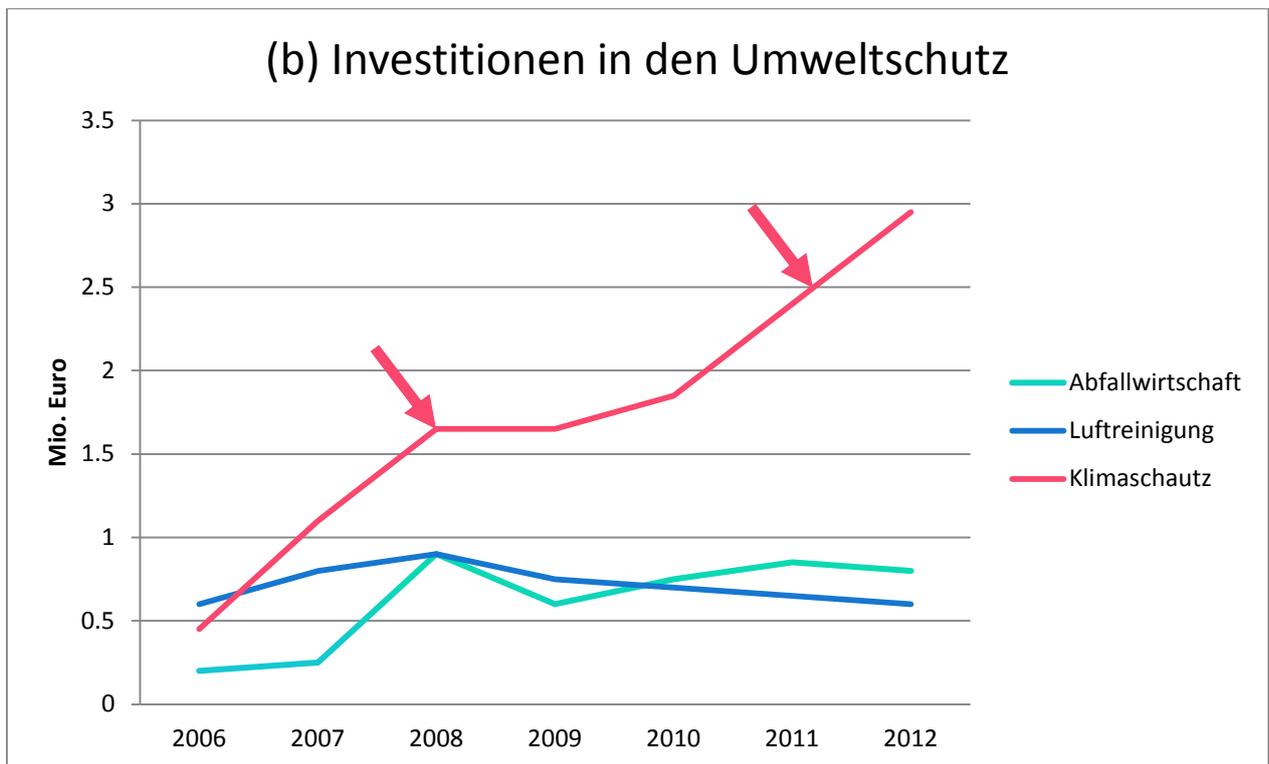
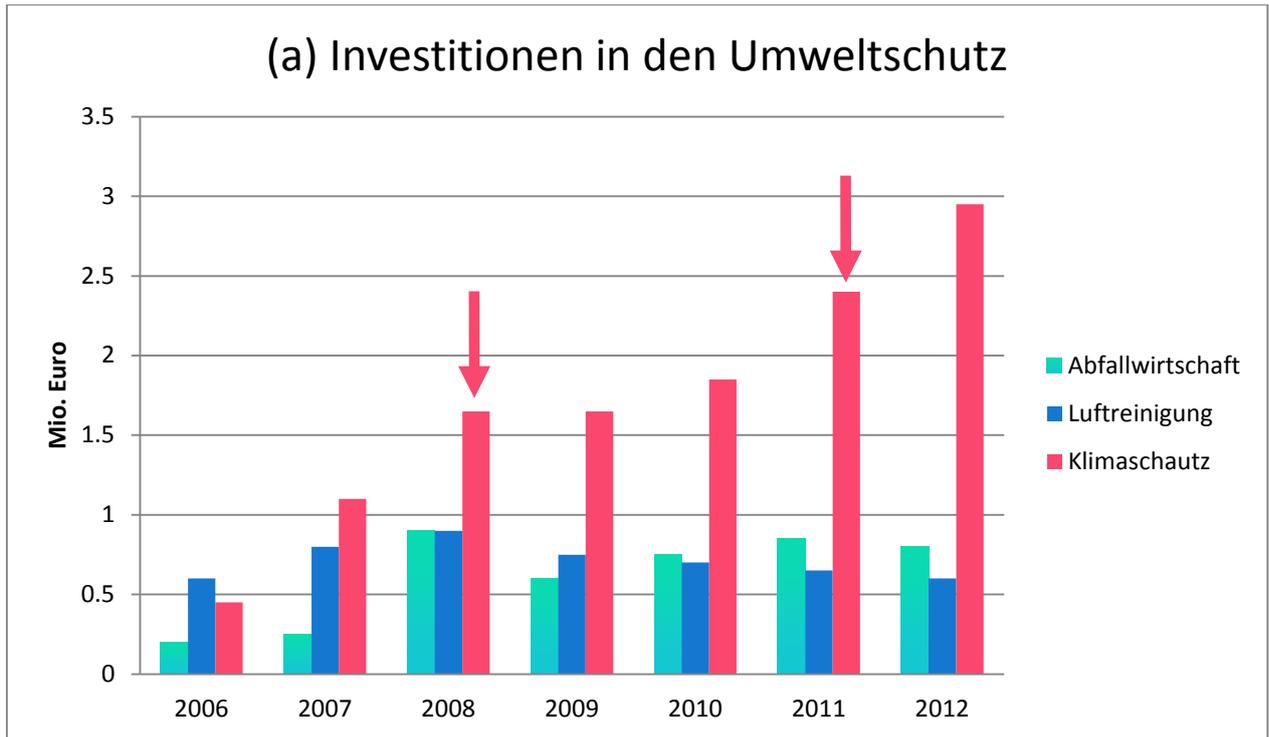


Abbildung 11. Perzeptuelle Einheiten (mit Pfeil markiert), die beim Ablesen von Punktvergleichen in (a) Säulendiagrammen und in (b) Liniendiagrammen berücksichtigt werden müssen

1.4 Eyetracking als Untersuchungsmethode

Eyetracking ist eine Methode in der kognitiven Psychologie und anderen Disziplinen, um die raumzeitliche Verteilung von Blickbewegungen zu studieren (für Einleitungen zum Thema Eyetracking siehe z. B. Duchowski, 2007; Holmqvist et al., 2011). Im Gegensatz zu typischen Leistungsmaßen wie Fehlerquoten und Bearbeitungszeiten, handelt es sich bei der Blickbewegung um ein (kontinuierliches) Prozessmaß. Blickbewegungsmuster werden hierbei als Indikator für zugrundeliegende kognitive Prozesse genommen. Die Eyetracking-Forschung stützt sich insbesondere auf die so genannte *Eye-Mind-Hypothesis* (Just & Carpenter, 1980). Dabei handelt es sich um die Annahme, dass der Fixationspunkt unseres Blicks auch den Fokus unserer Aufmerksamkeit widerspiegelt. Das bedeutet, wir verarbeiten kognitiv jene visuellen Informationen, die wir gerade betrachten. Obwohl die Eye-Mind-Hypothesis auch kritisch diskutiert wurde (siehe z. B. Hyönä, 2010; Wright & Ward, 2008), konnten zahlreiche Studien zeigen, dass Blickbewegungsmaße und kognitive Leistungen einen engen Zusammenhang aufweisen (z. B. Canham & Hegarty, 2010; Jarodzka, Scheiter, Gerjets, & van Gog, 2010; Lindner, Eitel, Strobel & Köller, 2017). Eine wichtige Voraussetzung dafür ist allerdings, dass die Verarbeitung visueller Informationen relevant für die Aufgabe ist, die bearbeitet werden soll (Mayer, 2010; Hyönä, 2010).

Mayer (2010) sieht in Eyetracking einen wertvollen Beitrag zur Erforschung von Lernen mit Abbildungen. Er führt an, dass Eyetracking-Maße Informationen bereitstellen können, die für die Gestaltung von Lehrmaterialien besonders relevant sind. In Experimenten kann man typischerweise Schlussfolgerung darüber ableiten, ob und wann die Manipulation eines Materials einen Effekt auf Leistungsmaße hat. Dabei bleibt häufig jedoch ungeklärt, wie die Verarbeitung des Materials sich durch die Manipulation verändert und welche Prozesse dem Effekt zugrunde liegen. Beispielsweise können Eyetracking-Daten für bekannte Effekte (z. B. den Seductive-Details-Effekt; siehe Abschnitt 1.2.3) neue Erklärungsansätze liefern oder bestehende validieren. Wissen darüber, wie Personen bestimmte Materialien verarbeiten, kann außerdem dazu beitragen, sie für den Einsatz in Lehr- und Lernkontexten zu verbessern (van Gog & Scheiter, 2010). Im Vergleich zu anderen Prozessmaßen (z. B. Lautes Denken; Ericsson & Simon, 1980; Konrad, 2010; Sandmann, 2014; van Gog, Kester, Nievelstein, Giesbers, & Paas, 2009) belastet Eyetracking das Arbeitsgedächtnis jedoch nicht zusätzlich, da videobasierte Eyetracking-Systeme unaufdringlich sind und Personen sich bei der Bearbeitung von Aufgaben natürlich verhalten können (Hyönä, 2010).

1.4.1 Eyetracking zur Untersuchung von Diagrammen

Die Erfassung von Blickbewegungen eignet sich vornehmlich zum Studium von visuellen Aufgaben, beispielsweise beim Lernen von Texten, Abbildungen und Animationen (Hyönä, 2010). Dies gilt für Diagramme in besonderem Ausmaß, da sie nicht nur visuell dargeboten werden, sondern auch nach hochgradig konventionalisierten, räumlichen Darstellungsprinzipien aufgebaut sind. So ist es Forscherinnen und Forschern gelungen, die Bearbeitungszeit bei Diagrammleseaufgaben zuverlässig der Verarbeitung verschiedener Subregionen zuzuordnen (z. B. Carpenter & Shah, 1998; Goldberg & Helfman, 2011; Kim & Lombardino, 2015; Kim, Lombardino, Cowles, & Altmann, 2014; Peebles & Cheng, 2003). Hierzu werden die Diagramme (und die zugehörigen Teile der Aufgabe) in verschiedene *Areas of Interest* (AoI) unterteilt. Dabei handelt es sich um Regionen, die in einem Stimulus definiert werden, um Blickdaten zusammenzufassen. AoIs müssen allerdings nicht alle Elemente in einer Abbildung abdecken. Es wird stattdessen empfohlen, nur tatsächlich interessierende Bereiche mit AoIs zu versehen (Goldberg & Helfman, 2010). In Diagrammen bietet es sich an, den verschiedenen funktionalen Bereichen AoIs zuzuordnen, beispielsweise den Achsen, Legenden, Titeln sowie Fragen und Antworten. In Abbildung 12 findet sich eine Übersicht der AoI-Aufteilungen in verschiedenen Diagrammstudien.

Das menschliche Auge scannt visuelle Szenen (z. B. eine Diagrammaufgabe am Computer), indem es eine Folge von schnellen Augenbewegungen (Sakkaden) macht, gefolgt von längeren Verweilzeiten (Fixationen). Während der Fixationen werden Informationen aufgenommen, Sakkaden dienen dazu, den Punkt der visuellen Aufmerksamkeit zu einer neuen Region zu verschieben. Je nach Art der Aufgabe, dauert eine Fixation etwa 200 bis 250 Millisekunden (siehe auch Hyönä, 2010). Von einem Eye-Tracker aufgezeichnete Blickbewegungsdaten liegen zunächst in einem Rohformat vor, das anschließend algorithmisch in Sakkaden und Fixationen aufgebrochen wird. Dabei wird häufig die Fixationsdauer auf festgelegten AoI als Maß herangezogen. Die Gesamtfixationszeit auf eine vordefinierte AoI ist definiert als die kumulative Dauer aller Fixationen, die während eines Trials in die jeweilige AoI gefallen sind (Holmqvist et al., 2011). Sie bieten somit ein Maß dafür, wie Personen ihre visuelle Aufmerksamkeit während einer Bearbeitungsphase verteilen (Hyönä, 2010).

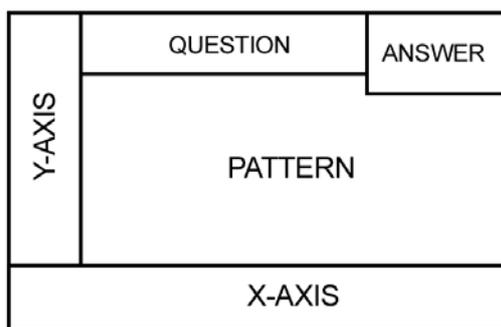
Gesamtfixationszeiten auf vordefinierte AoI haben sich bei der Aufgabenbearbeitung in pädagogischen Kontexten als valides Maß für die Aufmerksamkeitsverteilung und kognitive Verarbeitung erwiesen (siehe z. B. Lindner et al., 2014). Auch in der Diagrammforschung war diese Vorgehensweise fruchtbar. Bereits Shah (1998) verwendete

Kapitel 1 – Einleitung und theoretischer Rahmen

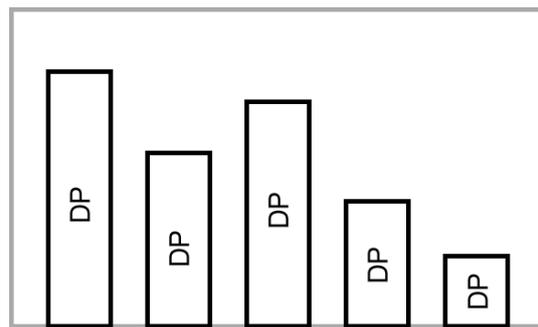
aggregierte Blickbewegungen auf vordefinierte AoI. Die Verteilung, Anzahl und Dauer der Fixationen in dieser Studie waren konsistent mit den postulierten Modellannahmen, dass Diagrammlesen Mustererkennung, Ablesen von Relationen und das Interpretieren der Zusammenhänge anhand ihrer Bezeichnungen beinhaltet.

In einer jüngeren Studie konnten Kim und Lombardino (2015) mithilfe von Fixationszeiten zeigen, dass Personen mehr Zeit mit der Betrachtung von Diagrammregionen verbracht haben, wenn sie eine höhere Datenkomplexität aufwiesen. Dies war sowohl für einfache Ablese-Aufgaben der Fall als auch für komplexere Vergleichsaufgaben (vgl. Abschnitt 1.2.1). Mit ähnlichen Methoden konnte die Arbeitsgruppe auch zeigen, dass Schülerinnen und Schüler mit Legasthenie nicht nur linguistische Informationen (Fragen, Antworten) länger verarbeiten, sondern auch nicht-linguistische (d.h. die Diagrammregionen), jedoch nicht bei erstmaliger Betrachtung, sondern nur bei wiederholter Prüfung der entsprechenden Regionen (Kim et al., 2014).

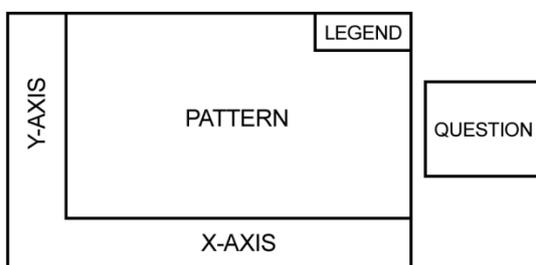
Peebles & Cheng (2003)



Goldberg & Helfman (2010)



Kim & Lombardino (2015)



Strobel, Lindner, Saß & Köller (2018)

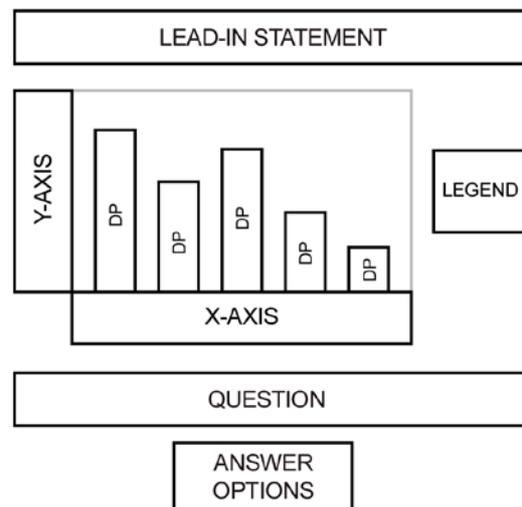


Abbildung 12. Schematische Übersicht über die AoI-Aufteilungen in einer Auswahl unterschiedlicher Diagrammstudien. AoI = Area of Interest, DP = Data Point.

Eyetracking eignet sich zur Untersuchung von Diagrammen aus mehrfachen Gründen: Zum einen bietet die Methode im Allgemeinen die Möglichkeit, typische Leistungsmaße (Fehlerquote, Bearbeitungszeit) um Prozessdaten zu ergänzen, denn die Methode ist non-invasiv und führt zu keiner zusätzlichen Belastung des Arbeitsgedächtnisses. Die Abbildung von Bearbeitungsprozessen kann wiederum Anhaltspunkte liefern, warum und auf welche Weise ein Effekt auf Leistungsmaße zustande kommt. Darüber hinaus können auch Änderungen im Bearbeitungsprozess identifiziert werden, die in ergebnis-orientierten Maßen nicht sichtbar werden. Zum anderen ist Eyetracking im Besonderen dazu geeignet, Bearbeitungsprozesse bei Diagrammleseaufgaben offenzulegen. Da in Diagrammen räumliche Strukturen mit funktionalen Strukturen in hoher Übereinstimmung stehen, lassen sich Blickbewegungsmuster den zugrundeliegenden Arbeitsprozessen mit hoher Konfidenz zuordnen. Studienergebnisse, die in diesem Abschnitt berichtet wurden, liefern darüber hinaus auch empirische Evidenz für die Eignung der Methode.

1.5 Forschungsfragen

Ziel dieser Forschungsarbeit ist es, den Einfluss visueller Gestaltungsmerkmale auf kognitive Prozesse bei der Bearbeitung von Aufgaben mit Diagrammen zu untersuchen. Auf Basis bisheriger Theoriemodelle und empirischer Befunde aus der Diagrammforschung wurden für diese Arbeit drei konkrete Forschungsfragen entwickelt und in jeweils einer psychologischen Experimentalstudie empirisch untersucht. Um dabei einen Einblick in die kognitiven Prozesse zu erhalten, wurden neben ergebnis-orientierten Maßen stets Prozessmaße mittels Eyetracking herangezogen. Bevor die Forschungsfragen im Einzelnen hergeleitet werden, seien sie zunächst kurz benannt:

- (1) Sind Personen dazu in der Lage, für gegebene Aufgaben dasjenige Diagramm zur Bearbeitung zu nutzen, das für die jeweilige Aufgabe computationale Vorteile bietet?
- (2) Beeinflusst eine hohe Datenkomplexität den Bearbeitungsprozess und die Leistung, wenn zusätzliche Daten vollständig irrelevant für die Bearbeitung einer Aufgabe sind?
- (3) Beeinflussen interessante, aber irrelevante Zusatzinhalte (Seductive Details) in Diagrammen den Bearbeitungsprozess und die Leistung von Personen, die Diagrammleseaufgaben bearbeiten?

1.5.1 Forschungsfrage 1: Sind Personen dazu in der Lage, für gegebene Aufgaben dasjenige Diagramm zur Bearbeitung zu nutzen, das für die jeweilige Aufgabe computationale Vorteile bietet?

Die Theoriemodelle von Pinker (1990) und Freedman und Shah (2002) erlauben Vorhersagen darüber, welche Diagrammtypen und –merkmale in Bezug auf bestimmte Aufgaben eine effiziente kognitive Verarbeitung und Lösung ermöglichen. Shah und Hoeffner (2002) fassen in ihrem Forschungsüberblick empirische Befunde zu Diagrammtypen und visuellen Merkmalen zusammen. Sie kommen zu dem Fazit, dass Punktdifferenzen leichter in Säulendiagrammen erkannt und interpretiert werden können als in Liniendiagrammen. Andersherum eignen sich Liniendiagramme besser als Säulendiagramme dazu, Trends zu erkennen (siehe z. B. Zacks & Tversky, 1999). Diese Befunde stehen im Einklang mit theoretischen Modellannahmen. So beschreibt Pinker (1990) beispielsweise, dass Trends deshalb leichter in Liniendiagrammen zu identifizieren sind, weil ein Trend dort durch ein einzelnes visuelles Attribut (Steigung) einer einzelnen perzeptuellen Entität (Linie) repräsentiert wird. Demgegenüber müssen Personen bei Säulendiagrammen mehrere Attribute (Höhen) von multiplen Entitäten (Säulen) im Kopf behalten, wodurch die Bearbeitung langsamer und fehleranfälliger sein kann.

Während diese Annahmen sowohl theoretisch als auch empirisch gut fundiert sind, existiert kaum Wissen darüber, ob Personen dazu in der Lage sind, für die jeweiligen Aufgaben auch die geeigneten Diagramme auszuwählen, um sie zu bearbeiten. Die Fähigkeit, geeignete Repräsentationsformen für die Bearbeitung einer Aufgabe zu verwenden, ist eine wichtige Voraussetzung für eine effektive Auseinandersetzung mit (wissenschaftlichen) Daten.

1.5.2 Forschungsfrage 2: Beeinflusst eine hohe Datenkomplexität den Bearbeitungsprozess und die Leistung, wenn zusätzliche Daten vollständig irrelevant für die Bearbeitung einer Aufgabe sind?

Der Einfluss von Datenkomplexität wurde bereits in zahlreichen Studien als Hintergrundvariable untersucht (z. B. Casali & Gaylin, 1988; Kim & Lombardino, 2015; Kumar & Benbasat, 2004; Meyer et al., 1997; Schutz, 1961a, 1961b; Spence & Lewandowsky, 1991). Die Ergebnisse dieser Untersuchungen zeigen, dass insbesondere die Anzahl der Datenpunkte und Datenreihen die Bearbeitung verschiedener Aufgabentypen negativ beeinflusst. So führt eine höhere Datenkomplexität zu längeren Bearbeitungszeiten und mitunter auch zu einer höheren Fehlerquote. Dies steht im Einklang mit theoretischen

Annahmen der Cognitive Load Theory (Sweller, 1994), die darauf beruht, dass das menschliche Gedächtnis ein Informationsverarbeitungssystem mit begrenzten Ressourcen ist (Baddeley & Hitch, 1974). So konnten Huang et al. (2009) empirisch bestätigen, dass die kognitive Belastung bei der Bearbeitung von Aufgaben mit Diagrammen anstieg, wenn die Datenkomplexität höher war. Das Diagrammlesen kann besonders dann herausfordernd sein, wenn die Datenkomplexität nicht angemessen für die gegebene Aufgabe ist (Huang et al., 2009; Demetriadis & Cadoz, 2005; Sedig & Parsons, 2013).

Bisherige Studien haben jedoch nicht differenziert oder explizit aufgeschlüsselt, ob die zusätzlichen Datenpunkte, die eine höhere Komplexität konstituierten, relevant für die gegebenen Aufgaben waren. Soll beispielsweise ein Trend in einer Datenreihe identifiziert werden und diese wird um weitere Datenpunkte verlängert, so müssen diese selbstverständlich für die Untersuchung und Interpretation des Trends berücksichtigt werden. Das heißt, die Aufgaben-inhärente Komplexität steigt durch Hinzugabe weiterer Punkte zwingend. Demgegenüber gibt es Fälle, in denen zusätzliche Datenpunkte irrelevant für eine Aufgabe sind, beispielsweise, wenn zwei Datenreihen gegeben werden, sich eine Aufgabe jedoch explizit nur auf eine der beiden Datenreihen bezieht. Hierbei stellt sich die Frage, ob auch Aufgaben-irrelevante Datenpunkte die kognitive Belastung erhöhen und Bearbeitungsprozesse erschweren. Dies ist besonders relevant für den Einsatz von Diagrammen im Unterricht, da Lehrkräfte häufig auf vorgefertigtes Material zugreifen, das Informationen beinhalten kann, die für das Lernziel nicht von Bedeutung sind.

1.5.3 Forschungsfrage 3: Beeinflussen interessante, aber irrelevante Zusatzinhalte (Seductive Details) in Diagrammen den Bearbeitungsprozess und die Leistung von Personen, die Diagrammleseaufgaben bearbeiten?

Computer-basiertes Lernen und Testen machen es besonders leicht, Lehr- und Testmaterialien mit zusätzlichen Inhalten anzureichern. Während interessantes, aber irrelevantes Zusatzmaterial das Interesse von Lernenden steigern kann (vgl. Harp & Mayer, 1997), quittieren empirische Studien überwiegend schädliche Effekte auf Behaltens- und Transferleistungen nach dem Lernen (z. B. Harp & Mayer, 1997; Lehman et al., 2007; Rowland-Bryant et al., 2009).

Der Einfluss von Seductive Details wurde bislang hauptsächlich in Lernkontexten und in Bezug auf Textverstehen untersucht (z. B. Garner, et al., 1989; Lehman et al., 2007), während die Diagramme vorwiegend hinsichtlich ihrer inhärenten Merkmale untersucht wurden (z. B. Legenden, Datenkomplexität, Größe und Position von Diagrammelementen;

Cleveland & McGill, 1987; Kim & Lombardino, 2015; Kumar & Benbasat, 2004; Shah & Hoeffner, 2002). Aber auch Diagramme können um zusätzliche Inhalte ergänzt werden, beispielsweise mit Abbildungen, die Datenpunkte veranschaulichen oder mit Texten, die zusätzliche Informationen darbieten (siehe auch Abschnitt 1.2.3; Abb. 8). Diagramme besitzen allerdings eine hochgradig konventionalisierte, räumliche Struktur, die eine mehr oder weniger feste Zahl an visuellen Elementen beinhaltet, beispielsweise Achsen, Beschriftungen und Datenpunkte. Im Gegensatz zu diesen Elementen handelt es sich bei Seductive Details um externe Entitäten, die den Leseprozess unterbrechen oder davon ablenken können (siehe auch Harp & Mayer, 1998; Rey, 2012).

Es stellt sich die Frage, ob und in welchem Ausmaß Seductive Details den Bearbeitungsprozess und die Leistung beim Lesen und Verstehen von Diagrammen beeinflussen. Erkenntnisse aus einer solchen Untersuchung können Aufschluss darüber geben, ob bekannte Befunde zu dem Seductive-Details-Effekt auch auf die Bearbeitung von Diagrammen übertragbar sind und welche Mechanismen ihm in diesem Kontext zugrunde liegen. In der Folge lässt sich aus Studien zur Bedeutung von Seductive Details in Diagrammleseaufgaben ableiten, ob bei der Gestaltung von Lehr- und Lernmaterialien auf interessante, aber irrelevante Zusatzinhalte in Diagrammen verzichtet werden sollte.

1.6 Literatur

- Baddeley, A. D. (1994). The magical number seven: Still magic after all these years? *Psychological Review*, *101*, 353–356. doi:10.1037/0033-295X.101.2.353
- Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, *8*, 47–89. doi:10.1016/s0079-7421(08)60452-1
- Baker, R. S., Corbett, A. T., & Koedinger, K. R. (2001). Toward a model of learning data representations. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society* (pp. 45-50). Mahwah, NJ: Erlbaum.
- Bell, A., & Janvier, C. (1981). The interpretation of graphs representing situations. *For the Learning of Mathematics*, *2*(1), 34–42.
- Bertin, J. (1983). *Semiology of graphics: Diagrams, networks, maps*. Madison, WI: University of Wisconsin Press.
- Brockmann, R. J. (1991). The unbearable distraction of color. *IEEE transactions on professional communication*, *34*(3), 153–159. doi:10.1109/47.84109
- Bundesministerium für Bildung und Forschung (2016). *Berufsbildungsbericht 2016*. Retrieved from https://www.bmbf.de/pub/Berufsbildungsbericht_2016.pdf
- Burke, M. C. (2007). A mathematician's proposal. *Carnegie Perspectives*. Menlo Park, CA: Carnegie Foundation for the Advancement of Teaching. Retrieved from <https://files.eric.ed.gov/fulltext/ED498954.pdf>
- Büchter, A., & Henn, H. W. (2006). *Elementare Stochastik: Eine Einführung in die Mathematik der Daten und des Zufalls*. Berlin, Germany: Springer.
- Bybee, R. W., & Fuchs, B. (2006). Preparing the 21st century workforce: A new reform in science and technology education. *Journal of Research in Science Teaching*, *43*, 349–352. doi:10.1002/tea.20147
- Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction*, *20*(2), 155–166. doi:10.1016/j.learninstruc.2009.02.014
- Carey, J. M., & White, E. M. (1991). The effects of graphical versus numerical response on the accuracy of graph-based forecasts. *Journal of Management*, *17*(1), 77–96. doi:10.1177/014920639101700106

Kapitel 1 – Einleitung und theoretischer Rahmen

- Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, 4(2), 75–100. doi:10.1037/1076-898X.4.2.75
- Casali, J. G., & Gaylin, K. B. (1988). Selected graph design variables in four interpretation tasks: A microcomputer-based pilot study. *Behaviour & Information Technology*, 7(1), 31–49. doi:10.1080/01449298808901861
- Cleveland, W. S., Diaconis, P., & McGill, R. (1982). Variables on scatterplots look more highly correlated when the scales are increased. *Science*, 216(4550), 1138–1141. doi:10.1126/science.216.4550.1138
- Cleveland, W. S., & McGill, R. (1985). Graphical perception and graphical methods for analyzing scientific data. *Science*, 229(4716), 828–833. doi:10.1126/science.229.4716.828
- Cleveland, W. S., & McGill, R. (1987). Graphical perception: The visual decoding of quantitative information on graphical displays of data. *Journal of the Royal Statistical Society. Series A (General)*, 150(3), 192–229. doi:10.2307/2981473
- Cucuo, A. A., & Curcio, F. (2001). *The role of representation in school mathematics: National Council of Teachers of Mathematics Yearbook*. Reston, VA: National Council of Teachers of Mathematics.
- Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education*, 18, 382–393. doi:10.2307/749086
- De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science*, 38, 105–134. doi:10.1007/s11251-009-9110-0
- Demetriadis, S., & Cadoz, C. (2005). A conceptual framework for the integrative design of adaptable representations for learning. In A. Méndez-Vilas, B. González-Pereira, J. Mesa González, & J. A. Mesa González (Eds.), *Recent research developments in learning technologies* (pp. 1250–1255). Badajoz, Spain: Formatex.
- Duchowski, A. (2007). *Eye tracking methodology: Theory and practice*. London, United Kingdom: Springer.

- ERGO Direkt (2012). *Infografik: Lost in Net - so nutzen wir das Internet (Teil 1)*. Retrieved from <https://ergodirekt.de/de/presse/mediathek/infografiken/umfragen/lin-internetnutzung.html>
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87(3), 215–251. doi:10.1037/0033-295X.87.3.215
- Freedman, E. G., & Shah, P. (2001, November). *Individual differences in domain knowledge, graph reading skills, and explanatory skills during graph comprehension*. Paper presented at the 42nd Annual Meeting of the Psychonomic Society, Orlando, FL.
- Freedman, E. G., & Shah, P. (2002). Toward a model of knowledge-based graph comprehension. In M. Hegarty, B. Meyer, & N. H. Narayanan (Eds.), *Diagrammatic representation and inference* (pp. 8–31). Berlin, Germany: Springer.
- Garner, R., Brown, R., Sanders, S., & Menke, D. J. (1992). “Seductive details” and learning from text. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The Role of Interest in Learning and Development* (pp. 239–254). Hillsdale, NJ: Erlbaum.
- Garner, R., Gillingham, M. G., & White, C. S. (1989). Effects of “seductive details” on macroprocessing and microprocessing in adults and children. *Cognition and Instruction*, 6(1), 41–57. doi:10.1207/s1532690xci0601_2
- Gattis, M., & Holyoak, K. J. (1996). Mapping conceptual to spatial relations in visual reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(1), 231–239.
- Glazer, N. (2011). Challenges with graph interpretation: A review of the literature. *Studies in Science Education*, 47, 183–210. doi:10.1080/03057267.2011.605307
- Greaney, J., & MacRae, A. W. (1997). Visual search and the detection of abnormal readings in graphical displays. *Acta Psychologica*, 95(2), 165–179. doi:10.1016/S0001-6918(96)00002-9
- Goldberg, J. H., & Helfman, J. I. (2010). Comparing information graphics: A critical look at eye tracking. In *Proceedings of the 3rd BELIV'10 Workshop: BEyond time and errors: novel evaluation methods for Information Visualization* (pp. 71-78). New York, NY: ACM. doi:10.1145/2110192.2110203

Kapitel 1 – Einleitung und theoretischer Rahmen

- Goldberg, J. H., & Helfman, J. I. (2011). Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. *Information Visualization, 10*, 182–195. doi:10.1177/1473871611406623
- Guthrie, J. T., Weber, S., & Kimmerly, N. (1993). Searching documents: Cognitive processes and deficits in understanding graphs, tables, and illustrations. *Contemporary Educational Psychology, 18*(2), 186–221. doi:10.1006/ceps.1993.1017
- Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: On the distinction between emotional interest and cognitive interest. *Journal of Educational Psychology, 89*(1), 92–102. doi:10.1037/0022-0663.89.1.92
- Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology, 90*(3), 414–434. doi:10.1037/0022-0663.90.3.414
- Hollands, J. G., & Dyre, B. P. (2000). Bias in proportion judgments: The cyclical power model. *Psychological Review, 107*(3), 500–524. doi:10.1037/0033-295X.107.3.500
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. United Kingdom: Oxford University Press.
- Huang, W., Eades, P., & Hong, S. H. (2009). Measuring effectiveness of graph visualizations: A cognitive load perspective. *Information Visualization, 8*(3), 139–152.
- Hyönä, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction, 20*, 172–176. doi:10.1016/j.learninstruc.2009.02.013
- Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction, 20*(2), 146–154.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329–354. doi:10.1037/0033-295X.87.4.329
- Kieler Verkehrsgesellschaft. (2017). *Liniennetz Kerngebiet Kiel*. Retrieved from http://www.kvg-kiel.de/pdf/2017-09-03_KVG_Liniennetzplan_Kiel.pdf
- Kim, S., & Lombardino, L. J. (2015). Comparing graphs and text: Effects of complexity and task. *Journal of Eye Movement Research, 8*(3), 1–17. doi:10.16910/jemr.8.3.2

Kapitel 1 – Einleitung und theoretischer Rahmen

- Kim, S., Lombardino, L. J., Cowles, W., & Altmann, L. J. (2014). Investigating graph comprehension in students with dyslexia: An eye tracking study. *Research in developmental disabilities, 35*(7), 1609–1622. doi:10.1016/j.ridd.2014.03.043
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163–182.
doi:10.1037/0033-295X.95.2.163
- Koerber, S. (2011). Der Umgang mit visuell-grafischen Repräsentationen im Grundschulalter. *Unterrichtswissenschaft, 39*, 49–62. doi:10.3262/UW1101049
- Konrad, K. (2010). Lautes Denken. In G. Mey & K. Mruck (Eds.), *Handbuch Qualitative Forschung in der Psychologie* (pp. 476–490). Wiesbaden, Germany: VS Verlag für Sozialwissenschaften. doi:10.1007/978-3-531-92052-8_34
- Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology, 3*, 185–225. doi:10.1002/acp.2350030302
- Kosslyn, S. M. (1994). *Elements of Graph Design*. New York, NY: Freeman
- Kumar, N., & Benbasat, I. (2004). The effect of relationship encoding, task type and complexity on information representation: An empirical investigation of 2D and 3D graphs. *MIS Quarterly, 28*, 255–281.
- Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65–100. doi:10.1111/j.1551-6708.1987.tb00863.x
- Lauer, T. W., & Post, G. V. (1989). Density in scatterplots and the estimation of correlation. *Behaviour & Information Technology, 8*(3), 235–244. doi:10.1080/01449298908914554
- Legge, G. E., Gu, Y., & Luebker, A. (1989). Efficiency of graphical perception. *Perception & Psychophysics, 46*(4), 365–374. doi:10.3758/BF03204990
- Lehman, S., Schraw, G., McCrudden, M. T., & Hartley, K. (2007). Processing and recall of seductive details in scientific text. *Contemporary Educational Psychology, 32*(4), 569–587. doi:10.1016/j.cedpsych.2006.07.002
- Lewandowsky, S., & Spence, I. (1989). Discriminating strata in scatterplots. *Journal of the American Statistical Association, 84*(407), 682–688.

Kapitel 1 – Einleitung und theoretischer Rahmen

- Lindner, M. A., Eitel, A., Strobel, B., & Köller, O. (2017). Identifying processes underlying the multimedia effect in testing: An eye-movement analysis. *Learning and Instruction, 47*, 91–102. doi:10.1016/j.learninstruc.2016.10.007
- Lindner, M. A., Eitel, A., Thoma, G.-B., Dalehefte, I. M., Ihme, J. M., & Köller, O. (2014). Tracking the decision making process in multiple-choice assessment: Evidence from eye movements. *Applied Cognitive Psychology, 28*, 738–752. doi:10.1002/acp.3060
- Lohse, G. L., Biolsi, K., Walker, N., & Rueter, H. H. (1994). A classification of visual representations. *Communications of the ACM, 37*, 36–49. doi:10.1145/198366.198376
- Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society, 200*, 269–294. doi:10.1098/rspb.1978.0020
- Mausfeld, R. (2005). Über die Bedingungen der Möglichkeit von Lernen. In M. L. Käsermann & A. Altorfer (Eds.), *Über Lernen. Ein Gedankenaustausch* (pp. 218-236). Bern, Switzerland: EditionSolo.
- Mausfeld, R. (2010). The perception of material qualities and the internal semantics of the perceptual system. In A. Liliana, T. Gert, & V. Dhanraj (Eds.), *Perception beyond Inference. The Information Content of Visual Processes* (pp. 159–200). Cambridge, MA: MIT Press.
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511811678
- Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction, 20*(2), 167–171. doi:10.1016/j.learninstruc.2009.02.012
- Merwin, D. H., Vincow, M. A., & Wickens, C. D. (1994, October). Visual analysis of scientific data: Comparison of 3D-topographic, color and gray scale displays in a feature detection task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 38*(4), 240–244. doi:10.1177/154193129403800411
- Meyer, J. (2000). Performance with tables and graphs: Effects of training and a visual search model. *Ergonomics, 43*(11), 1840–1865. doi:10.1080/00140130050174509

Kapitel 1 – Einleitung und theoretischer Rahmen

- Meyer, J., Shinar, D., & Leiser, D. (1997). Multiple factors that determine performance with tables and graphs. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 39(2), 268–286. doi:10.1518/001872097778543921
- Millar, R. (2006). Twenty first century science: Insights from the design and implementation of a scientific literacy approach in school science. *International Journal of Science Education*, 28, 1499–1521. doi:10.1080/09500690600718344
- Moore, C. M., & Egeth, H. (1997). Perception without attention: Evidence of grouping under conditions of inattention. *Journal of Experimental Psychology: Human Perception and Performance*, 23(2), 339–352. doi:10.1037/0096-1523.23.2.339
- National Council of Teachers and Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers and Mathematics.
- OECD (2016). *PISA 2015 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris, France: OECD.
- Oestermeier, U., & Hesse, F. W. (2000). Verbal and visual causal arguments. *Cognition*, 75(1), 65–104. doi:10.1016/S0010-0277(00)00060-3
- Palmer, S. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 259–303). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Peebles, D., & Cheng, P. C. H. (2003). Modeling the effect of task and graphical representation on response latency in a graph reading task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 45, 28–46. doi:10.1518/hfes.45.1.28.27225
- Pereira-Mendoza, L., Goh, S. L., & Bay, W. (2004, November). *Interpreting graphs from newspapers: Evidence of going beyond the data*. Paper presented at the ERAS Conference, Singapore. Retrieved from https://repository.nie.edu.sg/bitstream/10497/15549/1/ERAS-2004-277_a.pdf
- Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Attention, Perception, & Psychophysics*, 16(2), 283–290. doi:10.3758/BF03203943

- Pinker, S. (1990). A theory of graph comprehension. In R. O. Freedle (Ed.), *Artificial intelligence and the future of testing* (pp. 73–126). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Purchase, H. C. (2014). Twelve years of diagrams research. *Journal of Visual Languages & Computing*, 25, 57–75. doi:10.1016/j.jvlc.2013.11.004
- Rey, G. D. (2012). A review of research and a meta-analysis of the seductive detail effect. *Educational Research Review*, 7(3), 216–237. doi:10.1016/j.edurev.2012.05.003
- Roberts, D. A. (2007). Scientific literacy/science literacy. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 729–780). Mahwah, NJ: Lawrence Erlbaum Associates.
- Rowland-Bryant, E., Skinner, C. H., Skinner, A. L., Saudargas, R., Robinson, D. H., & Kirk, E. R. (2009). Investigating the interaction of graphic organizers and seductive details: Can a graphic organizer mitigate the seductive-details effect? *Research in the Schools*, 16(2), 29–40.
- Russell, C., & Driver, J. (2005). New indirect measures of “inattentive” visual grouping in a change-detection task. *Perception & Psychophysics*, 67(4), 606–623. doi:10.3758/BF03193518
- Sandmann A. (2014). Lautes Denken – die Analyse von Denk-, Lern- und Problemlöseprozessen. In D. Krüger, I. Parchmann, & H. Schecker (Eds.) *Methoden in der naturwissenschaftsdidaktischen Forschung* (pp. 179–188). Berlin, Germany: Springer Spektrum. doi:10.1007/978-3-642-37827-0_15
- Sá, W. C., West, R. F., & Stanovich, K. E. (1999). The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology*, 91(3), 497–510. doi:10.1037/0022-0663.91.3.497
- Schnotz, W. (1994). Wissenserwerb mit logischen Bildern. In B. Weidenmann (Ed.), *Wissenserwerb mit Bildern* (pp. 95–148). Bern, Switzerland: Huber.
- Schnotz, W. (2001). Wissenserwerb mit Multimedia. *Unterrichtswissenschaft*, 29(4), 292–318.

- Schnotz, W. (2002). Commentary: Towards an integrated view of learning from text and visual displays. *Educational Psychology Review*, *14*, 101–120.
doi:10.1023/A:1013136727916
- Schiano, D. J., & Tversky, B. (1992). Structure and strategy in encoding simplified graphs. *Memory & Cognition*, *20*(1), 12–20. doi:10.3758/BF03208249
- Schiepe-Tiska, A., Rönnebeck, S., Schöps, K., Neumann, K., Schmidtner, S., Parchmann, I., & Prenzel, M. (2016). Naturwissenschaftliche Kompetenz in PISA 2015 – Ergebnisse des internationalen Vergleichs mit einem modifizierten Testansatz. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme, & O. Köller (Eds.), *Eine Studie zwischen Kontinuität und Innovation*. Münster, Germany: Waxmann.
- Schutz, H. G. (1961a). An evaluation of formats for graphic trend displays—experiment II. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *3*(2), 99–107.
- Schutz, H. G. (1961b). An evaluation of methods for presentation of graphic multiple trends—experiment III. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *3*(2), 108–119.
- Sedig, K., & Parsons, P. (2013). Interaction design for complex cognitive activities with visual representations: A pattern-based approach. *AIS Transactions on Human-Computer Interaction*, *5*(2), 84–133.
- Shah, P. (2002). Graph comprehension: The role of format, content and individual differences. In *Diagrammatic Representation and Reasoning* (pp. 173–185). Springer London. doi:10.1007/978-1-4471-0109-3_10
- Shah, P., Freedman, E. G., & Vekiri, I. (2005). The comprehension of quantitative information in graphical displays. In P. Shah & A. Miyake (Eds.), *The Cambridge Handbook of Visuospatial Thinking* (pp. 426–476). doi:10.1017/CBO9780511610448.012
- Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review*, *14*, 47–69. doi:10.1023/A:1013180410169
- 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), 690–702. doi:10.1037/0022-0663.91.4.690

Kapitel 1 – Einleitung und theoretischer Rahmen

- Shomstein, S., Kimchi, R., Hammer, M., & Behrmann, M. (2010). Perceptual grouping operates independently of attentional selection: Evidence from hemispatial neglect. *Attention, Perception, & Psychophysics*, 72(3), 607–618. doi:10.3758/APP.72.3.607
- Simkin, D., & Hastie, R. (1987). An information-processing analysis of graph perception. *Journal of the American Statistical Association*, 82, 454–465. doi:10.1080/01621459.1987.10478448
- Spence, I., & Lewandowsky, S. (1991). Displaying proportions and percentages. *Applied Cognitive Psychology*, 5, 61–77. doi:10.1002/acp.2350050106
- Strobel, B., Lindner, M. A., Saß, S., & Köller, O. (in press). Task-irrelevant data impair processing of graph reading tasks: An eye tracking study. *Learning and Instruction*. doi:10.1016/j.learninstruc.2017.10.003
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12, 257–285. doi:10.1207/s15516709cog1202_4
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, 4(4), 295–312.
- Tuttle, B., & Kershaw, R. (1998). Information presentation and judgment strategy from a cognitive fit perspective. *Journal of Information Systems*, 12, 1–17.
- Tversky, B., & Schiano, D. J. (1989). Perceptual and conceptual factors in distortions in memory for graphs and maps. *Journal of Experimental Psychology: General*, 118(4), 387–398. doi:10.1037/0096-3445.118.4.387
- van Gog, T., Kester, L., Nieuvelstein, F., Giesbers, B., & Paas, F. (2009). Uncovering cognitive processes: Different techniques that can contribute to cognitive load research and instruction. *Computers in Human Behavior*, 25(2), 325–331. doi:10.1016/j.chb.2008.12.021
- van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction*, 20(2), 95–99. doi:10.1016/j.learninstruc.2009.02.009
- Wickens, C. D., Merwin, D. H., & Lin, E. L. (1994). Implications of graphics enhancements for the visualization of scientific data: Dimensional integrality, stereopsis, motion, and mesh. *Human Factors*, 36(1), 44–61. doi:10.1177/00187208940360010

Kapitel 1 – Einleitung und theoretischer Rahmen

Wainer, H. (1992). Understanding graphs and tables. *Educational Researcher*, 21, 14–23.
doi:10.3102/0013189X021001014

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. II. *Psychologische
Forschung*, 4(1), 301–350.

Winn, W. D. (1987). Charts, graphics and diagrams in educational materials. In D. Willows &
H. Houghton (Eds.), *The Psychology of Illustration* (pp. 152–198). New York; NY:
Springer.

Winn, W. D. (1990). A theoretical framework for research on learning from graphics.
International Journal of Educational Research, 14, 553–564.
doi:10.1016/0883-0355(90)90025-4

Wright, R. D., & Ward, L. M. (2008). *Orienting of attention*. New York, NY: Oxford
University Press.

Zacks, J., & Tversky, B. (1999). Bars and lines: A study of graphic communication. *Memory
& Cognition*, 27, 1073–1079. doi:10.3758/BF03201236

Kapitel 2

Studie 1

Do Graph Readers Prefer the Graph Type Most Suited to a Given Task? Insights from Eye Tracking

Strobel, B., Saß, S., Lindner, M.A., & Köller, O. (2016). Do Graph Readers Prefer the Graph Type Most Suited to a Given Task? Insights from Eye Tracking. *Journal of Eye Movement Research*, 9(4), 1-15. doi:10.16910/jemr.9.4.4

Abstract

Research on graph comprehension suggests that point differences are easier to read in bar graphs, while trends are easier to read in line graphs. But are graph readers able to detect and use the most suited graph type for a given task? In this study, we applied a dual representation paradigm and eye tracking methodology to determine graph readers' preferential processing of bar and line graphs while solving both point difference and trend tasks. Data were analyzed using linear mixed-effects models. Results show that participants shifted their graph preference depending on the task type and refined their preference over the course of the graph task. Implications for future research are discussed.

Keywords: Graph comprehension; graph preference; diagrams; eye tracking

Do Graph Readers Prefer the Graph Type Most Suited to a Given Task? Insights from Eye Tracking

Visualizations such as graphs and diagrams are ubiquitous in everyday human experience, and can be found in newspapers and television, as well as in science, engineering, and education (e.g., Blackwell, 2001; Glazer, 2011; Mayer, 2009; Pereira-Mendoza, Goh, & Bay, 2004; Purchase, 2014; Schnotz, 1994). They are especially important in the context of problem solving (Baker, Corbett & Koedinger, 2001), for teaching and learning mathematics (Cucuo & Curcio, 2001) and for understanding scientific data (Shah & Hoeffner, 2002).

To date, a large variety of different graph types have been developed (cf. Bertin, 1983; Kosslyn, 1989; Lohse, Biolsi, Walker, & Rueler, 1994). The advantages of graphs are computational because they support efficient computational processes (Larkin & Simon, 1987). De-spite some similarities between the formats, computational differences between different graph types have been identified, even if two graphs are informationally equivalent (e.g., Kosslyn, 1989; Pinker, 1990). In general, two representations are called informationally equivalent when they display the same relations between the same objects, “because they are indistinguishable in terms of the information they represent” (Palmer, 1978, p. 270). Furthermore, two representations are called computation-ally equivalent when they are both informationally equivalent and “any inference that can be drawn easily and quickly from the information given explicitly in one can also be drawn easily and quickly from the information given explicitly in the other, and vice versa” (Larkin & Simon, 1987, p. 67).

A certain graph may facilitate information processing in some tasks, but not in others. For example, Simkin and Hastie (1987) used three informationally equivalent graph types and found that speed and accuracy in a graph task depended on the combination of graph type and task requirements. More recently, Peebles and Cheng (2003) discovered that the computational advantages of a graph may even compensate for a reader’s unfamiliarity with a certain graph type as participants showed significant learning effects over time. These findings show that computational properties may affect one’s efficiency in completing graph tasks. Yet, there is little knowledge about graph readers’ ability to detect and use computationally advantageous representations for a given task when given the choice. Accordingly, the main goal of the current study was to investigate graph readers’ ability to adapt their processing strategy to the demands of the task (i.e., preferring the graph type that is most suitable for solving a particular task). To accomplish this, we systematically varied the

Kapitel 2 - Do Graph Readers Prefer the Graph Type Most Suited to a Given Task?

task type while presenting two informationally equivalent graphs that are known to be computationally different with regard to the tasks (see e.g., Pinker, 1990; Shah & Hoeffner, 2002). However, previous graph comprehension research using eye-tracking methodology has mainly focused on the processing of one graph at a time (e.g., Shah & Carpenter, 1995; Goldberg & Helfman, 2011; Kim & Lombardino, 2015; Peebles & Cheng, 2003). For the current study, we used a dual representation paradigm and eye-movement parameters to determine graph readers' preferential usage of two graph types while processing a graph task.

Visual attention, measured by eye tracking, has repeatedly been used to establish a measure of preference, for example, in market research (e.g., Meissner & Decker, 2010), usability research (e.g., Djamasbi, Siegel, Skorinko, & Tullis, 2011) and in educational assessment with multiple-choice questions (Lindner et al., 2014). Building on this approach, we applied eye-movements as a measure of preference to the domain of graph comprehension.

In the following, we briefly review three important aspects of graph comprehension. First, we (1) introduce typical graph reading tasks, providing helpful insights into the nature of the graph tasks in the present study. Then we (2) outline the computational properties of different graph types. Finally, we (3) present findings on graph reading and learning effects.

2.1 Graph reading tasks

When it comes to graph reading tasks, a great number of taxonomies have been developed in the past decades, comprised of highly similar concepts (i.e., Bertin, 1983; Curcio, 1987; Schnotz, 1994; Tan & Benbasat, 1990; Wainer, 1992). Reviewing these classifications reveals a remarkable consensus across decades and authors that provides a high level of confidence that basic graph reading tasks can be identified with a high reliability. Tasks involving the *recognition of relations between multiple elements* are in the focus of the current study as they have been repeatedly classified as typical graph tasks. This includes identifying trends (e.g. a development in sales numbers over years) as well as identifying point differences (e.g. a difference in sales numbers between distinct years).

2.2 Computational properties of graphs

In graphs, information is not represented by a resemblance to physical objects. Instead, spatial relations between visual objects are employed as an analogy to non-spatial relations (e.g., Koerber, 2011; Winn, 1990). Often, inferences from non-spatial representations are relatively more demanding, because some information has to be computed at great cognitive

costs (e.g., comparing numbers in a continuous text). Thus, graphs offer computational advantages in comparison to other forms of representations (Larkin & Simon, 1987).

Graphs are not only computationally different from other representations, such as text and tables, but also from each other (Schnotz, 2002; Wainer, 1992). A number of studies provide empirical support for the idea that differences in the computational characteristics of different graph types can explain task performance in typical graph tasks (e.g., Peebles & Cheng, 2003; Simkin & Hastie, 1987; Shah, Mayer, & Hegarty, 1999). In their review of graph comprehension literature, Shah and Hoeffner (2002) summarize that bar graphs emphasize point differences, while line graphs are more helpful for identifying trends. Taking a closer look at the processes that guide graph comprehension, Pinker (1990) outlines the computational differences between bar graphs and line graphs. He suggests that line graphs facilitate the observation of trends (*trend task*) because, in this format, a trend can correspond to a single visual attribute (e.g., *slope*) of a perceptual entity (*line*) in the display. In bar graphs, on the other hand, trends do not correspond to the attributes of single perceptual entities. Instead, the attributes (e.g., *height*) of multiple entities (*bars*) need to be kept in mind, which makes observing trends more effortful and error prone. An observation of point differences (*difference task*), on the other hand, can be made more easily using bar graphs, because bars are distinct entities and facilitate point estimates (Zacks & Tversky, 1999). For example, one might want to compare a company's sales numbers for a set of particular years. Rather than looking at the numbers and computing the differences, viewers can use spatial analogies to make inferences about a non-spatial concept. In bar graphs, the horizontal position of the bars corresponds to a value on the depicted scale (e.g., years), and their height corresponds to a value on another scale (e.g., sales numbers). Using this graph type, readers can easily identify particular years because they are represented by distinct visual entities (*bars*). In a line graph, however, extracting distinct points is much more difficult because perceptual grouping (Wertheimer, 1938) causes each line to be encoded as a single entity rather than as a set of distinct points (Pinker, 1990). In summary, bar and line graphs are not computationally equivalent; rather, they provide very different computational advantages regarding trend and difference tasks. Thus, in combination, they seem useful in eliciting a shift in graph readers' preferential processing.

2.3 Graph reading and learning effects

Basic graph reading skills can be observed as early as in preschool and middle school (e.g., Curcio, 1987; Koerber, 2011) and these skills develop over time (Lowrie & Diezmann,

Kapitel 2 - Do Graph Readers Prefer the Graph Type Most Suited to a Given Task?

2007). Nonetheless, unexperienced readers often make mistakes when interpreting graphs (Bell & Janvier, 1981; Shah & Carpenter, 1995; Shah & Hoeffner, 2002). Even among more experienced graph readers, task performance seems to vary (e.g. Ali & Peebles, 2013). As shown by Baker, Corbett and Koedinger (2001), graph readers' misconceptions may arise from a lack of familiarity with some graph types. In contrast, a study with more experienced graph readers by Peebles and Cheng (2003) demonstrated a tradeoff between familiarity and the computational advantages of a given graph. The authors compared informationally equivalent graph types (a function line graph and a parametric line graph) and found that the effectiveness of a particular graph in retrieving the required information depended on the task, exposing computational differences between the formats. Even though participants were more familiar with one of the two graph types, they showed a significant learning effect after repeated exposure to the less familiar graph and were able to use it more effectively. These findings emphasize the importance of the computational properties inherent to different graph types as well as the potential of computational advantages to outweigh the cost of familiarization.

Looking at research on educational assessment, there is additional evidence that students are able to improve their performance over the course of a test due to learning effects regarding the handling of the given tasks (e.g., Hartig & Buchholz, 2012; Ren, Wang, Altmeyer, & Schweizer, 2014). Based on this, we assume that graph readers can improve their ability to adapt their processing strategy to the demands of a task, showing an increasing preference for the computationally advantageous graph over time.

2.4 Research questions

According to a number of graph task classifications (Bertin, 1983; Curcio, 1987; Schnotz, 1994; Tan & Benbasat, 1990; Wainer, 1992), identifying trends and comparing point differences are regular tasks of graph reading that are often necessary for understanding and interpreting data sets. Furthermore, there is conclusive empirical evidence that task performance may be influenced by the graph type that is used to complete the task. For bar and line graphs, the computational advantages are very different: trends can be understood easier when reading line graphs instead of bar graphs, while the opposite is the case for understanding point differences. To date, there is little knowledge about graph readers' ability to detect and use the graph most suited to a given task. Therefore, the aim of our study was to compare the processing of graphs in difference tasks and trend tasks, while presenting both line graphs and bar graphs at the same time. Additionally, we investigated graph readers'

Kapitel 2 - Do Graph Readers Prefer the Graph Type Most Suited to a Given Task?

preference development as it unfolds across the experimental trials. Even though there is some evidence that graph readers may get more efficient in using the computational advantages of graphs over time (Peebles and Cheng, 2003), this second research question is comparatively novel and more explorative in nature. In summary, we expected the following:

- (1) *Preferential graph processing hypothesis*: Graph readers prefer the computationally advantageous graph type as a function of the task: (a) Graph readers prefer line graphs to bar graphs for tasks that require an observation of trends (*trend tasks*). (b) Graph readers prefer bar graphs to line graphs for tasks that require an observation of point differences (*difference tasks*).
- (2) *Preference development hypothesis*: Over the course of processing multiple difference and the trend tasks, graph readers' preference for the computationally advantageous graph type increases: (a) Graph readers' preference for line graphs increases over the course of processing trend tasks. (b) Graph readers' preference for bar graphs increases over the course of processing difference tasks.

2.5 Method

2.5.1 Sample and study design

The participants in our study were 32 students on the undergraduate level from different faculties (70% female, $M_{\text{age}} = 24.37$ years). All participants had normal or corrected to normal vision. Data from two participants (6 %) were excluded from the analysis due to the poor quality of the eye-tracking data, resulting in a final sample of $N = 30$ participants. The assessment comprised a set of graph tasks (difference and trend tasks) as well as a short paper-pencil questionnaire to assess demographic information as well as participants' graph literacy and overall preference for graph types as a control variable. For the graph tasks, we employed a within-subject design, in which we varied the task type (difference task vs. trend task) using different sets of true-false statements. In the difference task, participants had to evaluate the true-false statement by identifying point differences in the graphs (e.g., "In 2011, more new apartments were built than in 2013"). In the trend task, participants had to identify trends to complete an item (e.g., "Investments in climate protection have increased over the years"). There were 24 trials in each condition of the task type, resulting in a total of 48 trials per participant. Eye-movement data were recorded for each trial and each participant during processing of the graph tasks.

2.5.2 Material and measures

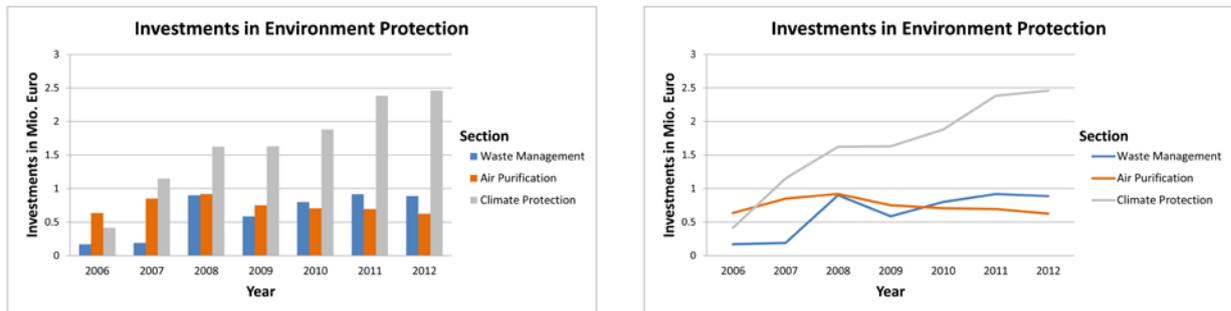
2.5.2.1 Graph tasks

We developed a dual representation paradigm for an indirect preference measure using eye-movement data. Participants were confronted with two representations at the same time, followed by either a difference or trend task that could only be completed with information derived from the given graphs and did not require any previous knowledge. Each trial consisted of a pair of graphs, next to each other in the center of the display; one a bar graph, the other a line graph. Positions (left or right) of the two graph types were balanced randomized. A statement was displayed below the pair of graphs, followed by the choices “true” and “false”, which could be selected by participants as a response with a mouse click. True-false items are known to function differently depending on whether they are true or false (Grosse & Wright, 1985; for a detailed overview see Haladyna, 2004), so we balanced the number of true and false statements throughout the test. Both graphs of each pair were constructed using the same data set, resulting in two informationally equivalent graphs. Data sets were taken from the Federal Statistical Office in Germany, using the office’s website (www.destatis.de). By using real data sets, we reduced the risk of disbelief resulting from an implausibility of the data. Data were refined to achieve comparable complexity throughout all items. Each pair of graphs displayed three variables employing the x-axis, y-axis and legend of the graphs. Data points were trimmed to show between five and eight occurrences on the x- and y-axis and three categories conveyed by differing colors. Identical labels and legends were used in both graph types. A translated example is given in Figure 1.

2.5.2.2 Paper-pencil questionnaire

A short questionnaire was employed to survey participants’ *graph literacy* and overall preference for the two graph types (bar graph vs. line graph) using a 5-point Likert scale. The graph literacy scale was comprised of two items (i.e., “I am good at reading graphs”, “I feel confident in reading graphs.”). Another two items were used to assess the overall preference. Here, one item was the inverse of the other (“Generally, I prefer bar graphs to line graphs” and “Generally I prefer line graphs to bar graphs”). To facilitate interpretation, the *overall graph preference* was calculated as the difference between the preference scores of the two items. Positive values indicate an overall preference for bar graphs, whereas negative values indicate an overall preference for line graphs.

Kapitel 2 - Do Graph Readers Prefer the Graph Type Most Suited to a Given Task?



When the investments in climate protection increased, the investments in air purification increased, too.

True

False

Figure 1. Sample stimulus of the graph tasks (translated from German).

2.5.3 Apparatus

Items were presented on a 19-inch screen with a 1280×1024 pixel resolution, using the software Experiment Center 3.4 from SensoMotoric Instruments (SMI, Teltow, Germany). Each item was presented on a single screen. While working on the test, participants sat in front of the screen at a distance of approximately 70 centimeters. The font size of the text was 2 centimeters (approx. 1.6° visual angle). Both graphs had an identical total size of 17.7 x 9.9 centimeters (approx. 14.4° visual angle in width and 8.1° visual angle in height) and were displayed next to each other with a gap of 0.8 centimeters (approximately 0.7° visual angle) in the center of the screen. Participants' eye-movements were recorded using a video-based remote eye-tracking system (SMI iView XTMRED-m; 120Hz sampling rate) and the corresponding SMI software iView XTM. The system was calibrated using an animated 8-point calibration image and subsequent validation. The calibration accuracy was below 0.49° of the visual angle for all participants on both the x and y coordinates (range: 0.09 to 0.48 ; $M_x = 0.32$, $SD_x = 0.08$; $M_y = 0.29$, $SD_y = 0.11$).

Kapitel 2 - Do Graph Readers Prefer the Graph Type Most Suited to a Given Task?

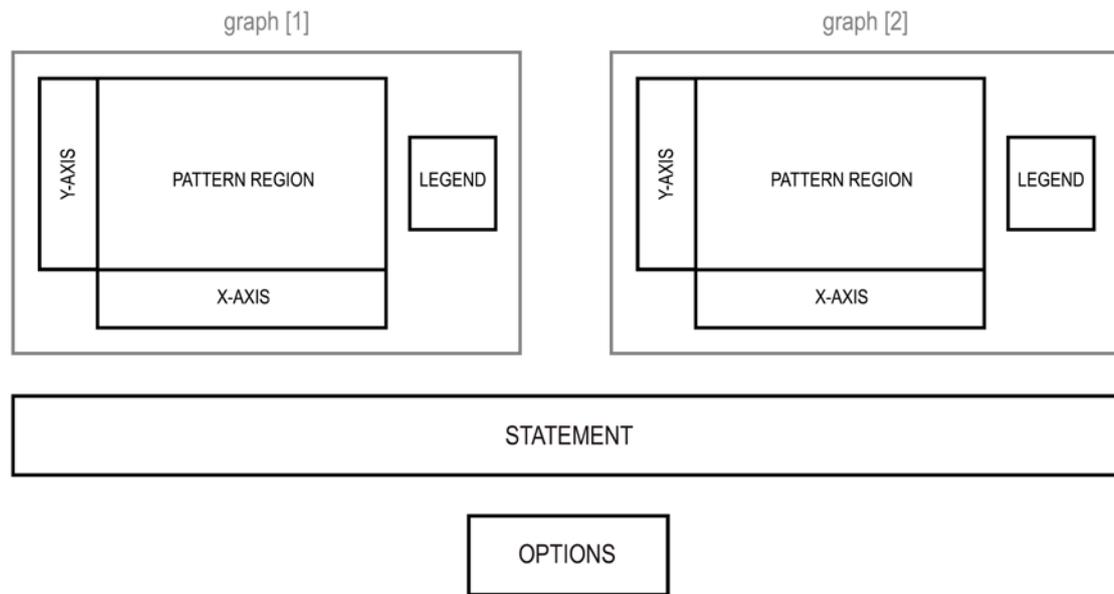


Figure 2. Composition of Areas of Interest (in capital letters) defined for the eye-tracking analysis.

2.5.4 Procedure

Students were tested in single sessions. First, they were familiarized with the procedure and the eye-tracking system; after this, they completed the graph tasks on a computer while their eye-movements were recorded. Participants were informed that both graphs in each item conveyed the same information. However, they received no particular instructions on how to choose between the given representations. Participants could therefore employ a solution strategy based on their individual preference instead of following a strategy that was either given or implied. The paper-pencil questionnaire was administered after participants completed the graph task. The whole cycle of assessment took about 30 minutes to complete.

2.5.5 Analysis

2.5.5.1 Eye-movement data pre-processing

Eye-movement recordings were analyzed using a dispersion-based algorithm implemented in the Begaze™ software, version 3.5, from SMI. A *fixation* was detected when eye movements lasted for at least 80 milliseconds on a position with a maximum dispersion of 100 pixels. To determine whether eye-movement data were recorded correctly, participants' scan paths were visually inspected using Begaze™ software, version 3.5, from SMI.

As gaze data, we used the *total fixation time* on predefined Areas of Interest (AoIs), which is defined as the cumulative duration of all AoI fixations from trial onset to task end,

Kapitel 2 - Do Graph Readers Prefer the Graph Type Most Suited to a Given Task?

indicating the total time devoted to a specific area (Holmqvist et al., 2011). Relying on the eye-mind hypothesis (Just & Carpenter, 1980), we assume that points of fixation represent the focus of attention, so that eye movements reflect the spatiotemporal encoding of visual information. Fixation times provide a valid indirect measure of attention distribution and cognitive processing in educational tasks, such as solving test items in a multiple-choice format (see e.g. Lindner et al., 2014).

As in previous graph studies (e.g., Carpenter & Shah, 1995; Kim & Lombardino, 2015; Peebles & Cheng, 2003), we divided the graph regions into four rectangular AoIs: x-axis, y-axis, legend, and pattern, defined separately for each of the two graphs, in addition to the statement and option (true/false) regions, resulting in a total of ten AoIs (Figure 2). The AoIs covering the pattern regions were about the same size - in every display and for the two graph types within a display (approx. 10 x 7 cm). Because the graph reading processes necessary to complete the tasks occur within the pattern area of the graph (e.g., comparing heights of bars or slopes of lines), we compared fixation times on the pattern regions.

Preferential processing of one graph over the other graph was defined as the difference between total fixation times on the pattern regions of the two graphs. However, not all fixations during a trial reflect preference. In order to decide which graph to use for task completion, participants must first read the statement to identify the task type. Accordingly, only fixations that occurred after the first fixation of the statement region were used to compute the preference measure.

2.5.5.2 Linear mixed-effects models

All data were analyzed using R, version 3.1.0 for Windows (R Development Core Team, 2015). With repeated measures nested within participants as well as in items, the data structure can be described as clustered or hierarchical (Snijders & Bosker, 2012). Since the repeated measures were nested within two higher-level units, the test design is called cross-classified (i.e., each measure belongs uniquely to one participant and one item; Snijders & Bosker, 2012). Regular approaches, such as ANOVA models, yield inflated Type I error rates when the data is clustered (Dorman, 2008). To account for the clustered data structure, we applied linear mixed-effects models (LMMs; for introductions, see Barr, 2008; Snijders & Bosker, 2012; Quené & van den Bergh, 2008). These models can be conceptualized “as a series of interrelated regression models that explain sources of variance at multiple levels of analysis, such as at the experimental stimuli and person levels” (Hoffman & Rovine, 2007, p. 102). More specific, LMMs can model fixed effects and random effects simultaneously.

Kapitel 2 - Do Graph Readers Prefer the Graph Type Most Suited to a Given Task?

While fixed effects aim to identify the typical rates of change in the outcome variable (e.g., following an experimental manipulation), random effects aim to identify unsystematic rates of change (e.g., due to differences between items and persons, respectively).

We used the R package lme4 (Bates, Maechler & Bolker, 2015) to perform a linear mixed-effects analysis of the relationship between the relative graph preference and several explanatory variables in four models: First, we computed an empty model with random intercepts for subjects and stimuli (M0) in order to gain insights into the variance structure of the data. To test our hypotheses, we included fixed effects for task type, trial number, and the task type by trial number interaction (M1). Then, we added the participants' overall graph preference as a control variable for individual preferences (M2). Finally, we added random effects by including by-subject random slopes for the effects of task type and trial number to account for differences in individual trajectories between tasks and over time (Barr, Levy, Scheepers & Tily, 2013; M3). Models were fitted by the Restricted Maximum Likelihood (REML) criterion since it yields better Type I error rates for smaller groups ($N \leq 50$) when testing fixed effects than estimates with the Maximum Likelihood (ML) criterion (Manor and Zucker, 2004; Snijders & Bosker, 2012).

2.6 Results

The ratings of graph literacy in the current sample were relatively high ($M = 3.55$, $SD = 0.83$, range = 2 to 5), with the graph literacy scale showing a good reliability (Cronbach's $\alpha = .82$). Participants achieved a correspondingly high average accuracy in the graph tasks, 89 % in the trend task and 90 % in the difference task. There was no significant difference in participants' performance between the task types ($t_{29} = 0.61$, $p = .55$). Because gaze data may differ for correct and incorrect responses, we removed data of incorrect responses from our analyses.

Average total fixation times and average percentages of processing time on the predefined AoIs across items and participants in the two task conditions are given in Table 1. Data showed a noticeable difference in total processing time between the two graph tasks. We compared the average total processing time across participants for trials in the difference task versus trials in the trend task using a paired t-test in order to confirm that this difference was significant ($t_{23} = 4.53$, $p < .001$, $d = 0.92$). This is also reflected by a significant difference in total fixation time on the statement area ($t_{23} = 5.90$, $p < .001$, $d = 1.20$). To account for this bias in processing time, we computed the total relative fixation time (percentage of processing time) for each trial of each participant as the total fixation time (on AoIs) relative to the total

Kapitel 2 - Do Graph Readers Prefer the Graph Type Most Suited to a Given Task?

processing time of the respective trial. As the preference measure, we computed the relative graph preference as the difference between total relative fixation time on the bar graph pattern region and total relative fixation time on the line graph pattern region in percentage points. Difference values were calculated for each trial from each participant. Positive values indicate preferential processing of the bar graph, whereas negative values indicate a preference for the line graph.

Regarding the linear mixed-effects models, a visual inspection of residual plots did not reveal obvious deviations from homoscedasticity or normality. Table 2 shows the fixed effects and variance components of the random effects in the linear mixed-effects models described above. Analyzing the empty model (M0), the intraclass correlations (ICC) for subjects and stimuli revealed that substantial portions of variance can be attributed to differences between subjects and stimuli, respectively ($ICC_{\text{Subject}} = 0.11$; $ICC_{\text{Stimulus}} = 0.19$). To perform likelihood ratio tests (LRT), each model was refitted using ML instead of REML because the LRT for fixed effects using REML is known to be inappropriate (Pinheiro & Bates, 2000; Snijders & Bosker, 2012). LRTs showed that each addition of explanatory variables (as described above) was a significant contribution to the model (see Table 2)¹. Regression lines of the fixed effects in M3 are given in Figure 3. The following sections focus on the fixed effects in Model M3 in order to address our research questions.

2.6.1 Overall graph preference

Even though participants varied in their overall graph preference, there was no global bias towards one of the two graph types ($M = 0.34$; $SD = 1.70$; Range = -3 to 3). To control for individual preferences for one of the graph types, we included participants' reported graph preference in the linear mixed-effects models. Comparing models M1 and M2, the inclusion of this variable explained a small portion of the subject-related variance in the random part of the model. The change in deviance between the two models was significant ($\chi^2[1] = 5.30$; $p < .05$). In the final LMM (M3), the overall graph preference showed a significant positive relation to the relative graph preference ($p < .05$). A change of one point in the overall graph preference score resulted, on average, in a change of 1.4 percentage points in the same direction in the estimated relative graph preference. This means that participants who reported a preference for one of the graph types also showed a slightly higher preference for that graph type during item processing.

¹ In addition, we included interaction terms in the fixed effects for overall graph preference, but the change in deviance was nonsignificant ($\chi^2[3] = 1.53$, $p = .68$). Thus, this model was omitted.

Kapitel 2 - Do Graph Readers Prefer the Graph Type Most Suited to a Given Task?

Table 1

Average total fixation times (in ms) and average percentages of processing time (in %) on predefined Areas of Interest for difference and trend task conditions across items (24 trials per condition) and participants (n=30)

Area of Interest	Average total fixation time (in ms)				Average percentage of processing time (in %)			
	Difference Task		Trend Task		Difference Task		Trend Task	
<i>Bar graph (total)</i>	6828.8	(6551.0)	2767.8	(4181.6)	22.9	(17.5)	13.8	(16.5)
Pattern	4044.3	(4236.0)	1823.1	(3006.5)	13.4	(11.5)	9.1	(12.0)
Legend	1731.0	(1971.4)	682.0	(1119.9)	6.0	(6.2)	3.6	(5.2)
X-axis	819.4	(1464.4)	165.4	(489.2)	2.8	(4.3)	0.8	(2.0)
Y-axis	234.1	(692.2)	97.4	(364.1)	0.6	(1.7)	0.4	(1.7)
<i>Line graph (total)</i>	4524.5	(4538.5)	5139.5	(4978.4)	17.7	(16.4)	30.3	(16.9)
Pattern	2548.1	(2857.0)	3324.6	(3712.0)	9.8	(10.1)	18.9	(13.2)
Legend	1231.5	(1517.3)	1477.1	(1289.1)	4.9	(5.9)	9.6	(6.8)
X-axis	539.5	(1004.8)	198.9	(445.3)	2.3	(4.0)	1.2	(2.5)
Y-axis	205.4	(554.0)	138.9	(433.8)	0.6	(1.6)	0.7	(2.0)
Statement	9004.7	(5737.1)	4341.8	(3582.5)	33.1	(13.1)	26.9	(15.2)
Options	1104.7	(636.1)	944.7	(571.3)	4.6	(3.0)	6.8	(4.8)
Total processing time	27685.8	(14533.3)	16988.2	(11274.9)				

Note. Standard deviations are given in parenthesis. Data of total processing time had one missing value (< 0.1%).

Table 2

Comparison of fixed effects and random effects in the linear mixed-effects models

Fixed effect	Model 0 (M0)			Model 1 (M1)			Model 2 (M2)			Model 3 (M3)		
	Estimate	SE	t-value	Estimate	SE	t-value	Estimate	SE	t-value	Estimate	SE	t-value
Intercept	-3.02	1.89	-1.60	9.84**	2.81	3.50	9.18**	2.79	3.30	9.23**	2.83	3.26
Trial no.				-0.53**	0.19	-2.83	-0.52**	0.19	-2.83	-0.53*	0.22	-2.39
Task type (trend)				-15.36***	3.57	-4.31	-15.36***	3.57	-4.30	-14.92**	4.06	-3.68
Task type x Trial no				0.16	0.27	0.60	0.16	0.27	0.26	0.14	0.27	0.52
Overall preference							1.79*	0.77	2.36	1.42*	0.64	2.23
Random effect	Variance Component			Variance Component			Variance Component			Variance Component		
Stimulus	83.03			28.78			28.80			31.00		
Subject	47.61			47.54			40.31			46.19		
Trial number										0.40		
Task type (trend)										108.35		
Residual	309.19			309.23			309.21			262.20		
Deviance	11191			11148			11142			11024		
LR-Test				M0-M1: $\chi^2(3) = 43.34***$			M1-M2: $\chi^2(1) = 5.30*$			M2-M3: $\chi^2(5) = 118.58***$		

Note. Degrees of freedom for statistical tests of fixed effects were determined by a conservative interpretation of a rule described in Snijders & Bosker (2012). That is, $df = N - q - 1$, where N is the minimum number of upper level units (here: $N = 30$ participants) and q is the number of explanatory variables included in the model.

* = $p < .05$ ** = $p < .01$ *** = $p < .001$

2.6.2 Preferential graph processing

First, we hypothesized that participants would prefer line graphs to bar graphs in trend tasks (Hypothesis 1a), whereas a preference of bar graphs was expected in difference tasks (Hypothesis 1b). Analyzing the fixed effects in the final LMM (M3), task type had a substantial impact on the relative graph preference: A change in task type from difference task to trend task lowered the estimated relative graph preference by about 15 percentage points. This difference was significant ($p < .001$; Table 2). For participants with no reported preference for either graph type, this change indicates, on average, a preference shift from bar graphs in difference tasks to line graphs in trend tasks. For example, in the first trial of the difference task, participants had, on average, a relative preference of 9 percentage points in favor of bar graphs. In the first trial of the trend task, however, participants showed a relative preference of 6 percentage points in favor of line graphs. In addition to the linear mixed-effects models, we conducted two paired t-tests in order to compare relative fixation times on bar graphs versus line graphs separately for each task condition. Here, total relative fixation times on bar graphs and on line graphs were aggregated across items for each participant. In the trend task, participants devoted significantly less time to the processing of the bar graph compared to the line graph ($t_{29} = -4.97, p < .001, d = 0.91$) and vice versa in the difference task ($t_{29} = 2.75, p < .05, d = 0.50$), with the effect size in the trend task being almost twice the size of that in the difference task. Both the results from the linear mixed-effects model and the additional t-tests are in line with our preferential graph-processing hypotheses (1a, 1b).

2.6.3 Preference development

Second, we expected participants to develop a stronger preference for line graphs in the trend task over time (Hypothesis 2a), whereas we expected participants to develop a stronger preference for bar graphs in the difference task (Hypothesis 2b). The LMM (M3) revealed a significant fixed effect of trial number ($p < .05$; Table 2), whereas the fixed effect of the trial number by task type interaction was not significant ($p = .52$; Table 2). From one trial to the next, on average, the estimated relative preference was reduced by about 0.5 percentage points in the difference task and by about 0.4 percentage points in the trend task, indicating that participants developed a stronger preference for line graphs both in the trend and in the difference task. However, for the difference task, we had predicted a change in the opposite direction. These findings are in line with hypothesis 2a, but could not confirm hypothesis 2b. Possible explanations for this finding are discussed below.

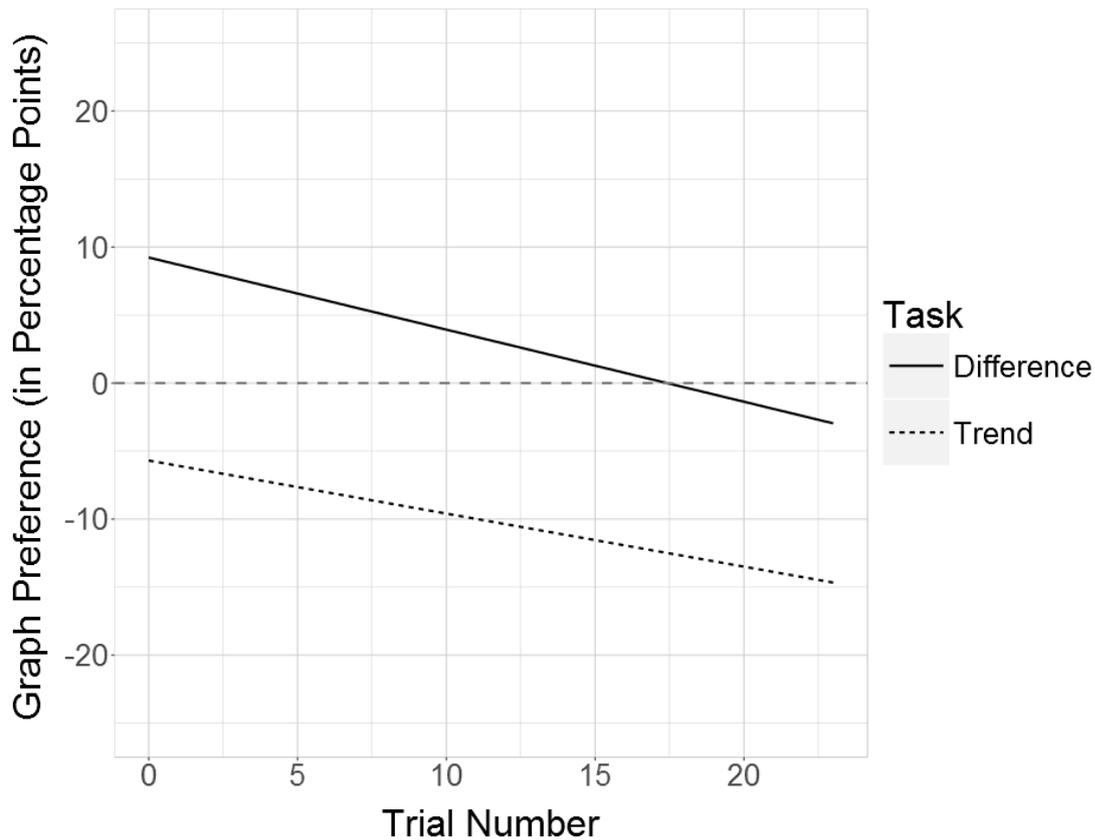


Figure 3. Regression lines of the fixed effects in Model 3, given a neutral (= 0) overall graph preference.

2.7 Discussion

In this study, we investigated the relationship between graph readers' preferential processing of bar versus line graphs when solving both difference tasks and trend tasks. Theoretical approaches and empirical studies suggest that bar graphs have computational advantages for difference tasks, whereas line graphs are advantageous for trend tasks, facilitating the task-directed processing of the depicted data (e.g., Pinker, 1990; Shah & Hoeffner, 2002). We investigated (1) participants' relative graph preference for bar versus line graphs, and (2) the preference development across time, in order to determine whether graph readers are able to detect and use the advantageous graph type when solving difference and trend tasks. In an experimental setting, we employed a dual representation paradigm (i.e., displaying two graph types at the same time) and eye-tracking measures to determine the preferential processing of bar versus line graphs.

Our results show that the graph readers in our sample of university students had, on average, a stronger preference for using bar graphs than for using line graphs in difference tasks and vice versa in trend tasks, indicating that they had a tendency to choose the

computationally advantageous graph type as a function of the task. However, we also found an increase in the preference for line graphs in both task conditions, resulting in a strong preferential processing of line graphs in trend tasks and a balanced use of both graph types in difference tasks.

2.7.1 Overall graph preference

To control for individual differences in preference beyond the experimental manipulation, we asked for participants' overall preference for bar graphs versus line graphs. The inclusion of this covariate into the LMM revealed a positive relation to the relative graph preference, which provides tentative evidence that the eye-movement-based measure of preference was a valid indicator of participants' actual graph preference. Additionally, it supports the conclusion by Shah (2002) that graph comprehension is a product of both bottom-up and top-down processes, which seem to consist not only of expectations and familiarity with content, but also of an individual preference for a certain type of representation (i.e., the graph type). However, the fixed effect of the overall graph preference was relatively small compared to the fixed effect of the task type (Table 2). In accordance with this, the context (e.g., the task type) in which a graph is presented seems to be far more predictive of graph readers' processing behavior than a judgment of their overall graph preference.

2.7.2 Preferential graph processing

The results from our final LMM (M3) show that participants shifted their preference from bar graphs in difference tasks to line graphs in trend tasks. Additional t-tests using aggregated data revealed that the difference in total fixation times on the two graphs was significant in both task conditions. However, the effect size was approximately twice as large in the trend task, suggesting that participants had an even stronger tendency towards the advantageous graph type when processing trend tasks. One explanation for this finding might be that the computational advantages of line graphs in trend tasks are stronger than those of bar graphs in difference tasks, due to working-memory limitations: When solving trend task items using bar graphs, participants always had to consider all five to eight data points at once to constitute a trend (Pinker, 1990), thereby operating at the limits of working memory (Baddeley, 1994; Baddeley & Hitch, 1972; Miller, 1956). On the other hand, when solving trend task items using line graphs, they only needed to inspect a single line, which puts relatively small demands on working memory, in comparison to the demands placed by bar

Kapitel 2 - Do Graph Readers Prefer the Graph Type Most Suited to a Given Task?

graphs. This difference in cognitive demands might explain the larger effect size in the trend task.

Looking at the difference task, even though comparing data points on a single line may be more difficult according to the Gestalt laws of perceptual grouping (Pinker, 1990) finding the correct positions among only five to eight data points may not have been too difficult. However, the true-false statements used in difference tasks may be responsible for the smaller effect size in this condition. Before examining the graphs, participants need to read the statement and keep in mind the relation that is described. While statements in the trend task only refer to single categories, those in the difference task explicitly mention multiple specific data points that must be compared. Thus, engaging with the statement in the difference task may be more difficult and demanding than in the trend task. This assumption is supported by the finding that participants spent significantly more time on the statement area when processing difference tasks. In the framework of Cognitive Load Theory (Sweller, 1988), it is assumed that cognitive processing is constrained by a limited working memory capacity and that the inherent characteristics of a task may put a higher demand on cognitive processes (intrinsic cognitive load; De Jong, 2010). Based on this, it seems possible that statements of the difference task put a greater cognitive load on participants. Since participants' cognitive resources were burdened with keeping multiple data points in mind before even tending to the graph itself, their ability to choose the advantageous graph type in difference tasks may have been impaired. However, there might be another explanation: Participants unsystematically reported that they sometimes used both graph types to double-check their answer. This is a strategy that has been reported in earlier research. Peebles and Cheng (2003) argued that such a strategy helped graph readers to memorize the chunk of information previously viewed. In the dual representation paradigm, participants may have expanded this strategy to repeatedly revisit the same regions in both graphs. A post-hoc analysis of gaze transitions between the statement area and the graph pattern areas revealed a significantly higher average number of gaze transitions in trials of the difference task compared to trials of the trend task ($M_{\text{Difference}} = 4.05$, $M_{\text{Trend}} = 2.11$, $t_{23} = 7.94$ $p < .001$, $d = 1.62$), supporting this assumption. In summary, the statements in the difference task seemed to be more demanding, causing participants to verify their answer by checking both graph types repeatedly.

2.7.3 Preference Development

We hypothesized that participants would develop a stronger preference for bar graphs when solving difference task items and a stronger preference for line graphs when solving trend task items. To investigate this, we compared participants' relative graph preference across the experimental trials. The results of the LMM (M3) suggest that participants developed a stronger preference for line graphs in both task conditions, confirming only one of our hypotheses (2b: preference for bar graphs increases over the course of processing difference tasks.), but not the other (2a: preference for line graphs increases over the course of processing trend tasks).

While the described development was expected in the trend task, it was a counterintuitive observation in the difference task. As there was no significant trial by task type interaction, the fixed effect of the trial number was very similar in both task conditions. Yet, the meaning of this development seems to be different for the two task types: In the trend task, on average participants started with a negative relative graph preference (i.e., preference for line graphs) that became stronger over time (this in line with our hypothesis 2b). In the difference task, on the other hand, participants started, on average, with a positive relative graph preference (i.e., preference for bar graphs) that decreased over time and became about zero (i.e., no preference for a particular graph, meaning fixation times on bar and line graphs were about equal). While this is contrary to our hypothesis 2a, it does not reflect a complete shift in preference, but rather a balanced use of both graph types when processing difference tasks. There might be more than one explanation for this finding: First, participants increasingly tended to use both graph types in the difference task, because verifying the answer proved helpful to evaluate the statements. However, because of the participants' generally high accuracy, it remains unclear if this strategy was actually advantageous. Or second, participants generally developed a higher preference for line graphs over the course of the trials. This might be due to a transfer effect (for transfer effects in graph tasks see Baker, Corbett & Koedinger, 2001) from the trend tasks in which line graphs were more clearly advantageous as suggested by the higher effect size in this condition.

2.8 Conclusion

Computational properties of graphs have been investigated extensively in the past (e.g. Peebles & Cheng, 2003; Pinker, 1990; Simkin & Hastie, 1987; Shah, Mayer, & Hegarty, 1999; Zacks & Tversky, 1999). It is well established that line graphs are advantageous for trend tasks, while bar graphs are more helpful for difference tasks. However, there is little

Kapitel 2 - Do Graph Readers Prefer the Graph Type Most Suited to a Given Task?

knowledge about graph readers' ability to detect and use these properties. The current study contributes to the field of graph comprehension by investigating graph readers' ability to choose the graph type most suited to a given task. Eye-tracking methodology was applied to assess graph readers' preferential processing of bar versus line graphs in a dual representation paradigm. Data showed that computational advantages of bar and line graphs were reflected by participants' preference towards the advantageous graph as a function of the task. Graph readers preferred to use the graph type most suited to a given task, suggesting that computational advantages in graphs are readily available to graph users, even in environments that provide alternative representations.

Still, there are limitations to our study that need to be considered. First, the true-false items used in the graph tasks were not designed to be sensitive to differences in graph readers' task performance and thus caused a ceiling effect in the performance data. Therefore, it remains unclear whether the observed preference patterns actually resulted in higher performance. Future studies could employ more complex tasks to investigate the connection between preferential processing and task performance.

Second, it is important to note that we used self-report data to assess participants' overall graph preference. Since it was not possible to verify the validity of participants' self-reports, this data should be interpreted with caution.

Third, we discovered that the true-false statements used in the difference task were relatively more demanding than those used in the trend task, which seems to have led participants to use both graphs in order to verify their answer. Additional research is needed to clarify the use of verifying strategies when multiple representations are available, for example, by complementing eye-tracking analyses with verbal reports (Ericsson & Simon, 1980; van Gog, Kester, Nieuvelstein, Giesbers, & Paas, 2009).

Fourth, we only investigated two different task types to analyze the preferential processing of bar graphs versus line graphs. Future studies could expand this research to additional tasks (especially more complex tasks) and more graph types (e.g., circle charts).

Finally, the number of participants in the current study was relatively small. However, by using a high number of experimental trials and employing linear mixed-effects models to account for the clustered data structure, we were able to compensate for this shortcoming. Still, future studies could collect data from larger samples to increase the reliability and generalizability of the results.

The findings and limitations of this study lead to some suggestions for future research. An analysis across trials revealed a general increase in the preference for line graphs, resulting

Kapitel 2 - Do Graph Readers Prefer the Graph Type Most Suited to a Given Task?

in a strong preference for line graphs in trend tasks and a balanced use of both graph types in difference tasks. This unexpected preference development towards line graphs in both graph tasks poses new questions: Is the use of two graphs superior to using only one graph? May the graph preference be influenced by transfer effects between different tasks? Future research should investigate processing of multiple graphs compared to single representations to clarify if the use of multiple graphs can be advantageous for some tasks. Alternative explanations, such as transfer effects between multiple tasks, could also be investigated.

Furthermore, it should be noted that the random effects in the LMMs showed a substantial variation in the observed graph preference between participants, emphasizing the need to further explore individual differences in the domain of graph comprehension. Additionally, high processing times in the difference task revealed that these items may put a higher demand on working memory. Even though working memory has been discussed as a limiting factor for graph comprehension (e.g., Lohse, 1997; Pinker, 1990; Shah & Hoeffner, 2002; Trickett & Trafton, 2006), it has rarely been focused on. Only recently, Halford, Baker, McCredden and Bain (2005) found that the human processing capacity is limited to four variables in one graphical display. Thus, future research should consider individual variables, such as working memory capacity, alongside processing and performance data. A combination of these approaches could provide a deeper insight into graph readers' ability to make use of computational properties, and into graph comprehension in general.

2.9 References

- Ali, N., & Peebles, D. (2013). The effect of gestalt laws of perceptual organization on the comprehension of three-variable bar and line graphs. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 55(1), 183-203.
- Baddeley, A. D. (1994). The magical number seven: still magic after all these years? *Psychological Review*, 101, 353–356. doi:10.1037/0033-295X.101.2.353
- Baddeley, A. D., & Hitch, G. (1974). Working Memory. *Psychology of Learning and Motivation*, 8, 47–89. doi:10.1016/s0079-7421(08)60452-1
- Baker, R. S., Corbett, A. T. & Koedinger, K. R. (2001). Toward a model of learning data representations. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society* (pp. 45-50). Mahwah, NJ: Erlbaum.
- Barr, D. J. (2008). Analyzing ‘visual world’ eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59, 457–474. doi:10.1016/j.jml.2007.09.002
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278. doi:10.1016/j.jml.2012.11.001
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, 1–48. doi:10.18637/jss.v067.i01.
- Bell, A., & Janvier, C. (1981). The Interpretation of Graphs Representing Situations. *For the Learning of Mathematics*, 2, 34–42. Retrieved from <http://eric.ed.gov/?id=EJ255422>
- Bertin, J. (1983). *Semiology of graphics: diagrams, networks, maps*. Madison, WI: University of Wisconsin Press.
- Blackwell, A. F. (2001). Introduction Thinking with Diagrams. In *Thinking with Diagrams* (pp. 1–3). Netherlands: Springer.
- Cucuo, A. A., & Curcio, F. (2001). *The role of representation in school mathematics: National Council of Teachers of Mathematics Yearbook*. Reston, VA: NCTM.
- Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education*, 18, 382–393. doi:10.2307/749086

Kapitel 2 - Do Graph Readers Prefer the Graph Type Most Suited to a Given Task?

- De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science*, 38, 105–134. doi:10.1007/s11251-009-9110-0
- Djamasbi, S., Siegel, M., Skorinko, J., & Tullis, T. (2011). Online viewing and aesthetic preferences of generation y and the baby boom generation: Testing user web site experience through eye tracking. *International Journal of Electronic Commerce*, 15(4), 121-158. doi:10.2753/JEC1086-4415150404
- Dorman, J. P. (2008). The effect of clustering on statistical tests: an illustration using classroom environment data. *Educational Psychology*, 28, 583–595. doi:10.1080/01443410801954201
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87, 215–251. doi:10.1037/0033-295X.87.3.215
- Glazer, N. (2011). Challenges with graph interpretation: A review of the literature. *Studies in Science Education*, 47, 183–210. doi:10.1080/03057267.2011.605307
- Goldberg, J., & Helfman, J. (2011). Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. *Information Visualization*, 10, 182–195. doi:10.1177/1473871611406623
- Grosse, M. E., & Wright, B. D. (1985). Validity and reliability of true-false tests. *Educational and Psychological Measurement*, 45, 1–13. doi:10.1177/0013164485451001
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items*. New York, NY: Routledge.
- Halford, G. S., Baker, R., McCredden, J. E., & Bain, J. D. (2005). How many variables can humans process?. *Psychological Science*, 16, 70–76. doi:10.1111/j.0956-7976.2005.00782.x
- Hartig, J., & Buchholz, J. (2012). A multilevel item response model for item position effects and individual persistence. *Psychological Test and Assessment Modeling*, 54, 418–431.
- Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: foundations and illustrative examples. *Behavior Research Methods*, 39, 101–117. doi:10.3758/BF03192848

Kapitel 2 - Do Graph Readers Prefer the Graph Type Most Suited to a Given Task?

- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & De Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford, UK: Oxford University Press.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological Review*, *87*, 329–354. doi:10.1037/0033-295X.87.4.329
- Kim, S. & Lombardino, L. (2015). Comparing graphs and text: Effects of complexity and task. *Journal of Eye Movement Research*, *8*(3), 1–17.
- Koerber, S. (2011). Der Umgang mit visuell-grafischen Repräsentationen im Grundschulalter. *Unterrichtswissenschaft*, *39*, 49–62. doi:10.3262/UW1101049
- Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, *3*, 185–225. doi:10.1002/acp.2350030302
- Larkin, J. H. & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, *11*, 65–100. doi:10.1111/j.1551-6708.1987.tb00863.x
- Lohse, G. L., Biolsi, K., Walker, N. & Rueter, H. H. (1994). A classification of visual representations. *Communications of the ACM*, *37*, 36–49. doi:10.1145/198366.198376
- Lindner, M. A., Eitel, A., Thoma, G.-B., Dalehefte, I. M., Ihme, J. M., & Köller, O. (2014). Tracking the decision making process in multiple-choice assessment: Evidence from eye movements. *Applied Cognitive Psychology*, *28*, 738–752. doi:10.1002/acp.3060
- Lowrie, T., & Diezmann, C. M. (2007). Solving graphics problems: Student performance in junior grades. *The Journal of Educational Research*, *100*, 369–378. doi:10.3200/JOER.100.6.369-378
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511811678
- Meissner, M., & Decker, R. (2010). Eye-tracking information processing in choice-based conjoint analysis. *International Journal of Market Research*, *52*, 591–610. doi:10.2501/S147078531020151X
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *101*, 343–352. doi:10.1037/h0043158

Kapitel 2 - Do Graph Readers Prefer the Graph Type Most Suited to a Given Task?

- Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 259–303). Hillsdale, NJ: Lawrence Erlbaum.
- Peebles, D., & Cheng, P. C. H. (2003). Modeling the effect of task and graphical representation on response latency in a graph reading task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 45, 28–46. doi: 10.1518/hfes.45.1.28.27225
- Pereira-Mendoza, L., Goh, S. L., & Bay, W. (November, 2004). Interpreting graphs from newspapers: Evidence of going beyond the data. Paper presented at the ERAS Conference, Singapore. Retrieved from https://repository.nie.edu.sg/bitstream/10497/15549/1/ERAS-2004-277_a.pdf
- Pinheiro, J., & Bates, D. (2006). *Mixed-effects models in S and S-PLUS*. New York, NY: Springer.
- Pinker, S. (1990). A theory of graph comprehension. In R. O. Freedle (Ed.), *Artificial Intelligence and the Future of Testing* (p. 73–126). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.
- Purchase, H. C. (2014). Twelve years of diagrams research. *Journal of Visual Languages & Computing*, 25, 57–75. doi:10.1016/j.jvlc.2013.11.004
- Quené, H., & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, 59, 413–425. doi:10.1016/j.jml.2008.02.002
- R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Ren, X., Wang, T., Altmeyer, M., & Schweizer, K. (2014). A learning-based account of fluid intelligence from the perspective of the position effect. *Learning and Individual Differences*, 31, 30–35. doi:10.1016/j.lindif.2014.01.002
- Schnotz, W. (1994). Wissenserwerb mit logischen Bildern. In B. Weidenmann (Ed.), *Wissenserwerb mit Bildern* (pp. 95–148). Bern: Huber.
- Schnotz, W. (2002). Commentary: Towards an integrated view of learning from text and visual displays. *Educational Psychology Review*, 14, 101–120. doi:10.1023/A:1013136727916

Kapitel 2 - Do Graph Readers Prefer the Graph Type Most Suited to a Given Task?

- Shah, P. (1997). A model of the cognitive and perceptual processes in graphical display comprehension. In *Proceedings of American Association for Artificial Intelligence Spring Symposium*, AAAI Technical Report FS-97-03. Stanford University.
- Shah, P. (2002). Graph comprehension: The role of format, content, and individual differences. In Anderson, M., Meyer, B., and Olivier, P. (Eds.), *Diagrammatic Representation and Reasoning*, Springer, New York.
- Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review*, *14*, 47–69. doi:10.1023/A:1013180410169
- Shah, P., & Carpenter, P. A. (1995). Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology*, *124*, 43–61. doi:10.1037/0096-3445.124.1.43
- 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*, 690–702. doi:10.1037/0022-0663.91.4.690
- Simkin, D. & Hastie, R. (1987). An information-processing analysis of graph perception. *Journal of the American Statistical Association*, *82*, 454–465. doi:10.1080/01621459.1987.10478448
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (2nd ed.). Los Angeles, CA: SAGE.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*, 257–285. doi:10.1207/s15516709cog1202_4
- Tan, J. K., & Benbasat, I. (1990). Processing of graphical information: A decomposition taxonomy to match data extraction tasks and graphical representations. *Information Systems Research*, *1*, 416–439.
- Trickett, S. B., & Trafton, J. G. (2006). Toward a comprehensive model of graph comprehension: Making the case for spatial cognition. In D. Barker-Plummer, R. Cox, & N. Swaboda (Eds.), *Diagrammatic representation and inference* (pp. 286–300). Berlin, Germany: Springer.
- van Gog, T., Kester, L., Nievelstein, F., Giesbers, B., & Paas, F. (2009). Uncovering cognitive processes: Different techniques that can contribute to cognitive load research and instruction. *Computers in Human Behavior*, *25*, 325–331. doi:10.1016/j.chb.2008.12.021

Kapitel 2 - Do Graph Readers Prefer the Graph Type Most Suited to a Given Task?

Wainer, H. (1992). Understanding graphs and tables. *Educational Researcher*, 21, 14–23.
doi:10.3102/0013189X021001014

Wertheimer, M. (1938). Laws of organization in perceptual forms. In W. D. Ellis (Ed.). *A source book of Gestalt psychology*. London: Routledge & Kegan Paul.

Winn, W. D. (1990). A theoretical framework for research on learning from graphics. *International Journal of Educational Research*, 14, 553–564. doi:10.1016/0883-0355(90)90025-4

Zacks, J., & Tversky, B. (1999). Bars and lines: a study of graphic communication. *Memory & Cognition*, 27, 1073–1079. doi:10.3758/BF03201236

Kapitel 3

Studie 2

Task-irrelevant Data Impair Processing of Graph Reading Tasks: An Eye Tracking Study

Strobel, B., Lindner, M. A., Saß, S., & Köller, O. (2018). Task-irrelevant data impair processing of graph reading tasks: An eye tracking study. *Learning and Instruction*. Advance online publication. doi:10.1016/j.learninstruc.2017.10.003

Abstract

For Instruction, teachers often rely on prefabricated material that may include irrelevant information. However, graphs can place a heavy burden on the cognitive system if their complexity is not suitable for a given task. In this study, we compared bar graphs showing task-irrelevant data points or task-irrelevant data series with a control condition using a within-subject design and eye tracking methodology. Data were analyzed using linear mixed-effects models. Results show that task-irrelevant data significantly elevated processing time, error rate and cognitive load. Even though perceptual grouping by color was expected to aid the process when a task-irrelevant data series was included in a graph, effects were strongest in this condition. Analyses of attention distribution using eye tracking measures revealed that task processing differed qualitatively between the conditions, yielding important implications for instruction.

Keywords: Graph comprehension; task processing; cognitive load; eye tracking; linear mixed-effects models

Task-irrelevant Data Impair Processing of Graph Reading Tasks: An Eye Tracking Study

Visualizations such as graphs and diagrams play an important role in everyday life and can be found in newspapers, science, engineering, and education (e.g., Glazer, 2011; Mayer, 2009; Pereira-Mendoza, Goh, & Bay, 2004; Purchase, 2014). They are especially important in the context of problem solving (Baker, Corbett, & Koedinger, 2001), for teaching and learning mathematics (Cucuo & Curcio, 2001) and for understanding scientific data (Shah & Hoeffner, 2002).

In learning, graphs provide cognitive support by offering computational advantages, such as perceptual grouping (see Wertheimer, 1938) and serving as external memory (Larkin & Simon, 1987; Tory & Möller, 2004). In contrast, inferences from non-spatial representations (e.g., text) are often more demanding, because some information must be computed at great cognitive expense (e.g., comparing numbers in a continuous text). However, graphs do not always make comprehension of information more effective and less demanding. In many learning situations graphs display more information than is relevant to readers, because teachers often rely on prefabricated material from textbooks and other sources. For example, in a graph that displays population sizes over several years, only a subset of years might be of interest to a learner (e.g., in a comparison task where a learner has to compare a number of specific data points). In this example, *task-irrelevant data* contribute to the overall *complexity* of the graphs, resulting in a higher complexity that is unnecessary for a learner to complete the given tasks.

Even though data complexity has rarely been the focus of graph comprehension research, it has been investigated as a background variable in several studies (e.g., Casali & Gaylin, 1988; Meyer, Shinar, & Leiser, 1997; Schutz, 1961a, 1961b; Spence & Lewandowsky, 1991). Still, more recently published studies provide some evidence that a high complexity negatively impacts performance in regular graph tasks (e.g., Kim & Lombardino, 2015; Kumar & Benbasat, 2004). Understanding a graph can be especially challenging if its complexity is not suitable for the task, placing a heavy burden on the cognitive system (Huang, Eades & Hong, 2009; Demetriadis & Cadoz, 2005; Sedig & Parsons, 2013). Because human memory is a limited capacity information processing system (Baddeley & Hitch, 1974), this burden may result in a cognitive overload for the reader

(Sweller, 1994). But does complexity affect graph processing and related task performance even if the additional data are completely irrelevant to the task at hand?

In the present study we focus on task-irrelevant data as a source of complexity in graphs. We compare different situations of task-irrelevant data: (1) When a given series of data points includes more data points than necessary to complete a task (*task-irrelevant data points*) and (2) when a second, task-irrelevant series of data points is presented next to a relevant data series (*task-irrelevant data series*). We investigate effects of task-irrelevant data on error rates, processing time and cognitive load. Additionally, (3) we explore underlying processes by applying eye tracking methodology, which has proven useful in previous graph comprehension studies (e.g. Kim & Lombardino, 2015; Strobel, Saß, Lindner, & Köller, 2016).

3.1 Data complexity

Data complexity is for the most part reflected in the number of data points and variables displayed in a graph. According to Meyer et al. (1997), data complexity can be understood as a result of three factors: (a) the number of data points in a graph, (b) the configuration of the data points (i.e., the organization of data points into data series), and (c) the regularity of the data².

Schutz (1961a; 1961b) was one of the first researchers to investigate the number of data points. Looking at tasks that require the identification of trends, he found that processing time increased with the number of irrelevant data points regardless of the graph type (bar and line graphs), but they did not exhibit differences in accuracy. In line with these findings, Kumar and Benbasat (2004) found elevated processing times when additional data points were included in a graph. In a series of three experiments, Spence and Lewandowsky (1991) compared multiple types of representations including bar and line graphs. In contrast to Schutz, they found that accuracy in regular comparison tasks (i.e., the comparison of two or more data points) was significantly lower when more data points were present.

Regarding the organization of data into data series, Schutz (1961b) found that processing time increased with the number of data series in line graphs, but only if multiple graphs were used (i.e., one graph per data series). Single graphs displaying multiple lines were unaffected by the data series number. Recently, Kim and Lombardino (2015) conducted an eye tracking study and varied the number of data series and the task type (i.e., single point

² Data regularity is not investigated in the present study, because task-irrelevant data is not the source of complexity in this factor (for findings on data regularity see Kumar & Benbasat, 2004; Schutz, 1961b).

location vs. point comparison). They found that processing time in both task types was significantly higher when an additional data series was present. This was also reflected in longer fixation times on the graph regions.

Data complexity has consistently shown effects on task performance. In addition to performance measures, Huang et al. (2009) exhibited cognitive load with a subjective rating scale for mental effort and found that reported mental effort increased with the level of data complexity. Along with processing times and task performance, cognitive load measures are important to reveal effects on the burden of working memory.

3.2 Cognitive load theory

Cognitive Load Theory (CLT; Sweller, 1988) describes how cognitive processing can be facilitated or inhibited under the constraints of a limited working memory capacity. Three types of cognitive load are distinguished in CLT: *Intrinsic load* refers to the inherent complexity of the information, especially the number of interactive elements that must be understood in relation to each other. *Extraneous load* originates from the instructional material and involves unnecessary processing of irrelevant or unrelated information. Finally, *germane load* describes the mental effort invested by the learner to comprehend the material and involves processes such as interpreting and organizing (de Jong, 2010; Sweller 1994).

A high data complexity in graphs may result in cognitive overload if the complexity is not suitable for the task (Huang et al., 2009; Demetriadis & Cadoz, 2005; Sedig & Parsons, 2013). However, in many of the presented studies the additional data were relevant to the task. In the context of CLT, there are two reasons to distinguish data that are required to solve a given task from task-irrelevant data. First, data necessary to complete a given task contribute to intrinsic cognitive load, because the information that must be extracted is inherently more complex. Task-irrelevant data on the other hand are expected to induce extraneous cognitive load because they give rise to unnecessary processing of irrelevant information. Second, while a higher burden on working memory is to be expected when a bigger amount of data *must* be processed (i.e., when additional data is required to complete the task), it is unclear if and to what extent task-irrelevant data are processed during task-oriented graph reading. For example, if the oil price of two specific years is of interest to a learner, giving the prices for a wide range of years (i.e., showing task-irrelevant data points) might induce additional load and make the task more difficult. When data points are inserted into an existing data series like in this example, readers have to refer to the corresponding labels (i.e., the specific years) in order to distinguish relevant from irrelevant data. Given an

additional data series on the other hand, a color-coded legend allows for perceptual grouping of information in the graph (e.g., Pinker, 1990; Freedman & Shah, 2002). Here, task-relevant and task-irrelevant data points are easily distinguishable by color, eliminating the need to check multiple labels and thus facilitating task processing.

In summary, task-irrelevant data can be expected to induce extraneous cognitive load, to extend the processing time of the task and to make this process less accurate. Based on the principles of perceptual grouping, these effects may be mitigated when task-irrelevant data can be identified by color.

3.3 Using eye tracking to gain insight into processing of graph reading tasks

The use of eye tracking (for an introduction, see Duchowski, 2007; Holmqvist et al., 2011) in cognitive psychology is based on the assumption that the location of eye-fixations represents the focus of attention. In other words, it is assumed that we process the visual information we are currently looking at. This idea is called the eye-mind hypothesis (Just & Carpenter, 1980). Even though shortcomings of the eye-mind hypothesis have been discussed (Hyönä, 2010; Wright & Ward, 2008), researchers have shown that eye-fixation measures and cognitive performance are closely related (e.g. Canham & Hegarty, 2010; Jarodzka, Scheiter, Gerjets, & van Gog, 2010; Lindner, Eitel, Strobel & Köller, 2017). In contrast to other process tracing methods (e.g., verbal protocols; Ericsson & Simon, 1980; van Gog, Kester, Nievelstein, Giesbers, & Paas, 2009), eye tracking does not place additional load on participants' working memory.

Mayer (2010) suggested that eye tracking technology can contribute to the study of learning with graphics by providing information relevant to the instructional design of graphics. In traditional experiments, we may draw conclusions about if and when a manipulation of the material has an effect on performance measures (e.g., accuracy, processing time). However, we often are clueless on *how* processing of the material changes. In the context of graph reading, eye tracking allows us to allocate processing time to the spatiotemporal attention of the graph reader. In a number of recent graph studies, researchers were able to successfully attribute processing time to important subregions of graph tasks (i.e., x-axis, y-axis, legend, pattern, question, answers; Kim & Lombardino, 2015; Peebles & Cheng, 2003; Strobel et al., 2016). For the present study, eye tracking enables us to estimate how much time can be attributed to the processing of specific regions in the display, such as task-irrelevant data points in graphs.

3.4 Research hypotheses

In this study, we compared two different types of data complexity: (1) task-irrelevant data points (DP) that are included in an existing data series and (2) a task-irrelevant data series (DS) presented next to an existing data series in comparison to a control condition with no task-irrelevant data. Additionally, (3) we explored differences in task processing by using eye tracking methodology. In summary, we expected the following:

- (1) *Task-Irrelevant Data Points*: If task-irrelevant DP are inserted into an existing DS, graph readers need to examine their labels in order to distinguish task-irrelevant from relevant data. Accordingly, we assume that graph readers are less accurate, show longer processing times and report a higher cognitive load in comparison with the control setting.
- (2) *Task-Irrelevant Data Series*: If a task-irrelevant DS is presented next to an existing DS, graph readers need to examine the legend in order to distinguish task-irrelevant from relevant data. We assume that graph readers are less accurate, show longer processing times and report a higher cognitive load in comparison with the control setting. These effects are expected to be smaller than those of task-irrelevant DP, because perceptual grouping is expected to facilitate the processing of relevant data.
- (3) *Differences in Task Processing*: We expect that task-irrelevant data (DP and DS) will require additional processing time as reflected in longer eye-fixations on task-irrelevant data. Furthermore we hypothesize that reading the labels and legends takes up more processing time than in the control setting. Processing times of the question and answers are expected to be unaffected by additional data.

3.5 Method

3.5.1 Sample and study design

The participants in our study were $N = 60$ university students from different faculties (66% female, $M_{\text{age}} = 23.35$ years). All participants had normal or corrected to normal vision. Data from 41 trials (5.6 %) were excluded from the analysis due to poor quality of the eye tracking data. The assessment comprised a set of graph tasks as well as a short paper-pencil questionnaire to assess demographic information and participants' graph literacy. For the

graph tasks we employed a within-subject design in which we varied the presence of task-irrelevant data in three qualities (no irrelevant data vs. irrelevant DP vs. an irrelevant DS). Participants completed four trials of the graph reading task in each experimental condition, resulting in a total of 12 trials.

As dependent variables we assessed total processing time, the error rate and a rating of cognitive load. Additionally, we collected eye tracking data to determine the amount of time devoted to different areas of the task material. Finally, we adapted a short graph literacy scale from Strobel et al. (2016) to measure the participants' level of experience in graph reading.

3.5.2 Material and measures

3.5.2.1 Graph task

The graph task was administered computer-based. We used only bar graphs to avoid effects of the graph type on task performance (see e.g. Peebles & Cheng, 2003; Strobel et al., 2016). To complete each item of the task, participants had to compare multiple data points in the graph, find an optimal combination according to a given criterion and then select the correct answer among four options in a classical multiple-choice item format. Each item of the graph task consisted of a lead-in statement, introducing the topic of the current task, followed by a bar graph, the question and the multiple choice options. The position of the correct answer was determined by a one-time randomization to avoid systematic tendencies that allow testwise participants to identify the correct answer without completing the actual task (see e.g. Dodeen, 2008). As a response, participants selected their answer by mouse click. Items of the graph task could only be completed using information derived from the given graph and did not require any specific prior knowledge. With regards to content, we used a variety of topics (e.g., food and sports).

In the bar graphs, the x-axis was categorial (e.g., food name), while the y-axis showed quantitative data (e.g., number of calories). In the control condition of the graph task, the graphs showed six data points that were required to identify the correct answer. There were no irrelevant data points. As experimental variations we either included four additional data points or one additional data series in the graph that were not necessary for completing the task. The position of irrelevant data points on the x-axis was chosen by a one-time randomization. Additional data series were identified by color and a corresponding legend. Each bar of the new data series was assigned to the corresponding value on the x-axis, next to the existing bar of the first data series. This way the axes regions of the graph remained unchanged to ensure comparability among the experimental conditions. Thus, perceptual

grouping was enabled by color, and not by proximity. Identical labels were used in all conditions. Lead-in statements, item stem and multiple choice options did not differ in the experimental conditions. Figure 1 shows an example item including graphs from all three conditions.

3.5.2.2 Cognitive load measurement

For the present study we used a rating scale for perceived difficulty which is an established measure of cognitive load (e.g., Ayres, 2006; Kalyuga, Chandler, & Sweller, 1998)³. Participants rated their perceived difficulty of each item on a unidimensional 5-point rating scale (from 1 = *low* to 5 = *high*). Unidimensional rating scales are known to be sensitive to small differences in cognitive load as well as to be valid, reliable and non-intrusive (e.g., Paas & van Merriënboer, 1994; Paas, van Merriënboer, & Adam, 1994).

3.5.2.3 Graph literacy

A short questionnaire was employed to survey participants' graph literacy using 5-point Likert-scaled items. The scale comprised four items (i.e., "I am familiar with bar graphs", "I have used bar graphs before", "I often use graphs", "I feel confident in reading graphs."). The graph literacy scale yielded a good reliability (Cronbach's $\alpha = .71$).

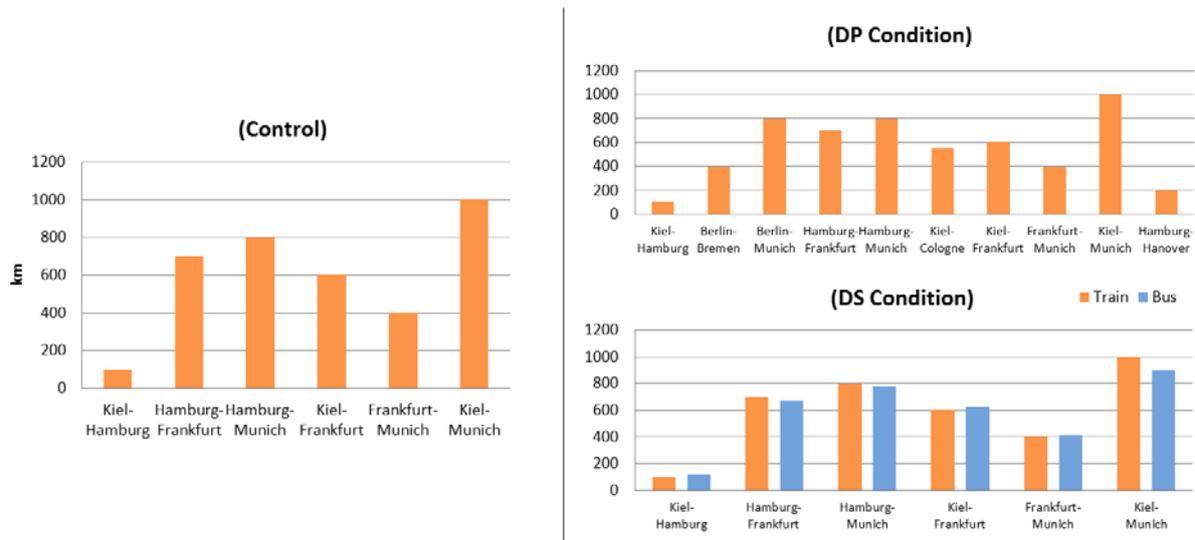
3.5.3 Apparatus

Items were presented on a 19-inch screen with a 1280×1024 pixel resolution, using the software Experiment Center 3.5 from SensoMotoric Instruments (SMI, Teltow, Germany). Each item was presented on a single screen. Participants sat in front of the screen at a distance of approximately 70 centimeters while working on the graph task. The font size of the text was 2 centimeters (approx. 1.6° visual angle). Participants' eye-movements were recorded using a video-based remote eye tracking system (SMI iView XTMRED-m; 120Hz sampling rate) and the corresponding SMI software iView XTM. The system was calibrated using an animated 8-point calibration image and subsequent validation. The calibration accuracy was below 0.5° visual angle for all participants on both the x and y coordinates (range: 0.08 to 0.49; $M_x = 0.31$, $SD_x = 0.10$; $M_y = 0.30$, $SD_y = 0.10$).

³ For a critical view on the use of this technique see Cierniak, Scheiter and Gerjets (2009) as well as Schnotz and Kürschner (2007).

Kapitel 3 - Task-irrelevant Data Impair Processing of Graph Reading Tasks

You want to take the train from Kiel to Munich. To protect the environment, you want to reduce emissions and take the shortest route. You check the train schedules and find four possible connections.



Which train connection takes the shortest route according to the schedules?

(A) From Kiel directly to Munich (B) From Kiel via Frankfurt to Munich (C) From Kiel via Hamburg and Frankfurt to Munich (D) From Kiel via Hamburg to Munich

Figure 1. Translated example item showing the graphs in all three conditions. DP = Data Points, DS = Data Series.

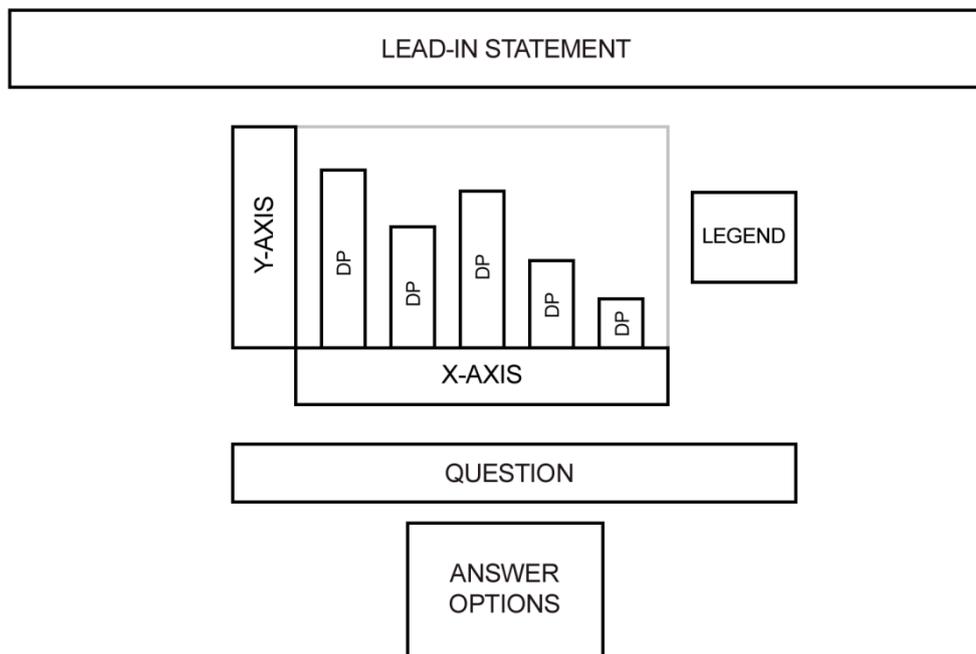


Figure 2. Schematic Representation of the Areas of Interest (AoI) defined for the eye tracking analysis in the control condition. In addition to the displayed AoIs, the experimental conditions include AoIs for irrelevant data. DP = Data Point

3.5.4 Procedure

Students were tested in single sessions. First, they were familiarized with the procedure and the eye tracking system; after this, they completed the graph task on a computer while their eye-movements were recorded. After each of the twelve trials, participants were asked to rate the perceived difficulty. After participants completed the graph task, a paper-pencil questionnaire was administered. The whole cycle of assessment took about 45 minutes to complete.

3.5.5 Analysis

3.5.5.1 Eye-Movement Data Pre-Processing

Eye-movement recordings were analyzed using a dispersion-based algorithm implemented in the Begaze™ software, version 3.5, from SMI. A fixation was detected when eye movements lasted for at least 80 milliseconds on a position with a maximum dispersion of 100 pixels. To determine whether eye-movement data were recorded correctly, participants' scan paths were visually inspected.

As gaze data we used the total fixation time on predefined Areas of Interest (AoIs), which is defined as the cumulative duration of all AoI fixations from trial onset to trial end, indicating the total time devoted to a specific area (Holmqvist et al., 2011). Total fixation times provide a valid indirect measure of attention distribution and cognitive processing in educational tasks using the multiple-choice item format (e.g. Lindner et al., 2014).

Similar to previous graph studies (e.g., Kim & Lombardino, 2015; Peebles & Cheng, 2003, Strobel et al., 2016), we divided the graph regions into multiple rectangular AoIs: x-axis, y-axis, legend and title (grouped together as *axes and labels*). Unlike traditional arrangements, each bar in the graph was covered by its own AoI in order to distinguish between fixations on relevant and irrelevant data points. Accordingly, these areas were combined into two groups, *relevant data* and *irrelevant data* (with the latter only occurring in the DP and DS conditions). Another six AoIs for the lead-in statement, the question and the four multiple-choice answer options were grouped together as *item stem and options*. All reported data refer to these AoI groups. See Figure 2 for a visualization of the AoI configuration.

3.5.5.2 Linear mixed-effects models

Data were analyzed using R, version 3.1.0 for Windows (R Core Team, 2016). With repeated measures nested within participants as well as in items, the data structure can be

described as clustered or hierarchical (Snijders & Bosker, 2012). Since measures were nested within two higher-level units, the design was cross-classified (i.e., each measure belongs uniquely to one participant and one item; Snijders & Bosker, 2012). Regular approaches, such as ANOVA models, yield inflated Type I error rates when the data are clustered (Dorman, 2008). To account for the clustered data structure, we applied linear mixed-effects models (LMMs; for introductions see Snijders & Bosker, 2012; Quené & van den Bergh, 2008). These models can be conceptualized “as a series of interrelated regression models that explain sources of variance at multiple levels of analysis, such as at the experimental stimuli and person levels” (Hoffman & Rovine, 2007, p. 102). More specifically, LMMs can model fixed effects and random effects simultaneously. While fixed effects aim to identify the typical rates of change in the outcome variable (e.g., following an experimental manipulation), random effects aim to identify unsystematic rates of change (e.g., due to differences between items and persons, respectively).

We used the R package lme4 (Bates, Maechler, Bolker, & Walker, 2015) to perform (generalized) linear mixed-effects analyses of the relationships between the experimental conditions and several outcome variables: error rate (M1), total processing time (M2), rating of cognitive load (M3), total fixation time on task regions (M4), total fixation time on axes and label regions (M5), total fixation time on relevant data points in the graph (M6) and total fixation time on irrelevant data points in the graph (M7). All models include random intercepts for both participants and stimuli to account for unsystematic differences between items and persons. Models were fitted by the Restricted Maximum Likelihood (REML) criterion, because it yields better Type I error rates for smaller groups when testing fixed effects than estimates with the Maximum Likelihood criterion (Manor & Zucker, 2004; Snijders & Bosker, 2012). All models were specified with the control condition being the reference group. We applied the delta method using the R package msm (Jackson, 2011) to calculate differences between the two experimental conditions. Degrees of freedom for statistical tests of fixed effects in the LMMs were determined by a conservative interpretation of a rule described in Snijders and Bosker (2012). That is, $df = N - q - 1$, where N is the minimum number of upper level units (here: $N = 12$ stimuli) and q is the number of explanatory variables included in the model (here: $q = 2$).

3.6 Results

Participants were asked to rate their graph literacy on a scale comprised of four items. Results show that the ratings of graph literacy in the current sample were relatively high ($M = 3.94$, $SD = 0.74$, range = 1.75 to 5)⁴.

Table

Fixed effects and random effects in the (generalized) linear mixed-effects models for error rate (M1), processing time (M2), and cognitive load (M3).

Fixed effect	Error rate (M1)			Processing time (M2)			Cognitive Load (M3)		
	Estimate	SE	z-value	Estimate	SE	t-value	Estimate	SE	t-value
Intercept	-1.30***	0.26	-5.11	74.18***	4.32	17.18	2.54***	0.12	20.58
Task-irrelevant DP	0.48*	0.22	2.20	5.86*	2.07	2.83	0.15*	0.06	2.33
Task-irrelevant DS	0.82***	0.22	3.80	13.21***	2.07	6.39	0.40***	0.06	6.29
Random effect	Variance Component			Variance Component			Variance Component		
Subject	0.57			248.30			0.30		
Stimulus	0.34			148.10			0.10		
Residual				480.30			0.50		

Note. * = $p < .05$ ** = $p < .01$ *** = $p < .001$

Table 2

Fixed effects and random effects in the linear mixed-effects models for total fixation times (in seconds) on task material (M4), axes and labels (M5), relevant data (M6) and irrelevant data (M7).

Fixed effect	Task material (M4)			Axes and labels (M5)			Relevant data (M6)			Irrelevant data (M7)		
	Estimate	SE	t-value	Estimate	SE	t-value	Estimate	SE	t-value	Estimate	SE	t-value
Intercept	30.40***	1.65	18.46	17.14***	1.49	11.50	15.30***	1.26	12.12	-0.03	0.50	-0.05
Task-irrelevant DP	1.34	0.76	1.77	3.42***	0.65	5.28	-2.91**	0.63	-4.65	2.97***	0.34	8.88
Task-irrelevant DS	4.40***	0.75	5.80	1.90*	0.65	2.95	-2.66**	0.62	-4.26	6.44***	0.33	19.25
Random effect	Variance Component			Variance Component			Variance Component			Variance Component		
Subject	33.95			27.82			17.41			1.18		
Stimulus	22.24			18.57			13.32			2.08		
Residual	64.83			46.92			43.89			12.60		

Note. * = $p < .05$ ** = $p < .01$ *** = $p < .001$

⁴ Originally, graph literacy was included in the linear mixed analyses, but the inclusion did not yield significant changes to the models. Thus, graph literacy was not included in the LMMs.

3.6.1 Error Rate

We used a generalized linear mixed-effects model (GLMM) to estimate the relationship between the participants' error rate and the experimental conditions (M1). We hypothesized that the error rate would be higher when task-irrelevant DP or a task-irrelevant DS were added to the graph. Effects in the DS condition are expected to be smaller than in the DP condition, because perceptual grouping may aid the graph reading process.

The estimates of the GLMM (Table 1) are given in log odds. The intercept shows the log odds for the control setting since this is the reference group in the model. Translating the intercept of the model into a frequency, the estimated error rate in the control setting was 21%. Compared to the control setting, we found that the estimated error rate was significantly elevated in the DP condition ($z = -2.20, p < .05$) as well as in the DS condition ($z = -3.81, p < .001$). Translated into changes in odds, the estimated error rate was 61% higher ($OR = 1.61$) in the DP condition, resulting in an error rate of 34%. In the DS condition, it was more than twice as big ($OR = 2.27$), resulting in an error rate of 47%. Using the delta method, we were able to confirm that the difference between the DP and DS conditions was also significant (Estimate = 0.34, SE = 0.21, $z = 1.67, p < .05$).

3.6.2 Total Processing Time

In Model 2 we estimated the relationship between the participants' total processing time of a graph task item and the experimental conditions (M2; Table 1). We hypothesized that the total processing time would be higher in the DP and DS conditions. Compared to the control setting, the estimated total processing times were significantly higher both in the DP condition ($t = 2.83, p < .05$) and in the DS condition ($t = 6.39, p < .001$). On average, the estimated total processing time was about 6 seconds longer when task-irrelevant DP were included and 13 seconds longer if an additional DS was included in the graph. Furthermore, the estimated processing time was significantly longer in the DS condition than in the DP condition (Estimate = 7.35, SE = 1.06, $t = 3.56, p < .01$).

3.6.3 Cognitive Load

Model 3 was used to estimate the relationship between the cognitive load rating and the experimental conditions (M3; Table 1). We hypothesized that cognitive load would be higher in the DP and DS conditions. Compared to the control condition, the estimated cognitive load rating was significantly higher both in the DP condition ($t = 2.33, p < .05$) and in the DS condition ($t = 6.29, p < .001$). On average, the estimated cognitive load rating

changed by 0.15 points on the 5-point rating scale when task-irrelevant DP were included and by 0.40 points when a task-irrelevant DS was included in the graph. Again, the delta method showed that the difference between the DP and the DS conditions was also significant (Estimate = 0.25, SE = 0.06, $t = 3.95$, $p < .01$).

3.6.4 Attention Distribution among AOIs

An overview of the average attention distribution among the AOIs for all three conditions is given in Figure 3. We used LMMs to model the relationships between the experimental conditions and the total fixations times on the four AOI groups, item stem and options (M4), axes and labels (M5), relevant data (M6) and irrelevant data (M7) to investigate the differences in item processing based on the experimental conditions. The results are given in Table 2.

3.6.4.1 Item Stem and Options

On average, the estimated total fixation time on item stem and options was about 30 seconds in the control setting. The estimated total fixation time did not differ significantly in the DP condition ($t = 1.77$, $p = .11$), but was significantly elevated in the DS condition ($t = 5.80$, $p < .001$). On average, the estimated total fixation time on item stem and options was about 4.4 seconds longer than in the control setting when a task-irrelevant DS was included in the graph. Applying the delta method we found that the estimated total fixation time in the DS condition was also significantly higher than in the DP condition (Estimate = 3.06, SE = 0.76, $t = 4.04$, $p < .01$).

3.6.4.2 Axes and Labels

The estimated total fixation time on axes and labels was, on average, about 17 seconds in the control condition. The estimated fixation time was significantly higher in both the data points condition ($t = 5.28$, $p < .001$) and in data series condition ($t = 2.95$, $p < .05$). On average, the estimated total fixation time on axes and labels was 3.5 seconds longer when irrelevant data points were included into the graph and about 2 seconds longer when an additional data series was included into the graph. The delta method revealed that the difference between the DP and DS conditions was significant (Estimate = 1.52, SE = 0.64, $t = 2.35$, $p < .05$).

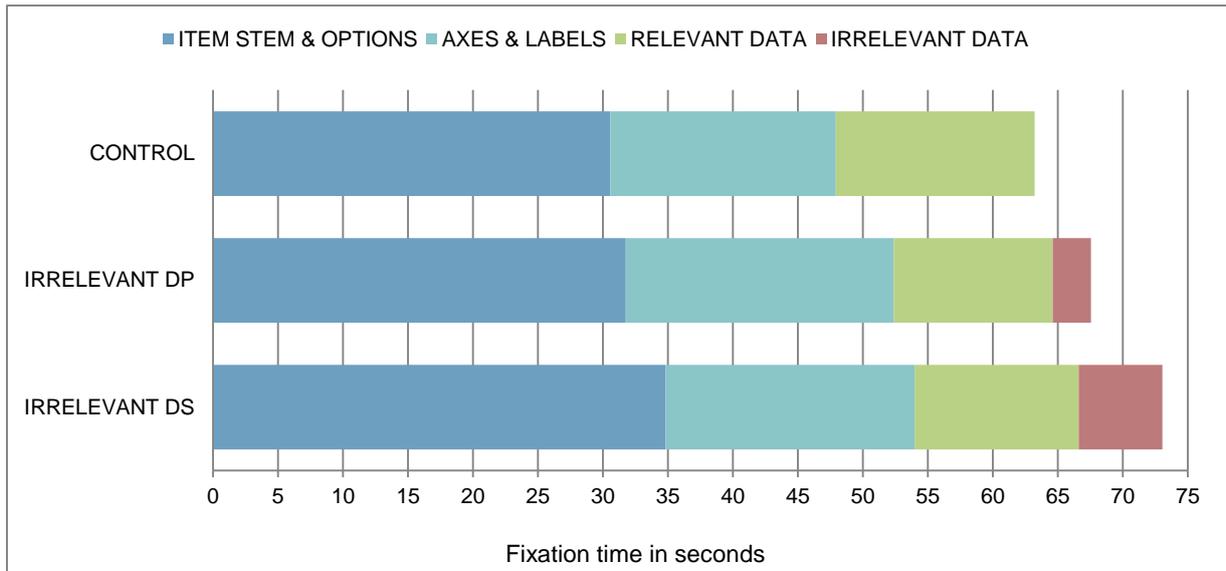


Figure 3. Average total fixation times on the four Areas of Interest groups across all participants, compared between the three experimental conditions. DP = Data Points, DS = Data Series.

3.6.4.3 Relevant Data

In the control condition, the estimate for the average total fixation time on relevant data points was about 15 seconds. The estimated total fixation time was significantly lower by about 3 seconds both in the DP condition ($t = -4.65, p < .01$) and in the DS condition ($t = -4.26, p < .01$). The difference between the DP and DS conditions was not significant (Estimate = 0.24, SE = 0.62, $t = 0.40, p = .70$).

3.6.4.4 Irrelevant Data

There were no irrelevant data points in the control condition, thus the total fixation time was zero by definition. However, estimated total fixation times on irrelevant data were significantly different from the control condition by about 3 seconds in the DP condition ($t = 8.88, p < .001$) and by about 6 seconds in the DS condition ($t = 19.25, p < .001$). The estimated total fixation time was also significantly higher in the DS condition than in the DP condition (Estimate = 3.46, SE = 0.33, $t = 10.37, p < .001$).

3.7 Discussion

In this study, we investigated the relationship between the processing of graph reading tasks and the presentation of task-irrelevant data. Conducting an experiment, we varied the amount of task-irrelevant data in three qualities (i.e., task-irrelevant data points, a task-irrelevant data series, and a control setting with no irrelevant data) and employed eye tracking

Kapitel 3 - Task-irrelevant Data Impair Processing of Graph Reading Tasks

measures to analyze attention distribution during task processing. We hypothesized that task-irrelevant data would negatively impact error rates, processing time and cognitive load.

In line with our hypotheses, we found that task-irrelevant data significantly elevated error rates, processing time and cognitive load ratings. Results show that similar findings by Spence and Lewandowsky (1991), Meyer et al. (1997) as well as Kim and Lombardino (2015) also apply to settings where additional data is always irrelevant to the task and that these findings are applicable to bar graphs as well (cf. Kumar & Benbasat, 2004 for findings on line graphs). As suggested by Sedig and Parsons (2013), additional complexity places a large burden on the perceptual and cognitive systems, resulting in higher ratings of cognitive load. However, we expected that these effects would be smaller in the DS condition than in the DP condition because perceptual grouping may aid the graph reading process. Unexpectedly, task-irrelevant DS impaired task processing much stronger than task-irrelevant DP. Regarding this unsuspected finding, our analysis of attention distribution revealed differences between the experimental conditions that provide some explanation which we will elaborate on in the next paragraphs.

In the DP condition, the estimated error rate was elevated by 61% and the estimated processing time by 6 seconds. Additionally, the rating of cognitive load was significantly, but only slightly higher compared to the control condition. The analysis of attention distribution among AoIs revealed that time spent with irrelevant data goes at the expense of time spent with relevant data. Furthermore, a major part of additional processing time was spent with reading axes and labels (about 3.5 seconds). This is not surprising because additional data points come with associated labels that need to be understood in order to distinguish relevant from irrelevant data points.

In the DS condition however, estimated error rates were more than twice as big as in the control setting, the processing time was elevated by 13 seconds and the rating of cognitive load was significantly higher. Moreover, the change in processing time could not be explained solely by additional time devoted to axes and labels (2 seconds). As shown by the fixation times, a considerable amount of additional processing time was associated with reading the item stem and answer options. Participants were spending more time with these regions than in the other conditions, possibly to identify which of the two data series is required for the task or to double check if only one or both data series are of interest. This finding indicates that, even though these regions did not change, the task demands did. Prior to comparing the data, graph readers had to identify which data series was required by the task, adding an additional step to the task completion process. In other words, adding irrelevant data points

seems to be a quantitative change while adding an irrelevant data series seems to be a qualitative change.

3.7.1 Implications for Instruction

The results of the present study demonstrate that different qualities of task-irrelevant data affect processing of graph reading tasks differently. For instructional design, it is important to identify critical scenarios and to adapt accordingly. In the following we provide some suggestions for instruction and the design of graphs as part of instructional material.

First, we found that task-irrelevant DP negatively impacted performance in the graph task as well as the rating of cognitive load, but to a relatively small degree. Results suggest that readers need more time to read the additional labels and distinguish relevant from irrelevant data. However, the time spent with the item stem and options was not significantly different from the control setting, indicating the task demands did not change fundamentally. The estimated rating of cognitive load differed, although significantly, just slightly from the control setting (by about 0.15 points on a 5-point Likert scale). Taken together, task-irrelevant DP mainly seem to add processing time for comprehension of additional labels, but show only small effects on other characteristics of the task. In this case, the benefits of excluding task-irrelevant data points from instructional material may not outweigh the costs of altering existing graphs or designing new material.

Second, we found that task-irrelevant DS qualitatively changed the task demands, induced higher cognitive load and impaired performance significantly. Considering typical teaching-learning settings, the provision of task-irrelevant data can be conceptualized as a conversation between teacher and student. By providing additional information in the graph that is unrelated to the task a teacher would be violating what is considered a conversation principle (i.e., Do not make your contribution more informative than required; Grice, 1975, 1978). For example, if students are to evaluate the economic growth of a specific country, a graph displaying multiple countries might be confusing. Grice (1975) argues that “such overinformativeness may be confusing in that it is liable to raise side issues; and there may also be an indirect effect, in that the hearers may be misled as a result of thinking that there is some particular POINT in the provision of the excess of information.” (p. 46). In the case of our graph task, readers may have been confused by the extra information in the graph, leading them to question and check if they missed something in the task description. Based on our results, we would recommend that teachers avoid the use of graphs that display multiple data series or variables when only one is relevant for the learning topic.

3.7.2 Implications for Graph Tasks in Tests

While the present findings are also applicable to the design of graph tasks in tests, design implications may differ to some degree. In the context of testing, the amount of task-irrelevant data can be understood as part of the item characteristics. Because the presence of irrelevant data, especially task-irrelevant DS, may affect test results, test designers should pay attention to whether the ability to distinguish relevant from irrelevant data is part of their understanding of the construct they aim to assess. For example, if a test is used to measure graph reading skills, distinguishing relevant from irrelevant data is likely to be included in the construct. However, if a test aims to assess reasoning based on a given data set, difficulties resulting from task-irrelevant data may be understood as construct-irrelevant variance (i.e., "aspects of the task that are extraneous to the focal construct make the test irrelevantly more difficult"; Messick, 1989, p. 34). We suggest that a decision to include task-irrelevant data in test items or not should be based on the understanding of the construct in question.

3.7.3 Limitations & Future Research

A number of limitations should be acknowledged for the present research. The scope of our paper is limited to a specific task (comparison of multiple data points) and a specific display type (bar graphs). Even though point comparison is a common graph reading task, there are a number of equally common tasks (e.g., point location and trend identification; see Bertin, 1983; Schnotz, 1994; Wainer, 1992) and well-established graph types (e.g., line graphs and pie charts; for overviews see Kosslyn, 1989; Lohse, Biolsi, Walker, & Rueter, 1994). Graph comprehension research has shown that performance in graph reading tasks is determined by an interaction between the task and the display type (e.g., Pinker, 1990; Shah & Hoeffner, 2002). For example, trend tasks may be affected differently by task-irrelevant data, because trends are evaluated more globally, not on a point-by-point level. Future studies should expand the research to additional tasks and more graph types in order to paint a more comprehensive picture.

The extent of task-irrelevant data was varied using a fixed number of additional data points and one additional data series. While we were able to find substantial effects using this manipulation, it remains unclear how these effects scale with smaller or larger numbers. For example, Spence and Lewandowsky (1991) found that performance with five or seven data points was about the same, but differed from graphs with only 4 data points. Future studies

should further vary the extent of task-irrelevant data to investigate the critical amount of irrelevant data in instructional material.

Finally, it is important to note that we examined university students that reported a consistently high level of graph literacy in this study. While advanced graph reading skills may compensate for high cognitive demands induced by task-irrelevant data, beginners may be more vulnerable to the detrimental effects we observed in the present study. This may be especially true for younger children because they are also more susceptible to misleading information (Newcombe & Dour, 2001). We suggest that future studies compare learners of different ages and different levels of graph literacy to help identify vulnerable groups.

3.8 Conclusion

In addition to supporting earlier findings that performance may be impacted by data complexity, the present study expands the graph comprehension research by identifying detrimental effects on both performance and task processing from data that is completely irrelevant to the given task. These new findings have important implications for the design of graph tasks both for instruction and for testing. Moreover, eye tracking measures yielded a first impression of the underlying mechanisms of the observed effects. While task-irrelevant data points had relatively small effects on cognitive performance, the presence of a task-irrelevant data series qualitatively elevated the task demands, resulting in higher error rates, longer processing times and a higher rating of cognitive load. For instructional material, the benefits of altering graphs with irrelevant data points may not outweigh the costs. However, instructors should consider altering graphs to display only task-relevant data series, especially when the task is designed for beginners.

3.9 References

- Ayres, P. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction, 16*(5), 389–400.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation, 8*, 47–89. doi:10.1016/s0079-7421(08)60452-1
- Baker, R. S., Corbett, A. T., & Koedinger, K. R. (2001). Toward a model of learning data representations. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society* (pp. 45–50). Mahwah, NJ: Erlbaum.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. doi:10.18637/jss.v067.i01
- Bertin, J. (1983). *Semiology of graphics: Diagrams, networks, maps*. Madison, WI: University of Wisconsin Press.
- Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction, 20*(2), 155–166.
- Casali, J. G., & Gaylin, K. B. (1988). Selected graph design variables in four interpretation tasks: A microcomputerbased pilot study. *Behaviour & Information Technology, 7*, 31–49.
- Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior, 25*(2), 315–324. doi:10.1016/j.chb.2008.12.020
- Cucuo, A. A., & Curcio, F. (2001). *The role of representation in school mathematics: National Council of Teachers of Mathematics Yearbook*. Reston, VA: National Council of Teachers of Mathematics.
- De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science, 38*, 105–134. doi:10.1007/s11251-009-9110-0
- Demetriadis, S., & Cadoz, C. (2005). A conceptual framework for the integrative design of adaptable representations for learning. In A. Méndez-Vilas, B. González-Pereira, J. Mesa González, & J. A. Mesa González (Eds.), *Recent research developments in learning technologies* (pp. 1250–1255). Badajoz, Spain: Formatex.

- Dodeen, H. (2008). Assessing test-taking strategies of university students: Developing a scale and estimating its psychometric indices. *Assessment & Evaluation in Higher Education*, 33, 409–419.
- Dorman, J. P. (2008). The effect of clustering on statistical tests: An illustration using classroom environment data. *Educational Psychology*, 28, 583–595. doi:10.1080/01443410801954201
- Duchowski, A. (2007). *Eye tracking methodology: Theory and practice*. London, UK: Springer.
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87(3), 215–251.
- Freedman, E. G., & Shah, P. (2002). Toward a model of knowledge-based graph comprehension. In M. Hegarty, B. Meyer, & N. H. Narayanan (Eds.), *Diagrammatic representation and inference* (pp. 8–31). Berlin, Germany: Springer.
- Glazer, N. (2011). Challenges with graph interpretation: A review of the literature. *Studies in Science Education*, 47, 183–210. doi:10.1080/03057267.2011.605307
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York, NY: Academic Press.
- Grice, H. P. (1978). Some further notes on logic and conversation. In P. Cole (Ed.), *Syntax and semantics 9: Pragmatics* (pp. 113–128). New York, NY: Academic Press.
- Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods*, 39, 101–117. doi:10.3758/BF03192848
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. UK: Oxford University Press.
- Huang, W., Eades, P., & Hong, S. H. (2009). Measuring effectiveness of graph visualizations: A cognitive load perspective. *Information Visualization*, 8(3), 139–152.
- Hyönä, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction*, 20, 172–176. doi:10.1016/j.learninstruc.2009.02.013

Kapitel 3 - Task-irrelevant Data Impair Processing of Graph Reading Tasks

- Jackson, C. H. (2011). Multi-State Models for Panel Data: The msm package for R. *Journal of Statistical Software*, 38(8), 1–29.
- Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the eyes of the beholder: how experts and novices interpret dynamic stimuli. *Learning and Instruction*, 20(2), 146–154.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329–354. doi:10.1037/0033-295X.87.4.329
- Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 40(1), 1–17.
- Kim, S. & Lombardino, L. J. (2015). Comparing graphs and text: Effects of complexity and task. *Journal of Eye Movement Research*, 8(3), 1–17.
- Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, 3, 185–225. doi:10.1002/acp.2350030302
- Kumar, N., & Benbasat, I. (2004). The effect of relationship encoding, task type and complexity on information representation: An empirical investigation of 2D and 3D graphs. *MIS Quarterly*, 28, 255–281.
- Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11, 65–100. doi:10.1111/j.1551-6708.1987.tb00863.x
- Lindner, M. A., Eitel, A., Strobel, B., & Köller, O. (2017). Identifying processes underlying the multimedia effect in testing: An eye-movement analysis. *Learning and Instruction*, 47, 91–102. doi:10.1016/j.learninstruc.2016.10.007
- Lindner, M. A., Eitel, A., Thoma, G.-B., Dalehefte, I. M., Ihme, J. M., & Köller, O. (2014). Tracking the decision making process in multiple-choice assessment: Evidence from eye movements. *Applied Cognitive Psychology*, 28, 738–752. doi:10.1002/acp.3060
- Lohse, G. L., Biolsi, K., Walker, N., & Rueter, H. H. (1994). A classification of visual representations. *Communications of the ACM*, 37, 36–49. doi:10.1145/198366.198376
- Manor, O., & Zucker, D. M. (2004). Small sample inference for the fixed effects in the mixed linear model. *Computational Statistics & Data Analysis*, 46(4), 801–817.
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511811678

Kapitel 3 - Task-irrelevant Data Impair Processing of Graph Reading Tasks

- Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction*, 20(2), 167–171. doi:10.1016/j.learninstruc.2009.02.012
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (p. 13–103). New York, NY: Macmillan.
- Meyer, J., Shinar, D., & Leiser, D. (1997). Multiple factors that determine performance with tables and graphs. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 39(2), 268–286.
- Newcombe, P. A., & Dour, T. M. (2001). Conversational influences on young children's responses to misleading questions. *Journal of applied developmental psychology*, 22(4), 363–378.
- Paas, F., & van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem solving skills: A cognitive-load approach. *Journal of Educational Psychology*, 86, 122–133.
- Paas, F., van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, 79, 419–430.
- Peebles, D., & Cheng, P. C. H. (2003). Modeling the effect of task and graphical representation on response latency in a graph reading task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 45, 28–46. doi: 10.1518/hfes.45.1.28.27225
- Pereira-Mendoza, L., Goh, S. L., & Bay, W. (2004, November). Interpreting graphs from newspapers: Evidence of going beyond the data. Paper presented at the ERAS Conference, Singapore. Retrieved from https://repository.nie.edu.sg/bitstream/10497/15549/1/ERAS-2004-277_a.pdf
- Pinker, S. (1990). A theory of graph comprehension. In R. O. Freedle (Ed.), *Artificial intelligence and the future of testing* (p. 73–126). Hillsdale, NJ, England: Lawrence Erlbaum Associates.
- Purchase, H. C. (2014). Twelve years of diagrams research. *Journal of Visual Languages & Computing*, 25, 57–75. doi:10.1016/j.jvlc.2013.11.004

- Quené, H., & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*, 413–425. doi:10.1016/j.jml.2008.02.002
- R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Schnotz, W. (1994). Wissenserwerb mit logischen Bildern [Knowledge acquisition with logical pictures]. In B. Weidenmann (Ed.), *Wissenserwerb mit Bildern* (pp. 95–148). Bern: Huber.
- Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review*, *19*(4), 469–508. doi:10.1007/s10648-007-9053-4
- Schutz, H. G. (1961a). An evaluation of formats for graphic trend displays—experiment II. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *3*(2), 99–107.
- Schutz, H. G. (1961b). An evaluation of methods for presentation of graphic multiple trends—experiment III. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *3*(2), 108–119.
- Sedig, K., & Parsons, P. (2013). Interaction design for complex cognitive activities with visual representations: A pattern-based approach. *AIS Transactions on Human-Computer Interaction*, *5*(2), 84–133.
- Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review*, *14*, 47–69. doi:10.1023/A:1013180410169
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Los Angeles, CA: SAGE.
- Spence, I., & Lewandowsky, S. (1991). Displaying proportions and percentages. *Applied Cognitive Psychology*, *5*, 61–77.
- Strobel, B., Saß, S., Lindner, M. A., & Köller, O. (2016). Do graph readers prefer the graph type most suited to a given task? Insights from eye tracking. *Journal of Eye Movement Research*, *9*(4), 1–15. doi:10.16910/jemr.9.4.4
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*, 257–285. doi:10.1207/s15516709cog1202_4

Kapitel 3 - Task-irrelevant Data Impair Processing of Graph Reading Tasks

- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, 4(4), 295–312.
- Tory, M., & Möller, T. (2004). Human factors in visualization research. *IEEE transactions on visualization and computer graphics*, 10(1), 72–84.
- Van Gog, T., Kester, L., Nievelstein, F., Giesbers, B., & Paas, F. (2009). Uncovering cognitive processes: Different techniques that can contribute to cognitive load research and instruction. *Computers in Human Behavior*, 25, 325–331. doi:10.1016/j.chb.2008.12.021
- Wainer, H. (1992). Understanding graphs and tables. *Educational Researcher*, 21, 14–23. doi:10.3102/0013189X021001014
- Wertheimer, M. (1938). Laws of organization in perceptual forms. In W. D. Ellis (Ed.), *A source book of Gestalt psychology*. London: Routledge & Kegan Paul
- Wright, R. D., & Ward, L. M. (2008). *Orienting of attention*. New York, NY: Oxford University Press.

Kapitel 4

Studie 3

Do Seductive Details do Their Damage in the Context of Graph Comprehension? Insights from Eye Movements

Strobel, B., Grund, S., & Lindner, M. A. (2018). Do Seductive Details do Their Damage in the Context of Graph Comprehension? Insights from Eye Movements. *Applied Cognitive Psychology*. Advance online publication. doi:10.1002/acp.3491

Abstract

Computer based learning and testing provide accessible ways to enrich educational material with additional multimedia content. In educational research, interesting but irrelevant materials are often considered seductive details. Seductive details can increase students' task interest but they are also suspected to have detrimental effects on learning and retention. While seductive details have been mostly examined in the context of learning and text comprehension, such elements are also used in graphs (e.g., for depicting data points or providing additional information). In the present experiment, we investigated both seductive text and seductive pictures in the context of graph comprehension compared to a control condition without additional information, using a within-subject design. We recorded $N = 68$ students' eye movements while they processed several bar graphs with and without seductive details. The data were analyzed with linear mixed-effects models. Our results show that the seductive details did not affect students' graph reading performance, but they prolonged the task processing time. The eye-fixation measures revealed that additional processing time was best explained by attention distraction towards the seductive material rather than by disruption of the process. Implications for both graph comprehension and seductive details research are discussed.

Keywords: Seductive details effect; graph comprehension; distraction and disruption; eye tracking; linear mixed-effects models

Do Seductive Details do Their Damage in the Context of Graph Comprehension? Insights from Eye Movements

Computer based learning and testing make it easy to enrich educational material with additional multimedia content. To describe interesting but irrelevant information used to spice up educational material, Garner et al. coined the term *seductive details* (Garner, Brown, Sanders, & Menke, 1992; Garner, Gillingham, & White, 1989). Harp and Mayer (1997, 1998) further distinguished between *seductive text* to describe interesting but irrelevant text and *seductive illustrations* to describe interesting but irrelevant pictures.

While there is evidence that additional, but irrelevant material may increase student interest (cf. Harp & Mayer, 1997; Lenzner, Schnotz, & Müller, 2013), seductive details are also known for their detrimental effects on retention and transfer in learning (e.g., Harp & Mayer, 1997; Lehman, Schraw, McCrudden, & Hartley, 2007; Rowland-Bryant et al., 2009). Typically, researchers have examined the *seductive details effect* in the context of text comprehension (e.g., Mayer, Griffith, Jurkowitz, & Rothman, 2008; Rey, 2011) while research on graph comprehension was mainly concerned with visual features that are part of the graphical system (e.g., legends, data complexity, size, position and more; Cleveland & McGill, 1987; Kim & Lombardino, 2015; Kumar & Benbasat, 2004; Shah & Hoeffner, 2002). However, digital media such as presentations, websites and computer based learning material provide accessible ways to supplement graphs in the same ways as text (e.g., depicting the content of data points or providing additional information). There are reasons to assume that seductive details may affect graph comprehension as well. By design, graphs are highly conventionalized visual systems in which spatial relations between visual objects are employed as an analogy to non-spatial relations (Schnotz, 2002; Winn, 1990). Thus, extraneous material may be distracting and disruptive for the process of graph reading that relies on the structural conventions and spatial analogies employed in graphs.

Theoretical explanations for the seductive details effect are linked to the fact that working memory capacity (WMC) is limited (Baddeley & Hitch, 1974) and can easily be overloaded by extraneous material (i.e., extraneous cognitive load, see e.g. Paas & Sweller, 2014). For example, Sanchez and Wiley (2006) found that individuals with low WMC were especially vulnerable to the seductive details effect as they were less able to direct their attention towards the relevant information. A number of underlying mechanisms have been

proposed (Harp & Mayer, 1998), but a meta-analysis by Rey (2012) showed overall mixed results as to how the seductive details effect is caused.

In the present experiment, we integrated two types of seductive material into various graph comprehension tasks—seductive text and seductive illustrations—and investigated how these additional materials would affect students' error rate and processing time, while accounting for individual differences in WMC. In addition, we aimed to gain deeper insights into the processes that take place during the solution of graph tasks with and without seductive materials. Therefore, we applied eye tracking methodology, which has proven useful in previous graph comprehension studies (e.g. Kim & Lombardino, 2015; Peebles & Cheng, 2003; Strobel, Saß, Lindner, & Köller, 2016).

4.1 The seductive details effect

In educational research, seductive details are described as interesting but irrelevant details related to a given topic but not necessary to achieve the instructional objective (Harp & Mayer, 1998; Mayer & Fiore, 2014). To introduce a classic example, Harp and Mayer (1997) used an educational text on lightning formation. In this paradigmatic approach, a story about a football player who got struck by lightning and corresponding illustrations were added to the informational text about the genesis of lightning. Even though the additional material has topical relevance (i.e., it is related to the topic of lightning), it has little conceptual relevance (i.e., it is not related to the instructional objective, learning the process of lightning formation; Mayer, 2009). Results across multiple studies consistently revealed that adding seductive details to an informational passage had detrimental effects on retention of important information (e.g., Doolittle & Altstadter, 2009; Lehman et al., 2007). Typically, students remembered less of the main ideas of the text and more of the seductive details (for a meta-analysis on seductive details see Rey, 2012). In the context of graph comprehension, seductive details may interfere with the graph reading process as the extraneous elements and the structural important graph elements compete for a reader's limited cognitive capacity (see e.g., Mayer & Fiore, 2014; Paas & Sweller, 2014).

Harp and Mayer (1998) postulate three potential causes of the seductive details effect that adhere to cognitive load theory (Sweller, 1994; Paas & Sweller, 2014): distraction, disruption and diversion. According to the *distraction hypothesis*, seductive details divert a reader's attention away from the relevant information and toward the more interesting but irrelevant information. The *disruption hypothesis* suggests that seductive details interrupt the mental model construction (e.g., Schnotz, 2002) as readers proceed through the material,

resulting in incomplete or incoherent representations of the learning material. Finally, the *diversion hypothesis* attributes the detrimental effects of seductive details to the activation of inappropriate representations: While readers are still able to construct a coherent mental model, they are not building their representation around the important information in the text. Instead, seductive details encourage the activation of inappropriate prior knowledge, leading to a mental model construction around the irrelevant information.

While Harp and Mayer (1998) found some evidence in favor of the diversion hypothesis, more recent studies challenged this notion. Presenting seductive material *before* a text instead of embedding it *within* the text has shown to mitigate the seductive details effect (e.g., Wiley, 2003; Wright, Milroy, & Lickorish, 1999). If poorer comprehension due to seductive material was caused by diversion, the position of the seductive material should have no effect on students' learning performance. These findings suggest that interference is somehow involved in the mechanism of the seductive details effect, highlighting the role of distraction and disruption. Interference via distraction shifts a reader's attention back and forth from relevant to irrelevant (i.e., seductive) information; interference via disruption is supposed to manifest in an interruption of the mental model construction process.

Sanchez and Wiley (2006) argue that both distraction and disruption are related to low WMC. In their study on the relationship between seductive details and working memory they found that WMC affected learning outcomes only when seductive details were present in the material. More precisely, the effect was most evident in the group of students with low WMC, showing that these readers were especially vulnerable to the seductive details effect. The authors conclude that individuals with lower WMC were less able to deal with distracting information and irrelevant material, and consequently focused less on the relevant conceptual information.

4.2 Seductive details in the context of graph comprehension

Traditionally, researchers have examined the seductive details effect in the context of learning, often associated with text comprehension and the recall of important information (e.g., Garner et al., 1989; Lehman et al., 2007). With this study, we apply the research on learning with seductive details to graph reading tasks and immediate performance. This implies tackling two challenges:

First, we expand the traditional research on seductive details to the area of graph comprehension. Research on graph comprehension is often concerned with computational differences between various graph types (e.g., Peebles & Cheng, 2003; Pinker, 1990; Shah &

Hoeffner, 2002; Shah, Mayer, & Hegarty, 1999), between graphs and other representations (e.g., Wainer, 1992), as well as visual features that affect graph comprehension and learning with graphs (e.g., color, position, size, legends, data complexity and more; Cleveland & McGill, 1987; Kim & Lombardino, 2015; Kumar & Benbasat, 2004; Shah & Hoeffner, 2002). While there has been no research on irrelevant, but interesting details in graphs, there are findings on irrelevant data. For example, Kim and Lombardino (2015) conducted an eye tracking study and varied the number of data series presented in bar graphs. They found that processing time was significantly higher when an additional data series was present. This was also reflected in longer eye-fixation times on the graph regions. However, graphs are highly conventionalized visual systems (i.e., they consist of a more or less fixed number of visual features, such as axes, labels and data points). Visually inspecting these spatial systems is crucial to successfully extracting information from a graph. In contrast to irrelevant data points, seductive details can be considered extraneous material that may interfere with the process of graph reading through distraction or disruption (see Section 4.1). In graphs, seductive material may not only be distracting in regard to its content but also in regard to spatial aspects (i.e., seductive details add irrelevant visual elements to an otherwise highly structured spatial system).

Secondly, we apply the principles of the seductive details effect from learning and retention to the area of graph reading tasks that require students' immediate performance. Nevertheless, mechanisms identified in the context of learning should be applicable to immediate performance at least to some extent because processing and understanding the material have similar fundamental requirements to those necessary for learning, namely encoding the given information and constructing a coherent mental model in order to solve the task (see also Lindner, Ihme, Saß, & Köller, 2016). In their study on the effect of representational pictures, Lindner et al. (2016) showed that the multimedia effect on learning (see Mayer, 2009; Mayer & Fiore, 2014) also occurred during problem solving in a multiple choice test and affected students' test performance. Seductive details may likewise affect the processing of graphs because the graph reading process relies heavily on the visual processing of different elements in the graph, making it vulnerable to effects of distraction and disruption. In the presence of seductive details, we expect the graph reading process to become more difficult and error-prone, resulting in higher error rates and longer processing times.

4.3 Using eye tracking to gain insight into the graph reading process

Using eye tracking as means to gain insight into cognitive processes has a long tradition that revolves around an idea called the *eye-mind hypothesis* (Just & Carpenter, 1980; for introductions on the method of eye tracking, see e.g. Duchowski, 2007; Holmqvist et al., 2011). Basically, the eye-mind hypothesis embraces the assumption that the location of eye-fixation also represents the focus of attention (i.e., it is assumed that we process the visual information we are currently looking at). While researchers have discussed some shortcomings of the method (e.g., Hyönä, 2010; Wright & Ward, 2008), a great number of empirical studies have shown that eye-fixation measures and cognitive performance are closely related (e.g. Canham & Hegarty, 2010; Jarodzka, Scheiter, Gerjets, & van Gog, 2010; Lindner, Eitel, Strobel & Köller, 2017). In contrast to other process tracing methods (e.g., verbal protocols; Ericsson & Simon, 1980; van Gog, Kester, Nievelstein, Giesbers, & Paas, 2009), remote eye tracking does not place additional load on participants' working memory and is unobtrusive for the participant (e.g. Hyönä, 2010).

Eye tracking is especially useful for the study of attentional processes that occur during processing of multimedia and learning materials because it provides detailed insight into the allocation of visual attention (Mayer, 2010; van Gog & Scheiter, 2010). In the context of graph reading, eye tracking allows allocating the processing time to the current spatiotemporal attention of the graph reader. In a number of recent eye tracking graph studies, researchers were able to successfully attribute processing time to important subregions of graph tasks (i.e., x-axis, y-axis, legend, pattern, question, answers; Kim & Lombardino, 2015; Peebles & Cheng, 2003; Strobel et al., 2016). In the present study, eye tracking enables us to track how much time can be attributed to the processing of specific regions in the graph, such as seductive details or specific graph regions. Furthermore, eye fixation patterns allow us to distinguish between potential explanations of the seductive details effect (cf. Harp & Mayer, 1998): If seductive details affect graph readers via distraction, this should be reflected in both fixations on the seductive details and longer total processing times. However, if seductive details affect graph processing via disruption (i.e., the transition from one element to the next is interrupted), readers need to revisit regions of the graph they inspected before the interruption occurred. Thus, longer fixation times on other graph regions (e.g., axes and labels) would be expected in addition to eye fixations on seductive details.

4.4 Research hypotheses

In this study we examined the effects of two types of seductive details placed in regular bar graphs in the context of graph comprehension, seductive text and seductive illustrations. We also focused on the role of students' working memory capacity in processing graph reading tasks with the different variations of seductive details compared to regular graph tasks. We applied eye tracking methodology to investigate two potential explanations for the seductive details effect: distraction and disruption. Based on the theoretical and empirical framework, we formulated the following hypotheses:

- (1) *Seductive Details Hypothesis*: Seductive details (text or pictures) in graphs are expected to distract students' attention from the functional elements of the graph or disrupt the graph reading process. Accordingly, we expected that graph readers make more mistakes and take more time to complete a task when they interpret graphs with seductive details compared to conventional graphs.
- (2) *Distraction Hypothesis*: If the seductive material affects graph reading via distraction from the relevant information, students should require additional processing time as reflected in eye-fixations on the seductive details. Additionally, we expected that fixation times on other regions will be unaffected by seductive details in a graph.
- (3) *Disruption Hypothesis*: If the seductive material affects graph reading via disruption, students should require additional processing time, reflected in longer eye-fixations on both the seductive details *and* other graph regions (e.g., axes, data points). This is because the mental model construction would take additional time when students are interrupted in the process of building their mental model.
- (4) *Vulnerability Hypothesis*: When solving graph reading tasks, students need to hold multiple variables in working memory, making the process especially difficult for individuals with a low WMC. As the processing of seductive material may induce additional cognitive load, students with a lower WMC will be more severely affected by seductive details than students with a higher WMC.

4.5 Method

4.5.1 Sample and study design

In our experiment, we examined $N = 68$ university students from different faculties (81% female, $M_{\text{age}} = 22.81$ years). All students had normal or corrected to normal vision. Originally, 73 students were approached but data from 5 students (7 %) were excluded from the analysis due to poor quality of the eye tracking data. The assessment comprised a set of graph task items as well as a paper-pencil questionnaire to assess demographic information. We employed a repeated measures within-subject design in which we varied the presence of seductive details in three conditions: no seductive details vs. seductive pictures vs. seductive text. The students completed five items in each experimental condition, resulting in a total of 15 graph task trials.

We assessed total processing time and error rate as dependent variables while controlling for working memory capacity. Additionally, we collected eye tracking data to determine the amount of time devoted to different functional areas of the task material. Finally, we adapted a short graph literacy scale to measure the students' level of experience in graph reading (see Strobel et al., 2016).

4.5.2 Material and measures

4.5.2.1 Graph task

For the computer-based graph task, we exclusively used bar graphs to avoid effects of the graph type on task performance (see e.g. Peebles & Cheng, 2003; Strobel et al., 2016). Each of the 15 graph task items consisted of a lead-in statement, followed by a single bar graph, the question, and four multiple-choice answer options (Figure 1).

To complete an item, students had to compare multiple data points in the graph, find an optimal solution in regard to the question (i.e., data points that fit a given criterion), and select the correct answer among four options in a classical multiple-choice format (single choice). The position of the one correct answer was determined by a one-time randomization (i.e., one randomization that was used for all students) to avoid that systematic positioning of the correct answers could help *test-wise* students to identify the correct answers without effortful completion of the actual tasks (for research on test-wiseness see e.g. Dodeen, 2008). For each student, the order of the 15 items was randomized. Regarding the item contents, we drew a sample from a variety of different topics (e.g., food, sports), but no specific prior knowledge was required to solve the tasks.

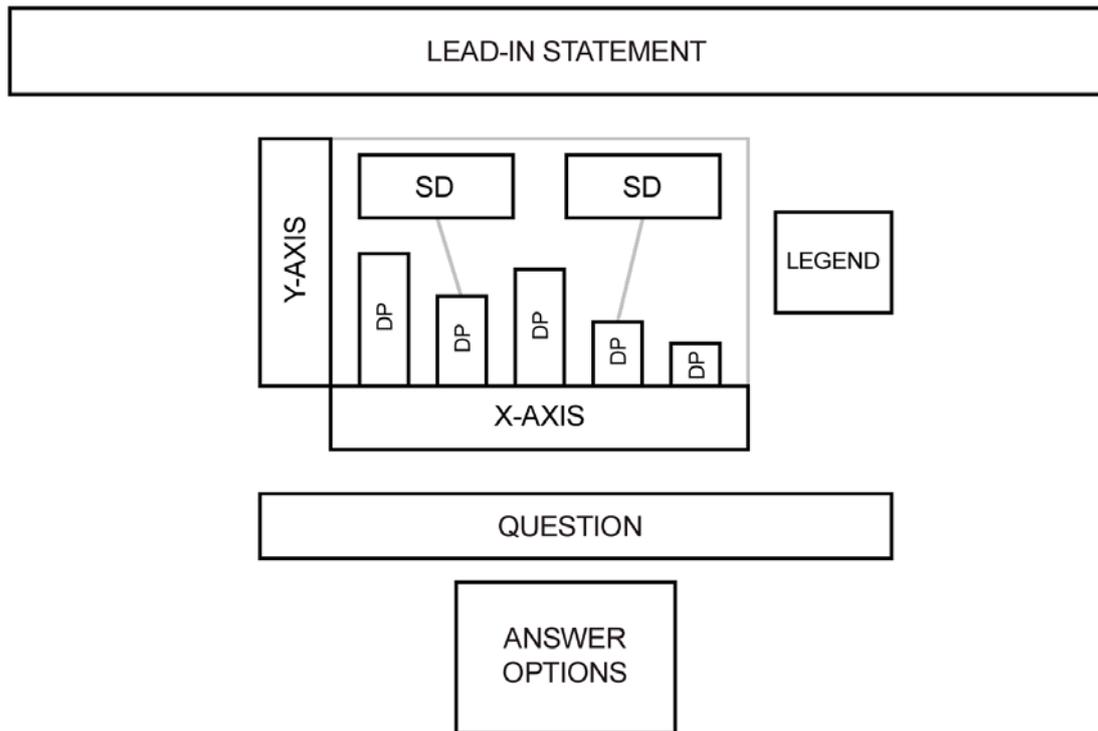


Figure 1. Schematic Representation of the Areas of Interest (AoI) defined for the eye tracking analysis in the experimental conditions. In addition to the displayed AoIs, the experimental conditions include AoIs for irrelevant data. DP = data point, SD = seductive detail.

All bar graphs were constructed in the same manner: The bar graphs consisted of a categorical x-axis (e.g., food name) and a continuous y-axis that showed quantitative data (e.g., number of calories). All graphs showed six data points, each of which had to be considered in order to identify the correct answer. As experimental variations we either included two pictures in the graph, visually illustrating two of the data points, or two short texts describing two of the data points (e.g., “If heated for a longer time period, raw eggs may explode in the microwave.” illustrating a data point for “eggs”). Both the pictures and texts had a rectangular shape of similar size and were included in the pattern area of the graph, directly above the bars. Seductive details were carefully positioned so that they would not occlude any data points or functional parts of the graph. Identical labels were used in all conditions. Lead-in statements, item stem, and multiple choice options did not differ between the experimental and control conditions.

4.5.2.2 Working memory capacity

We adapted a computerized spatial working memory capacity (SWMC) task described in Oberauer, Süß, Schulze, Wilhelm, and Wittmann (2000). The task is constructed “as a

spatial equivalent to reading span” (p. 1026). Every item consisted of a transformation instruction and a series of simple patterns that were presented sequentially on the computer screen. The instruction was either a 90° rotation to the left or to the right. Patterns were drawn on the screen by partially filling the cells of a 3 x 3 matrix. Students had to mentally rotate the pattern according to the instruction and remember the outcome. After a series of patterns, students had to fill a row of blank 3 x 3 matrices with the correct results in the same order they were presented. Each matrix was displayed for three seconds, followed by an inter stimulus interval of 200 ms. The task included two practice items, followed by eight items with two and six items with three patterns, respectively. One point was scored for every matrix that was completed correctly. Oberauer et al. (2000) argued that “the main requirement of this task should be simultaneous storage and transformation, although retention of spatial patterns can be assumed also to require coordination of the single cells into an integrated structure” (p. 1026).

4.5.2.3 Graph literacy

A short questionnaire was employed to assess the students’ graph literacy using 5-point Likert-scaled items. The scale comprised four items (i.e., “I am familiar with bar graphs”, “I have used bar graphs before”, “I often use graphs”, “I feel confident in reading graphs.”). The reliability of the graph literacy scale was satisfactory (Cronbach’s $\alpha = .74$).

4.5.3 Apparatus

Items of the graph task were presented on a 19-inch screen with a 1280 × 1024 pixel resolution, using the software Experiment Center 3.5 from SensoMotoric Instruments (SMI, Teltow, Germany). Each item appeared on a single screen. Students were seated in front of the computer screen at a distance of approximately 70 centimeters. The font size of the text was about 2 centimeters (approximately 1.6° visual angle). The students’ eye-movements were recorded using a video-based remote eye tracking system (SMI iView X™RED-m; 120Hz sampling rate) and the corresponding SMI software iView X™. The system was calibrated for each session using an animated 8-point calibration image and subsequent validation. The calibration accuracy was below 0.5° visual angle for all students on both the x and y coordinates (range: 0.04 to 0.49; $M_x = 0.29$, $SD_x = 0.09$; $M_y = 0.30$, $SD_y = 0.10$).

4.5.4 Procedure

Students were tested in single sessions. Prior to the experiment, they were familiarized with the procedure and the eye tracking system. Subsequently, the students completed the graph task on a computer while their eye-movements were recorded. The paper-pencil questionnaire was administered after completion of the graph task. The complete assessment cycle took about 45 minutes. Students later received ten euros for their voluntary participation in the study.

4.5.5 Analysis

4.5.5.1 Eye-movement data pre-processing

We analyzed the eye-movement recordings using a dispersion-based algorithm implemented in the Begaze™ software, version 3.5, from SMI. A fixation was detected when an eye movement lasted for at least 80 milliseconds on a position with a maximum dispersion of 100 pixels. In addition, we conducted an exhaustive visual inspection of the students' scan paths to determine the quality of the eye-movement data. We used total fixation time on predefined Areas of Interest (AoI) as the eye movement measure for our analyses. Total fixation time indicates how much time is spent fixating on each of the AoIs from trial onset to trial end (i.e., the total time devoted to a specific area; Holmqvist et al., 2011). For educational tasks using the multiple-choice item format, total fixation times provide a valid indirect measure of attention distribution and cognitive processing (e.g., Lindner et al., 2014).

In the tradition of recent graph studies (e.g., Kim & Lombardino, 2015; Peebles & Cheng, 2003, Strobel et al., 2016), we divided the graph regions into multiple rectangular AoIs. The x-axis and y-axis, including the respective labels, were grouped together as *axes & labels*. Each bar in the graph was also covered by its own AoI and all six bars were grouped together as *data*. Another six AoIs for the lead-in statement, the question and the four multiple-choice answer options were grouped together as *item stem & options*. All reported data refer to these AoI groups. For a visualization of the AoI configuration, see Figure 1.

4.5.5.2 Linear mixed-effects models

Data were analyzed using R, version 3.4.3 for Windows (R Core Team, 2017). Repeated measures were nested both within students and items, with each measure belonging uniquely to one student and one item. This data structure can be described as clustered or hierarchical with cross-classified effects for students and items (Snijders & Bosker, 2012). Regular ANOVA models can yield inflated Type I error rates when the data are clustered in

this manner (Dorman, 2008), so we applied *linear mixed-effects models* (LMMs) to account for the clustered structure of the data (for introductions see Snijders & Bosker, 2012; Quené & van den Bergh, 2008). This type of model can be understood “as a series of interrelated regression models that explain sources of variance at multiple levels of analysis, such as at the experimental stimuli and person levels” (Hoffman & Rovine, 2007, p. 102). LMMs can model fixed effects and random effects simultaneously: Fixed effects on the one hand aim to identify typical rates of change in the criterion variable that can be attributed to aspects of the experimental design (e.g., the manipulation). Random effects on the other hand aim to identify unsystematic rates of change (e.g., due to differences between items and students, respectively).

We used the R package lme4 (Bates, Maechler, Bolker, & Walker, 2015) to perform (generalized) linear mixed-effects analyses of the relationships between the experimental conditions and several outcome variables: error rate, total processing time, total fixation time on item stem and options, total fixation time on axes and labels, total fixation time on data, and total fixation time on seductive details in the graph. The LMMs include random intercepts for both students and stimuli. Models were fitted by the Restricted Maximum Likelihood (REML) criterion, because it yields better Type I error rates for smaller groups when testing fixed effects than estimates with the Maximum Likelihood criterion (Manor & Zucker, 2004; Snijders & Bosker, 2012). The control condition was chosen as the reference group in all LMMs. To calculate differences between the two experimental conditions (i.e., seductive text and seductive pictures) in the form of post-hoc contrasts, we made use of the “delta method” implemented in the R package car (Fox & Weisberg, 2011). Degrees of freedom for statistical tests of fixed effects in the LMMs were determined in accordance with the rule described in Snijders & Bosker (2012), that is, as $df = N - q - 1$, where N is the minimum number of upper-level units in the sample (here: $N = 15$ stimuli), and q is the number of explanatory variables included in the model.

4.5.5.3 Model comparisons and Bayes factors

For all our dependent variables, the model comparisons followed the same structure: Starting with an empty model including random intercepts for students and stimuli (Model 0), we first added SWMC as a covariate (Model 1), followed by a model that included the fixed effects for experimental conditions regarding the seductive details integration (Model 2) and finally we considered potential interaction effects of seductive details and SWMC (Model 3). For the first two dependent variables (error rate and total processing time) we report the full

models with all coefficients in Tables 1 and 2. For the remaining models that address the eye movement patterns, we only report model comparison parameters to facilitate inspection and save space (see Table 3).

We computed likelihood-ratio tests and multiple R^2 measures for the LMMs to assert if the inclusion of predictor variables constitutes a significant addition to the model. For the LMMs, we computed the two R^2 measures suggested by Snijders and Bosker (2012) and Nakagawa and Schielzeth (2013). For the generalized mixed effects models (GLMMs), we calculated the corresponding R^2 measure by Nakagawa and Schielzeth and the pseudo R^2 measure suggested by Nagelkerke (1991).

In addition to the more traditional indicators, we computed *Bayes factors* to aid with the comparisons of our LMMs using the R package BayesFactor (Morey & Rouder, 2015). Bayes factors allow for a simplified Bayesian approach to model comparisons by evaluating the posterior odds in favor of and against the models included in the comparison (Raftery, 1995). In the context of this study, Bayes factors were deemed useful because they provide a continuous measure of “evidence” in favor and against the null and alternative hypotheses on an intuitive scale (Kass & Raftery, 1995; Raftery, 1995). As a rough guide, the literature suggests that Bayes factors > 3 and $< .33$ may indicate notable support in favor of one of the two models involved in the comparison. For a more detailed introduction on Bayes factors, see Kass and Raftery (1995).

4.6 Results

Along with other biographic information, students were asked to rate their graph literacy on a scale comprised of four items, each of which used a five point rating scale. The mean rating of graph literacy in the current sample was $M = 3.39$ ($SD = 0.81$, range = 1.25 to 5) with a negative skewness (-0.38) indicating that the mass of the distribution was concentrated on the upper half of the scale (i.e., the ratings were relatively high).

4.6.1 Error rate

We used a series of generalized linear mixed-effects models (GLMMs) to estimate the relationship between the students’ error rate and the experimental conditions while controlling for working memory (Table 1). First, we computed an empty model with random intercepts for students and stimuli (M0). The estimates of the GLMMs are given in log odds (i.e., the logarithm of the odds, $p / (1 - p)$, where p represents a given probability). Translated into a frequency, the estimated error rate across all conditions was about 20 % (i.e., 80 % correct

answers). The main effect of SWMC was significant according to the likelihood-ratio test (Model 1; $\chi^2 = 5.39$, $df = 1$, $p = .020$). However, differences in the R^2 measures were small ($< .01$). As indicated by the model parameters, an increase of one standard deviation in SWMC resulted in an estimated decrease in error rate by approximately 4 percentage points. The subsequent addition of the experimental conditions did not yield significant changes in the likelihood-ratio, neither for main effects (M2) nor for the interaction effects (M3).

We hypothesized that the error rate would be higher in the seductive details conditions (i.e., when a picture or a fact is included in the graph) compared to the control setting with no additional content and that students with a low working memory capacity would be more vulnerable to the seductive details effect. Contrary to our expectations, error rates did not differ between experimental and control conditions and SWMC did not affect the presence of the seductive details effect.

4.6.2 Processing time

A series of LMMs was used to estimate the relationship between the students' processing time and the experimental conditions as well as working memory capacity (Table 2). We computed an empty model again with random intercepts for students and stimuli (M0). The estimates are given in seconds. The inclusion of SWMC was a significant improvement of the model (M1; $\chi^2 = 4.96$, $df = 1$, $p = .026$) as was the subsequent inclusion of condition main effects (M2; $\chi^2 = 8.14$, $df = 2$, $p = .017$). Again, differences in the R^2 measures were relatively small. Also, the inclusion of interaction effects did not improve the model significantly (M3; $\chi^2 = 5.94$, $df = 2$, $p = .051$).

We hypothesized that the processing time would be higher in the seductive details conditions compared to the control condition. Examining Model 2, the estimated processing time in the control condition was, on average, about 64.2 seconds. In this model, processing time was significantly longer by about 3.9 seconds in the seductive text condition ($p = .007$) and by about 3.3 seconds in the seductive pictures condition ($p = .002$). In addition, the assessment of post-hoc contrasts revealed that the conditions for seductive text and seductive pictures did not differ significantly from one another (Estimate = 0.58, $SE = 1.44$, $t = 0.40$, $p = .699$). These findings are in line with the seductive details hypothesis. We examined Model 3 with regards to the vulnerability hypothesis. Model comparison parameters unanimously suggest that the inclusion of interaction effects did not improve the model, indicating that SWMC did not affect the extent of the seductive details effect.

Kapitel 4 - Seductive Details in the Context of Graph Comprehension

Table 1

Comparison of fixed effects and random effects in the generalized linear mixed-effects models for error rate.

Fixed effect	Model 0 (M0)			Model 1 (M1)			Model 2 (M2)			Model 3 (M3)		
	Estimate	SE	z-value	Estimate	SE	z-value	Estimate	SE	z-value	Estimate	SE	z-value
Intercept	-1.36***	0.19	-7.05	-1.36***	0.19	-7.20	-1.47***	0.22	-6.60	-1.46***	0.22	-6.56
SWMC				-0.30*	0.13	-2.38	-0.30*	0.13	-2.37	-0.19	0.17	-1.08
Pictures							0.12	0.20	0.63	0.10	0.20	0.53
Text							0.20	0.20	1.03	0.20	0.20	0.97
SWMC x Pictures										-0.03	0.03	-1.06
SWMC x Facts										-0.02	0.03	-0.67
Random effect	Variance Component			Variance Component			Variance Component			Variance Component		
Student	0.74			0.63			0.63			0.63		
Stimulus	0.28			0.27			0.27			0.28		
Residual	-			-			-			-		
R^2_{NK}	0.00			0.01			0.01			0.01		
R^2_{NS}	0.00			0.02			0.02			0.02		
LR-Test ^a				$\chi^2(1) = 5.39^*$			$\chi^2(2) = 1.03$			$\chi^2(2) = 1.12$		

Note. SWMC = spatial working memory capacity (scaled). LR = likelihood ratio. R^2_{NK} = R^2 measure suggested by Nagelkerke (1991). R^2_{NS} = R^2 measure suggested by Nakagawa & Schielzeth (2013). ^a in comparison to the previous model.

* $p < .05$ *** $p < .001$

Table 2

Comparison of fixed effects and random effects in the linear mixed-effects models for total processing time.

Fixed effect	Model 0 (M0)			Model 1 (M1)			Model 2 (M2)			Model 3 (M3)		
	Estimate	SE	t-value	Estimate	SE	t-value	Estimate	SE	t-value	Estimate	SE	t-value
Intercept	66.60***	3.14	21.22	66.60***	3.08	21.62	64.21***	3.20	20.04	64.28***	3.20	20.09
SWMC				5.44*	2.42	2.24	5.48*	2.32	2.25	3.83	2.59	1.48
Pictures							3.28*	1.45	2.26	3.21*	1.45	2.21
Text							3.85*	1.45	2.66	3.82*	1.45	2.64
SWMC x Pictures										0.21	0.23	0.91
SWMC x Facts										0.57*	0.24	2.41
Random effect	Variance Component			Variance Component			Variance Component			Variance Component		
Student	395.40			371.40			372.80			375.64		
Stimulus	54.01			54.01			54.66			53.58		
Residual	342.01			342.01			339.56			338.02		
R^2_{SB}	0.00			0.03			0.03			0.03		
R^2_{NS}	0.00			0.04			0.04			0.04		
LR-Test ^a				$\chi^2(1) = 4.96^*$			$\chi^2(2) = 8.14^*$			$\chi^2(2) = 5.94$		
Bayes Factor ^a				2.19			0.59			0.19		

Note. SWMC = Spatial Working Memory Capacity (scaled). LR = likelihood ratio. R^2_{SB} = R^2 measure suggested by Snijders & Bosker (2012). R^2_{NS} = R^2 measure suggested by Nakagawa & Schielzeth (2013). ^a in comparison to the previous model. Bayes factors > 3 and < .33 are printed in bold type.

* $p < .05$ ** $p < .01$ *** $p < .001$

4.6.3 AoI analysis

Figure 2 shows a stacked bar chart of the average total fixation times, grouped by experimental condition and with blocks pertaining to the four AoI groups (item stem & options, axes & labels, data, and seductive details). Differences in fixation times on the AoI groups can help us distinguish between effects of distraction and disruption. We hypothesized that additional fixation times on the seductive details regions but not on other regions would support the distraction hypothesis, because this pattern would indicate that students attended to the seductive details, but the graph reading process was otherwise unaffected. In contrast, additional fixation times on both seductive details and the other regions would be in favor of the disruption hypothesis. This pattern would suggest that students needed to revisit previously inspected regions of the graph after the reading process was disrupted. Results of the model comparisons are provided in Table 3 (to facilitate inspection and save space, we only report model comparison parameters).

In the control condition, the average total fixation time on item stem and options was approximately 27 seconds, which was also the longest fixation time compared to the other AoI groups. The average total fixation time on axes and labels regions was approximately 18 seconds and 10 seconds on the data regions. Compared to the other conditions, there were virtually no differences in total fixation times between the experimental conditions except for a slight decrease in the time on the axes and labels region in the text condition by 1 second. Additional fixation times on the seductive details regions were approximately 3 seconds in the text condition and 1.5 seconds in the pictures condition.

For each of the AoI groups, we computed a series of LMMs for the relationship between total fixation times and the experimental conditions using the same procedures we employed for error rate and processing time. As the descriptive data already suggested, there were little to no differences in the total fixation times between the experimental and control conditions. Regarding the total fixation time on item stem and answer options, neither the inclusion of main nor the addition of interaction effects yielded a significant change in the likelihood ratio, which is mirrored by only small changes in the R^2 measures. The Bayes factors for Model 2 and Model 3 likewise indicated that the experimental conditions did not affect the total fixation times on the AoI regions for item stem and options. Model comparisons for total fixation times on axes and labels as well as the fixation times on data regions paint a similar picture with virtually no differences between experimental and control conditions as indicated by the negligible differences between models in terms of R^2 and Bayes factors consistently below .33.

Kapitel 4 - Seductive Details in the Context of Graph Comprehension

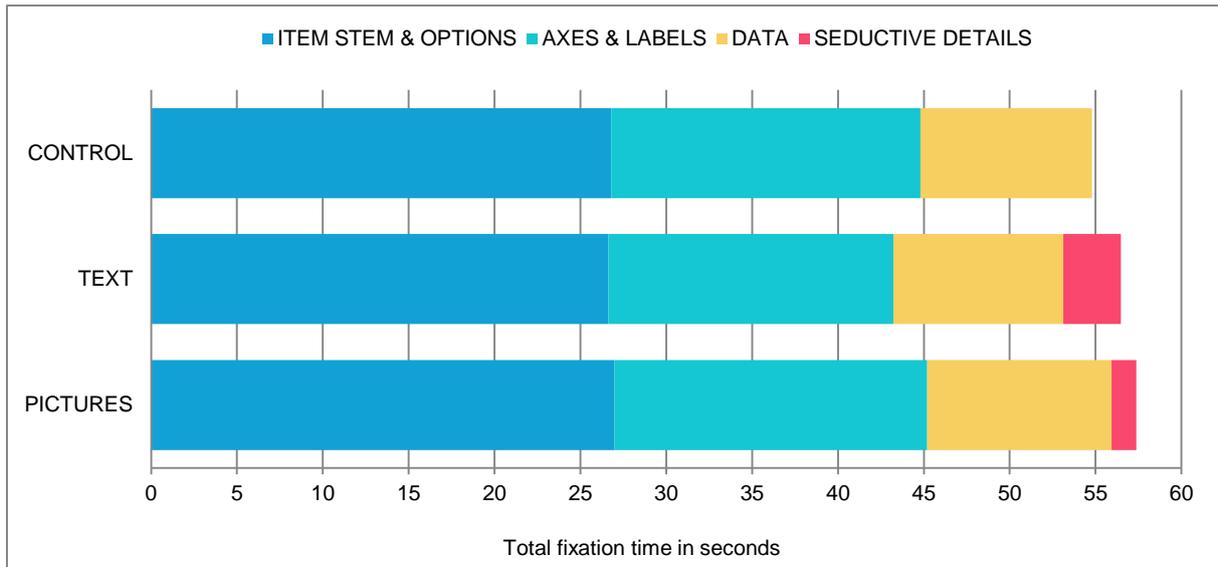


Figure 2. Average total fixation times on the four Areas of Interest (AoI) groups across all participants, grouped by condition and with blocks pertaining to AoI groups.

Table 3

Model comparisons of the linear mixed-effects models for total fixation times on the four AOI groups.

	Model 0 (M0)	Model 1 (M1)	Model 2 (M2)	Model 3 (M3)
ITEM STEM & OPTIONS				
R^2_{SB}	0.00	0.02	0.02	0.02
R^2_{NS}	0.00	0.03	0.03	0.03
LR-Test ^a		$\chi^2(1) = 3.71$	$\chi^2(2) = 1.14$	$\chi^2(2) = 4.43$
Bayes Factor ^a		1.22	0.02	0.10
AXES & LABELS				
R^2_{SB}	0.00	0.02	0.02	0.02
R^2_{NS}	0.00	0.03	0.03	0.03
LR-Test ^a		$\chi^2(1) = 3.89^*$	$\chi^2(2) = 7.01^*$	$\chi^2(2) = 5.75$
Bayes Factor ^a		1.27	0.36	0.19
DATA				
R^2_{SB}	0.00	0.01	0.01	0.01
R^2_{NS}	0.00	0.01	0.02	0.02
LR-Test ^a		$\chi^2(1) = 2.47$	$\chi^2(2) = 5.93$	$\chi^2(2) = 3.97$
Bayes Factor ^a		0.63	0.21	0.09
SEDUCTIVE DETAILS				
R^2_{SB}	0.00	0.00	0.28	0.28
R^2_{NS}	0.00	0.00	0.28	0.28
LR-Test ^a		$\chi^2(1) = 0.03$	$\chi^2(2) = 353.77^{***}$	$\chi^2(2) = 0.25$
Bayes Factor ^a		0.10	> 1.000	0.02

Note. M0: empty model with random intercepts for students und stimuli. M1: including main effects of spatial working memory capacity (SWMC). M2: including main effects of condition. M3: including interaction effects of SWMC and conditions. R^2_{SB} = R^2 measure suggested by Snijders & Bosker (2012). R^2_{NS} = R^2 measure suggested by Nakagawa & Schielzeth (2013). ^a comparison to the previous model. Bayes factors > 3 and < .33 are printed in bold type.

* $p < .05$ ** $p < .01$ *** $p < .001$

The analysis of total fixation times on the seductive details regions is the only exception to this pattern. However, this was to be expected because there were, by design, no seductive details present in the control condition. All indicators show evidence for main effects of condition on the total fixation time on the seductive details regions (M2). On average, the estimated total fixation time in the pictures conditions was approximately 1.4 seconds ($SE = 0.12$, $t = 8.73$, $p < .001$) and approximately 3.4 seconds in the text condition ($SE = 0.16$, $t = 20.76$, $p < .001$), reflecting the differences in total processing time. However, the addition of interaction effects with SWMC (M3) did not significantly improve the model. This finding indicates that students with different amounts of SWMC were not reacting differently to the presence of seductive details.

4.7 Discussion

In this study we examined the seductive details effect in the context of graph comprehension while accounting for individual differences in (spatial) working memory capacity. In addition, we employed eye tracking methodology to gain insight into the process of graph reading and how it is affected by our experimental manipulation of the graph tasks with either seductive text or seductive pictures.

Students completed multiple computerized graph reading items in a within-subject repeated measures design. The items either included thematically related but task-irrelevant pictures, texts, or no seductive details at all. Measures of error rate and processing time were examined to determine the extent of the seductive details effect in the context of graph comprehension. However, performance in the graph reading task was essentially unaffected by the inclusion of seductive details in the graph task items. As indicated by the fixation times on AoIs, the graph comprehension process did not differ in the experimental conditions with one notable exception: students exhibited additional processing time that can be attributed to fixations on the seductive details regions. Higher levels of working memory capacity on the other hand did significantly lower the error rates, but SWMC did not moderate the presence, absence, or size of the seductive details effect. In the following section, we will discuss the seductive details effect, the role of working memory capacity, and the scope of this study separately and in greater detail.

4.7.1 The seductive details effect in the context of graph comprehension

Results indicate that seductive details did not unfold all of their anticipated detrimental effects (cf. Harp & Mayer, 1997; Lehman et al., 2007; Rowland-Bryant et al., 2009) in the

context of graph comprehension. We hypothesized that the processing of graph reading tasks would take longer and would be more error-prone. By analyzing multiple LMMs we found that the error rate was unaffected by the inclusion of seductive details, while processing times were elevated by about 3 seconds in the seductive pictures condition and about 4 seconds in the seductive text condition. Model comparisons for the seductive details effect on eye-fixation times (see Table 3) revealed that processing of the graph task items was mostly unaffected by the experimental conditions, which was also evident by examining the descriptive statistics of total fixation times (see Figure 2). The only noticeable exception to this pattern was a few additional seconds of fixation time on the seductive details regions. This finding suggests that students (shortly) directed their attention to the seductive details, adding a few seconds of additional time needed to complete the task. Otherwise, task processing was mostly unaffected by the presence of seductive details.

We hypothesized that seductive details might be distracting and/or disruptive because diagrams are highly conventionalized graphical systems and thus might be vulnerable to the inclusion of external elements that are not part of the convention. Contrary to our expectations students had no difficulty solving the graph task items even when seductive details were present. In comparison, as another example, the addition of task-irrelevant data points in graphs can noticeably impair the graph reading process (Strobel, Lindner, Saß, & Köller, in press). This is probably because they are an integral part of the graph and initially indistinguishable from the important information. However, a recent study on seductive details suggests that readers pay less attention to seductive material when they are made aware that the details only show task-irrelevant information (Peshkam, Mensink, Putman, & Rapp, 2011). Accordingly, a possible explanation for our finding might be that students were immediately able to tell that the seductive material was irrelevant to the task. This idea is supported by two notions: First, we placed the seductive material in such a way that it did not include elements or paths that are essential to the graph reading process. Thus, it was possible to largely ignore the seductive material, given a person was fully aware of the task and the conventions of the graphical system. Second, the graph readers in our sample reported moderate to high graph literacy, indicating they were familiar with the bar graph format that we used in this study.

Taken together, the highly conventionalized nature of graphs (in conjunction with high graph literacy) may have had a shielding effect against negative influences of external elements such as seductive details. Thus, they may not pose a big threat for the graph comprehension process for individuals with at least moderate levels of graph literacy.

4.7.2 Distraction and disruption

In the seductive details literature, multiple causes of the seductive details effect are discussed (see e.g., Harp & Mayer, 1998; Sanchez & Wiley, 2006). We identified distraction and disruption as the most likely causes of the seductive details effect in the context of graph comprehension. Distraction happens when a reader's attention is seduced away from the relevant information and towards the more interesting but irrelevant material. Disruption describes the hypothesis that the detrimental effects of seductive details are the result of an interruption of the mental model construction process during comprehension.

In our study, eye-movement data were used to help us identify the most likely cause of the seductive details effect in the context of graph comprehension. If seductive details affect readers through distraction alone, additional eye-fixations would occur on the seductive details regions, but not on others. Instead, if the graph reading process is affected by disruption, additional eye-fixations would occur on other graph regions as well, because readers would need to start over where they left off (Sanchez & Wiley, 2006).

While seductive details showed only small effects on the graph comprehension process in general, there is tentative evidence for the distraction hypothesis over the disruption hypothesis. Our analyses of total fixation times on different graph regions revealed that additional processing time can be almost completely attributed to additional fixation time on the seductive details regions in the two experimental conditions. Model comparisons for the remaining AOI groups unanimously suggest there were no substantial differences in the fixation times on these regions. Taken together, the eye-fixation data in this study provide first evidence that supports the distraction hypothesis rather than the disruption hypothesis. While we found a short distraction in the range of a few seconds on average, there was no evidence for disruption as the fixation times on other regions were largely unaffected by both the experimental conditions, seductive text and seductive pictures.

4.7.3 The role of working memory capacity

Spatial working memory capacity was related to both error rate and total processing time. A high SWMC was associated with a smaller error rate, indicating that working memory capacity aids graph reading performance. Furthermore, a high SWMC was associated with longer processing time as well. These findings suggest that a higher level of SWMC enabled readers to inspect the graphs more thoroughly and make fewer mistakes.

In addition, our results indicated that working memory capacity did not affect the absence, presence or size of the seductive details effect. This was true for both error rates and

total processing time as well as for the fixation times on different regions of the graph. While Sanchez and Wiley (2006) found that individuals with a low WMC were especially vulnerable to the seductive details effect in the context of text comprehension, we did not find the same effect on the performance in the graph reading task. However, it is important to note that there were no main effects of seductive material on error rates in the first place. Also, the relatively high level of graph literacy in our sample may have caused variance restrictions, meaning that vulnerability to the seductive details effects might require a lower level of WMC than that observed in our sample. If future research is able to identify conditions under which seductive details in graphs affect performance in graph reading tasks, the role of working memory should be evaluated in greater detail once again.

4.7.4 Limitations & future research

While traditional research on the seductive details effect was focused on text comprehension on the one hand and effects on retention on the other hand (e.g., Lehman et al., 2007; Mayer et al., 2008; Rey, 2011; Rowland-Bryant et al., 2009), we evaluated the effects of seductive details on immediate performance in the context of graph comprehension. It can be argued that understanding and processing share the same fundamental requirements with learning and retention (i.e., encoding information and constructing a coherent mental model; Lindner et al., 2016), but they are not one and the same. For example, retention also requires the storage of information in long term memory and the ability to access it at a later point. Future studies of the seductive details effect in graphs could evaluate measures of retention alongside performance indicators. Furthermore, we placed seductive material directly in the graphs. Future studies could evaluate the effects of seductive details on graph comprehension in more traditional compositions (i.e., as part of the accompanying text) to paint a more comprehensive picture.

It is important to note that the present study embraces a specific graph reading task (comparison of multiple data points) and only one graph type (bar graphs). While point comparison is a typical task, graph reading also comprises a number of equally common tasks, such as point location and identification of trends in the data (see e.g., Bertin, 1983; Schnotz, 1994; Wainer, 1992), and other well-established graph types, such as line graphs and pie charts (see e.g., Kosslyn, 1989; Lohse, Biolsi, Walker, & Rueter, 1994). Future studies should expand the research of the seductive details effect in the context of graph comprehension to additional tasks and graph types in order to evaluate the scope of our current findings.

Finally, it should be noted that we examined a sample of university students that reported a relatively high level of graph literacy. We argue that seductive details in graphs can be easily identified by educated graph readers because graphs feature highly conventionalized compositions (i.e., axes, labels and data points) that are easily distinguishable from external elements such as seductive text and illustrations. We suggest that future studies compare learners of different ages and different levels of graph literacy to help identify groups that may be more vulnerable to the effect of seductive details in graphs. This is underlined by the fact that the linear mixed-effects analyses revealed large variance components for students, highlighting the importance of individual factors for the graph comprehension process.

4.8 Conclusion

Expanding both the research on seductive details and on graph comprehension, this study evaluated the effect of seductive details on the performance in graph reading tasks. Results indicate that neither seductive text nor seductive pictures affected the students' performance. However, seductive details elevated the processing time of the graph reading task. Analyses of eye-fixation parameters further revealed that a large portion of additional processing time could be allocated to the processing of seductive details, but not to other regions of the display. This pattern of fixation times aligns well with one of the two mechanisms that were considered in this study: distraction rather than disruption. Since graphs are highly conventionalized systems, seductive details may likely be identified as task-irrelevant by educated readers, preventing them from being disrupted after a brief distraction.

Seductive details did not seem to do their damage when placed in graphs. While working memory capacity did not affect the seductive details effect, it was positively related to performance in the graph task. We suggest that future research should focus on the study of less experienced graph readers and individual factors in general.

4.9 References

- Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, 8, 47–89. doi:10.1016/s0079-7421(08)60452-1
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi:10.18637/jss.v067.i01
- Bertin, J. (1983). *Semiology of graphics: Diagrams, networks, maps*. Madison, WI: University of Wisconsin Press.
- Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction*, 20, 155–166.
- Cleveland, W., & McGill, R. (1987). Graphical perception: The visual decoding of quantitative information on graphical displays of data. *Journal of the Royal Statistical Society. Series A (General)*, 150(3), 192–229. doi:10.2307/2981473
- Dodeen, H. (2008). Assessing test-taking strategies of university students: Developing a scale and estimating its psychometric indices. *Assessment & Evaluation in Higher Education*, 33, 409–419.
- Doolittle, P. E., & Altstaedter, L. L. (2009). The effect of working memory capacity on multimedia learning: Does attentional control result in improved performance? *Journal of Research in Innovative Teaching*, 2(1), 7–23.
- Dorman, J. P. (2008). The effect of clustering on statistical tests: An illustration using classroom environment data. *Educational Psychology*, 28, 583–595. doi:10.1080/01443410801954201
- Duchowski, A. (2007). *Eye tracking methodology: Theory and practice*. London, UK: Springer.
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87(3), 215–251.
- Fox, J., & Weisberg, S. (2011). *An {R} Companion to Applied Regression, Second Edition*. Thousand Oaks, CA: Sage. Retrieved from <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>

Kapitel 4 - Seductive Details in the Context of Graph Comprehension

- Garner, R., Brown, R., Sanders, S., & Menke, D. J. (1992). "Seductive details" and learning from text. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The Role of Interest in Learning and Development* (pp. 239–254). Hillsdale, NJ: Erlbaum.
- Garner, R., Gillingham, M. G., & White, C. S. (1989). Effects of "seductive details" on macroprocessing and microprocessing in adults and children. *Cognition and Instruction*, 6(1), 41–57. doi:10.1207/s1532690xci0601_2
- Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: On the distinction between emotional interest and cognitive interest. *Journal of Educational Psychology*, 89(1), 92–102. doi:10.1037/0022-0663.89.1.92
- Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, 90(3), 414–434. doi:10.1037/0022-0663.90.3.414
- Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods*, 39, 101–117. doi:10.3758/BF03192848
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. UK: Oxford University Press.
- Hyönä, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction*, 20, 172–176. doi:10.1016/j.learninstruc.2009.02.013
- Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, 20, 146–154.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329–354. doi:10.1037/0033-295X.87.4.329
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795.
- Kim, S., & Lombardino, L. J. (2015). Comparing graphs and text: Effects of complexity and task. *Journal of Eye Movement Research*, 8(3), 1–17.
- Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, 3, 185–225. doi:10.1002/acp.2350030302

Kapitel 4 - Seductive Details in the Context of Graph Comprehension

- Kumar, N., & Benbasat, I. (2004). The effect of relationship encoding, task type and complexity on information representation: An empirical investigation of 2D and 3D graphs. *MIS Quarterly*, 28, 255–281.
- Lehman, S., Schraw, G., McCrudden, M. T., & Hartley, K. (2007). Processing and recall of seductive details in scientific text. *Contemporary Educational Psychology*, 32(4), 569–587. doi:10.1016/j.cedpsych.2006.07.002
- Lenzner, A., Schnotz, W., & Müller, A. (2013). The role of decorative pictures in learning. *Instructional Science*, 41(5), 811–831. doi:10.1007/s11251-012-9256-z
- Lindner, M. A., Eitel, A., Strobel, B., & Köller, O. (2017). Identifying processes underlying the multimedia effect in testing: An eye-movement analysis. *Learning and Instruction*, 47, 91–102. doi:10.1016/j.learninstruc.2016.10.007
- Lindner, M. A., Eitel, A., Thoma, G.-B., Dalehefte, I. M., Ihme, J. M., & Köller, O. (2014). Tracking the decision making process in multiple-choice assessment: Evidence from eye movements. *Applied Cognitive Psychology*, 28, 738–752. doi:10.1002/acp.3060
- Lindner, M. A., Ihme, J. M., Saß, S., & Köller, O. (2016). How representational pictures enhance students' performance and test-taking pleasure in low-stakes assessment. *European Journal of Psychological Assessment*. Advance online publication. doi:10.1027/1015-5759/a000351
- Lohse, G. L., Biolsi, K., Walker, N., & Rueter, H. H. (1994). A classification of visual representations. *Communications of the ACM*, 37, 36–49. doi:10.1145/198366.198376
- Manor, O., & Zucker, D. M. (2004). Small sample inference for the fixed effects in the mixed linear model. *Computational Statistics & Data Analysis*, 46(4), 801–817.
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511811678
- Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction*, 20, 167–171. doi:10.1016/j.learninstruc.2009.02.012
- Mayer, R. E., & Fiore, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity

Kapitel 4 - Seductive Details in the Context of Graph Comprehension

- principles. In R. E. Mayer (Ed.), *Cambridge handbooks in psychology. The Cambridge handbook of multimedia learning* (pp. 279–315). doi:10.1017/CBO9781139547369.015
- Mayer, R. E., Griffith, E., Jurkowitz, I. T. N., & Rothman, D. (2008). Increased interestingness of extraneous details in a multimedia science presentation leads to decreased learning. *Journal of Experimental Psychology: Applied*, *14*(4), 329–339. doi:10.1037/a0013835
- Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes Factors for common designs. R package version 0.9.12-2. Retrieved from <https://CRAN.R-project.org/package=BayesFactor>
- Nagelkerke, N. J. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, *78*(3), 691–692.
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142.
- Oberauer, K., Süß, H. M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity—facets of a cognitive ability construct. *Personality and Individual Differences*, *29*(6), 1017–1045. doi:10.1016/S0191-8869(99)00251-2
- Paas, F., & Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbooks in psychology. The Cambridge handbook of multimedia learning* (pp. 27–42). doi:10.1017/CBO9781139547369.015
- Peebles, D., & Cheng, P. C. H. (2003). Modeling the effect of task and graphical representation on response latency in a graph reading task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *45*, 28–46. doi: 10.1518/hfes.45.1.28.27225
- Peshkam, A., Mensink, M. C., Putnam, A. L., & Rapp, D. N. (2011). Warning readers to avoid irrelevant information: When being vague might be valuable. *Contemporary Educational Psychology*, *36*(3), 219–231. doi:10.1016/j.cedpsych.2010.10.006
- Pinker, S. (1990). A theory of graph comprehension. In R. O. Freedle (Ed.), *Artificial intelligence and the future of testing* (pp. 73–126). Hillsdale, England: Lawrence Erlbaum Associates.

Kapitel 4 - Seductive Details in the Context of Graph Comprehension

- Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*, 413–425. doi:10.1016/j.jml.2008.02.002
- R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Raftery, A. (1995). Bayesian Model selection in social research. *Sociological Methodology*, *25*, 111–163. doi:10.2307/271063
- Rey, G. D. (2011). Seductive details in multimedia messages. *Journal of Educational Multimedia and Hypermedia*. *20*(3), 283–314. Retrieved from <https://www.learntechlib.org/p/36221/>
- Rey, G. D. (2012). A review of research and a meta-analysis of the seductive detail effect. *Educational Research Review*, *7*(3), 216–237. doi:10.1016/j.edurev.2012.05.003
- Rowland-Bryant, E., Skinner, C. H., Skinner, A. L., Saudargas, R., Robinson, D. H., & Kirk, E. R. (2009). Investigating the interaction of graphic organizers and seductive details: Can a graphic organizer mitigate the seductive-details effect? *Research in the Schools*, *16*(2), 29–40.
- Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory & Cognition*, *34*(2), 344–355. doi:10.3758/BF03193412
- Schnotz, W. (1994). Wissenserwerb mit logischen Bildern [Knowledge acquisition with logical pictures]. In B. Weidenmann (Ed.), *Wissenserwerb mit Bildern* (pp. 95–148). Bern, Switzerland: Huber.
- Schnotz, W. (2002). Commentary: Towards an integrated view of learning from text and visual displays. *Educational Psychology Review*, *14*, 101–120. doi:10.1023/A:1013136727916
- Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review*, *14*, 47–69. doi:10.1023/A:1013180410169
- 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*, 690–702. doi:10.1037/0022-0663.91.4.690

Kapitel 4 - Seductive Details in the Context of Graph Comprehension

- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Los Angeles, CA: Sage.
- Strobel, B., Saß, S., Lindner, M. A., & Köller, O. (2016). Do graph readers prefer the graph type most suited to a given task? Insights from eye tracking. *Journal of Eye Movement Research, 9*(4), 1–15. doi:10.16910/jemr.9.4.4
- Strobel, B., Lindner, M. A., Saß, S., & Köller, O. (in press). Task-irrelevant data impair processing of graph reading tasks: An eye tracking study. *Learning and Instruction*.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295–312. doi:10.1016/0959-4752(94)90003-5
- van Gog, T., Kester, L., Nievelstein, F., Giesbers, B., & Paas, F. (2009). Uncovering cognitive processes: Different techniques that can contribute to cognitive load research and instruction. *Computers in Human Behavior, 25*, 325–331. doi:10.1016/j.chb.2008.12.021
- van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction, 20*, 95–99. doi:10.1016/j.learninstruc.2009.02.009
- Wainer, H. (1992). Understanding graphs and tables. *Educational Researcher, 21*, 14–23. doi:10.3102/0013189X021001014
- Wiley, J. (2003). Cognitive and educational implications of visually-rich media: Images and imagination. In: M. Hooks & M. Kendrick (Eds.), *Eloquent images: Writing visually in new media* (pp. 201–218). Cambridge, MA: MIT Press.
- Winn, W. D. (1990). A theoretical framework for research on learning from graphics. *International Journal of Educational Research, 14*, 553–564. doi:10.1016/0883-0355(90)90025-4
- Wright, P., Milroy, R., & Lickorish, A. (1999). Static and animated graphics in learning from interactive texts. *European Journal of Psychology of Education, 14*(2), 203–224. doi:10.1007/BF03172966
- Wright, R. D., & Ward, L. M. (2008). *Orienting of attention*. New York, NY: Oxford University Press.

Kapitel 5 - Zusammenfassung und Gesamtdiskussion

Das Lesen und Verstehen von Diagrammen ist ein wichtiger Bestandteil der naturwissenschaftlichen Grundbildung (z. B. Glazer, 2011; Millar, 2006; National Council of Teachers and Mathematics, 2000; OECD, 2016). Der Erwerb entsprechender Fähigkeiten soll dazu beitragen, Herausforderungen der fortschreitenden technischen Entwicklung und Digitalisierung zu meistern und Individuen eine vollwertige gesellschaftliche Teilhabe zu ermöglichen (Bybee & Fuchs, 2006; Roberts, 2007).

Die vorliegende Dissertation hat sich dem Einfluss visueller Gestaltungsmerkmale auf kognitive Prozesse bei der Bearbeitung von Diagrammleseaufgaben gewidmet. Diese Einflüsse spielen im Kontext von Lehren und Lernen in den Naturwissenschaften eine besondere Rolle, da für den Einsatz in der Lehre vorgefertigtes Material verwendet oder neues gestaltet werden muss. Dabei ist zunächst jedoch ungeklärt, ob das Material im Anbetracht seiner Gestaltungsmerkmale überhaupt für den geplanten Einsatz geeignet ist. Diese Arbeit leistet mit drei empirischen Studien einen Beitrag dazu, Kenntnisse darüber zu gewinnen, wie geeignete Repräsentationsformen für gegebene Aufgaben identifiziert werden können und wie Merkmale von Diagrammen so angepasst werden können, dass diese für einen intendierten Zweck geeignet sind.

In den folgenden Abschnitten werden die Hintergründe, Fragestellungen und Ergebnisse der empirischen Dissertationsstudien zusammenfassend dargestellt. Die erste Studie widmete sich der Frage, ob Personen dazu in der Lage sind, für gegebene Aufgaben dasjenige Diagramm zur Bearbeitung zu nutzen, das für die Lösung der jeweiligen Aufgabe am besten geeignet ist (Abschnitt 5.1). In der zweiten Studie wurde untersucht, ob eine hohe Datenkomplexität den Bearbeitungsprozess und die Leistung in Aufgaben mit Diagrammen beeinflusst, auch wenn zusätzliche Daten vollständig irrelevant für die Bearbeitung einer Aufgabe sind (Abschnitt 5.2). Die dritte Studie lieferte erste Erkenntnisse zu der Frage, ob interessante, aber irrelevante Zusatzinhalte (Seductive Details) in Diagrammen den Bearbeitungsprozess und die Leistung von Personen beeinflussen, die Diagrammleseaufgaben bearbeiten (Abschnitt 5.3). Schließlich werden die Ergebnisse der drei Experimentalstudien integrativ betrachtet und gemeinsam vor dem Hintergrund theoretischer und empirischer Arbeiten aus der Diagrammforschung diskutiert (Abschnitt 5.4).

5.1 Zusammenfassung Studie 1

Die erste Dissertationsstudie mit dem Titel „Do Graph Readers Prefer the Graph Type Most Suited to a Given Task? Insights from Eye Tracking“ fußt auf den Theorien und empirischen Befunden zu computationalen Unterschieden verschiedener Diagrammtypen (z. B. Freedman & Shah, 2002; Pinker, 1990; Shah & Hoeffner, 2002) und widmet sich der Frage, ob Personen in der Lage sind, bei der Bearbeitung unterschiedlicher Aufgaben jeweils den Diagrammtyp auszuwählen, der für die Aufgabe am besten geeignet ist (siehe auch Abschnitt 1.5.1).

Aus vorangehenden empirischen Untersuchungen und theoretischen Arbeiten ist bereits ersichtlich, dass zwei informational äquivalente Diagramme sich dennoch computational unterscheiden können (Kosslyn, 1989; Larkin & Simon, 1987; Pinker, 1990). Das heißt, Aufgaben, die mit einem Diagramm schnell und einfach bearbeitet werden können, dauern länger und sind schwieriger mit einem anderen Diagramm zu lösen, obwohl beide die gleichen Daten repräsentieren. Dies ist beispielsweise bei Säulen- und Liniendiagrammen der Fall. So ist es einfacher, Trends in Liniendiagrammen zu identifizieren, während Säulendiagramme die Bestimmung von Punktdifferenzen erleichtern (Pinker, 1990; Shah & Hoeffner, 2002; Zacks & Tversky, 1999; siehe auch Abschnitt 1.3.3). Allerdings ist wenig darüber bekannt, ob Personen dazu in der Lage sind, die jeweiligen computationalen Unterschiede zu erkennen und bei der Bearbeitung einer entsprechenden Aufgabe zu nutzen. Das Hauptziel dieser Studie war es, zu untersuchen, ob Personen ihre Bearbeitungsstrategie an die Anforderungen der Aufgabe anpassen können, wenn ihnen multiple Repräsentationen zur Verfügung stehen.

Zur Untersuchung des Bearbeitungsprozesses wurden Blickbewegungen mittels Eyetracking aufgezeichnet. Während bisherige Eyetracking-Studien hauptsächlich die Verarbeitung einzeln präsentierter Diagramme untersucht haben (siehe z. B. Shah & Carpenter, 1995; Goldberg & Helfman, 2011; Kim & Lombardino, 2015; Peebles & Cheng, 2003), war eine besondere Herausforderung dieser Studie, Personen die Wahl zwischen zwei Diagrammtypen bei der Bearbeitung der Aufgaben zu überlassen. Um die bevorzugte Nutzung zweier Diagrammtypen bei der Bearbeitung unterschiedlicher Aufgaben zu untersuchen, wurde ein eigens dafür konzipiertes *Dual-Representation*-Paradigma in Verbindung mit Blickbewegungsmaßen eingesetzt, sodass zwei unterschiedliche Diagrammtypen (mit identischer Datengrundlage) gleichzeitig angezeigt und zur Bearbeitung genutzt werden konnten.

In einem Within-Subject-Messwiederholungsdesign wurde der Aufgabentyp (Identifikation eines Trends vs. Bestimmung einer Punktdifferenz) experimentell variiert. Wenn Personen in der Lage sind, differenzielle Verarbeitungsstrategien in Abhängigkeit des Aufgabentyps anzuwenden, sollten sie für Trendaufgaben bevorzugt das Liniendiagramm nutzen und für Differenzaufgaben bevorzugt das Säulendiagramm betrachten. $N = 32$ Studierende haben an der Studie teilgenommen. Sie erhielten für jede Aufgabe zwei inhaltsgleiche Diagramme; ein Säulendiagramm und ein Liniendiagramm. Um die Aufgaben zu lösen, mussten die Teilnehmenden Aussagen zu Trends und Punktdifferenzen im True-False-Antwortformat bewerten. Als abhängige Variable wurde auf Basis der Blickbewegungsdaten ein Präferenzmaß für die Nutzung der Diagramme berechnet. Hierzu wurde zunächst die Summe aller Einzelfixationen auf die Diagrammbereiche getrennt nach Diagrammtyp berechnet. Anschließend wurde die Differenz der beiden Summen als relative Diagrammpräferenz gebildet. Als Kontrollvariable wurde mittels eines Fragebogens die generelle Präferenz für die beiden Diagrammtypen erhoben. Die Auswertung der Daten erfolgte mit Mehrebenenmodellen, in denen die Zugehörigkeit eines Trials zu je einer Person und je einer konkreten Aufgabe durch die Mehrebenenstruktur berücksichtigt wurde (siehe z. B. Hoffman & Rovine, 2007; Snijders & Bosker, 2012).

Die Ergebnisse zeigen, dass sich die relative Diagrammpräferenz der Teilnehmenden in Abhängigkeit des Aufgabentyps signifikant unterschied, wenngleich sich die Präferenz im Testverlauf unabhängig vom Aufgabentyp weiter zum Liniendiagramm hin verlagerte. Im Mittel zeigten die Teilnehmenden eine signifikant stärkere Präferenz für die Verwendung von Liniendiagrammen, wenn sie eine Trendaufgabe bearbeiteten. Erhielten sie hingegen eine Differenzaufgabe, so zeigte sich eine signifikant höhere Präferenz für das Säulendiagramm. Zusätzlich wurde die Präferenz bei der Aufgabenbearbeitung auch von der selbstberichteten, aufgabenunabhängigen Präferenz für einen der beiden Diagrammtypen beeinflusst. Im zeitlichen Verlauf verlagerte sich die Präferenz weiter in Richtung des Liniendiagramms. Dies resultierte in einer verstärkten Präferenz von Liniendiagrammen in Trendaufgaben und einer ausgewogenen Nutzung beider Diagrammtypen bei Differenzaufgaben. Weitere Ergebnisse aus den Fragebogendaten zeigten, dass die Teilnehmenden ihre eigene Diagrammkompetenz relativ hoch einschätzen, sodass man davon ausgehen kann, dass es sich um eine Stichprobe erfahrener Diagrammleserinnen und -leser handelte. Dies spiegelte sich auch in einer hohen Lösungsrichtigkeit bei der Aufgabenbearbeitung wider.

In der Zusammenschau lässt sich festhalten, dass erfahrene Diagrammleserinnen und -leser tatsächlich denjenigen Diagrammtyp bevorzugt zur Lösung einer Aufgabe verwendeten,

der für diese aus theoretischer Sicht auch besser geeignet ist. Der Effekt der selbstberichteten, aufgabenunabhängigen Präferenz auf die tatsächliche Präferenz war statistisch signifikant, jedoch klein im Vergleich zum Effekt des Aufgabentyps. Der Aufgabentyp erwies sich in dieser Studie somit als stärkerer Prädiktor für das Verhalten bei der Aufgabenbearbeitung als die selbstberichtete Präferenz für einen der beiden Diagrammtypen.

Die Entwicklung der Präferenz bei der Bearbeitung von Differenzaufgaben hin zu einer ausgewogenen Nutzung beider Diagrammtypen lässt sich dadurch erklären, dass Differenzurteile insgesamt schwieriger zu fällen waren als Trendurteile, weil für Differenzen immer mehrere Datenpunkte berücksichtigt werden mussten, während Trends anhand einer einzigen Linie abgelesen werden können (siehe Pinker, 1990). Das ist ein plausibler Grund dafür, dass Teilnehmende bei diesen Aufgaben zunehmend dazu neigten, ihre Urteile anhand beider Diagramme zu überprüfen, indem sie häufiger zwischen den beiden Diagrammen hin- und herschauten. Dieses Ergebnis legt nahe, dass für einige Aufgabentypen auch die integrative Nutzung mehrerer Repräsentationen eine nützliche Strategie sein kann, um die Aufgaben erfolgreich zu lösen. Die integrative Nutzung multipler Diagrammtypen ist bislang kaum erforscht worden und stellt somit ein vielversprechendes Desiderat der Diagrammforschung dar.

5.2 Zusammenfassung Studie 2

In der zweiten Studie mit dem Titel „Task-irrelevant Data Impair Processing of Graph Reading Tasks: An Eye Tracking Study“ wurde der Einfluss von irrelevanten Datenpunkten und Datenreihen auf die Aufgabenbearbeitung untersucht (siehe auch Abschnitt 1.5.2). Aus der Literatur ist bekannt, dass eine hohe Datenkomplexität zu einer starken kognitiven Belastung führen kann, insbesondere dann, wenn die Komplexität nicht angemessen für die gestellte Aufgabe ist (Demetriadis & Cadoz, 2005; Huang, Eades, & Hong, 2009; Sedig & Parsons, 2013). Im Gegensatz zu früheren Studien, die sich mit Komplexität von Daten in Diagrammen beschäftigt haben (z.B. Kim & Lombardino, 2015; Kumar & Benbasat, 2004; Spence & Lewandowski, 1991), lag der Fokus dieser Dissertationsstudie auf dem Einfluss *aufgaben-irrelevanter* Datenpunkte und Datenreihen. Dabei stellte sich die Frage, ob zusätzliche Daten die Bearbeitungszeit, Fehlerquote und kognitive Belastung auch dann erhöhen, wenn sie vollständig irrelevant für die gestellte Aufgabe sind. Dahinter stand die Annahme, dass der Einfluss irrelevanter Datenpunkte sich von demjenigen einer irrelevanten Datenreihe unterscheidet. Da Personen bei zusätzlichen Datenpunkten die Beschriftungen aller Punkte prüfen müssen, um relevante von irrelevanten Punkten zu unterscheiden, sollte die

Aufgabenbearbeitung mehr Zeit benötigen, fehleranfälliger sein und zu einer höheren kognitiven Belastung führen. Wenn eine bestehende Datenreihe um eine zusätzliche, irrelevante Datenreihe ergänzt wird, erlaubt ein Farbcode mittels Legende die perzeptuelle Gruppierung der Datenpunkte zu zwei Datenreihen (Pinker, 1990; Wertheimer, 1923). Personen müssen daher lediglich die Legende prüfen, um festzustellen, welche der beiden Datenreihe relevant und welche nicht relevant für die Lösung einer gestellten Aufgabe ist. Die negativen Effekte der irrelevanten Daten auf die Fehlerquote, Bearbeitungszeit und die kognitive Belastung sollten daher geringer sein als bei aufgaben-irrelevanten Datenpunkten.

$N = 60$ Studierende bearbeiteten in einem Within-Subject-Messwiederholungsdesign computerisierte Aufgaben zu Säulendiagrammen im klassischen Multiple-Choice-Format mit einer korrekten Antwort unter vier Alternativen, die entweder (a) irrelevante Datenpunkte, (b) eine irrelevante Datenreihe oder (c) keine aufgaben-irrelevanten Daten enthielten. Zum Lösen der Aufgaben mussten die Probanden mehrere Datenpunkte ablesen, diese vergleichen und unter vier Antwortalternativen die korrekte auswählen. Die Auswertung der Daten erfolgte in kreuzklassifiziertem Mehrebenenmodellen mit Random Intercepts für Personen und Stimuli (siehe z. B. Hoffman & Rovine, 2007; Snijders & Bosker, 2012). Als abhängige Maße wurden neben Fehlerquote, Bearbeitungszeit und kognitiver Belastung je Aufgabe mittels Eyetracking auch die Gesamtfixationszeiten auf unterschiedliche Bereiche der Diagrammdarstellung untersucht (Aufgabenmaterial, Achsen und Label, relevante Daten und irrelevante Daten). Mittels Fragebogen haben die Teilnehmenden dieser Studie wie schon in Studie 1 eine Einschätzung ihrer Fähigkeiten im Diagrammlesen abgegeben.

Die Ergebnisse zeigen kleine, aber signifikante Effekte von irrelevanten Datenpunkten auf die Bearbeitungszeit, die Fehlerquote und die empfundene kognitive Belastung. Die Darbietung einer irrelevanten Datenreihe resultierte hingegen in Effekten die zunächst überraschend waren: So wiesen die Modelle auf große Leistungseinbußen in den Diagrammleseaufgaben hin, wenn eine irrelevante Datenreihe zu sehen war, obwohl diese durch eine farbliche Markierung prinzipiell hätte leicht identifiziert werden können. Teilnehmende benötigten entgegen der Erwartung signifikant länger für die Bearbeitung, machten mehr Fehler und gaben eine höhere kognitive Belastung an. Anders als in Studien, in denen nur reine Outcome-Maße zur Verfügung stehen, konnten in dieser Untersuchung allerdings Prozessmaße dazu verwendet werden, mögliche Gründe für die Unterschiede zu identifizieren (siehe z. B. Mayer, 2010). In der Tat lieferte die Analyse der Blickbewegungsdaten eine mögliche Erklärung für die zunächst unerwarteten Befunde: Wenn irrelevante Datenpunkte zu sehen waren, verbrachten Probanden erwartungsgemäß zwar mehr

Zeit mit dem Lesen der Achsenbeschriftungen, aber insgesamt genauso viel Zeit mit dem Lesen der Aufgabe und dem Ablesen der Daten. War jedoch eine zusätzliche Datenreihe im Diagramm zu sehen, erhöhte sich nicht nur die Beschäftigung mit den irrelevanten Daten deutlich, sondern auch die Zeit, die Probanden mit der Aufgabenstellung verbrachten. Dies deutet darauf hin, dass die Präsenz einer zusätzlichen Datenreihe die Anforderungen der Aufgabe verändert hat. Nach Grice (1975, 1978) stellt es nämlich einen Verstoß der Kommunikationsregeln dar, wenn mehr Informationen dargeboten werden als nötig. Personen können durch das Überangebot an Informationen verunsichert werden und zu dem Schluss kommen, dass es einen Grund dafür gibt, weshalb zusätzliche Informationen zur Verfügung stehen. Zur Identifikation der relevanten Datenreihe war deshalb möglicherweise ein zusätzlicher (und fehleranfälliger) Arbeitsschritt nötig, nämlich die erneute Prüfung des Aufgabentextes. Auf diese Weise schienen verunsicherte Teilnehmende erneut zu überprüfen, welche Datenreihe sie tatsächlich zur Lösung der Aufgabe benötigten.

Man kann zusammenfassen, dass aufgaben-irrelevante Datenpunkte in dieser Studie nur einen kleinen Einfluss auf die Bearbeitung von Diagrammleseaufgaben ausübten. Zusätzliche Datenreihen veränderten die Bearbeitung der Aufgabe jedoch deutlich und führten zu großen Leistungseinbußen. Es liegt daher nahe, dass einer unnötigen kognitiven Belastung in Lernmaterialien vorgebeugt werden kann, indem Darstellungen auf relevante Datenreihen bzw. Variablen beschränkt werden.

5.3 Zusammenfassung Studie 3

Die dritte Studie mit dem Titel „Do seductive details do their damage in the context of graph comprehension? Insights from eye movements“ widmete sich der Forschungsfrage, ob und in welchem Ausmaß der Seductive-Details-Effekt (Harp & Mayer, 1997) bei der Bearbeitung von Aufgaben mit Diagrammen auftritt (siehe auch Abschnitt 1.5.3). Als Seductive Details werden solche Inhalte bezeichnet, die als interessant empfunden werden und im thematischen Zusammenhang zum übrigen Material stehen, jedoch irrelevant für das Ziel der Aufgabe sind (Garner, Brown, Sanders, & Menke, 1992; Garner, Gillingham, & White, 1989). Harp und Mayer (1997, 1998) unterscheiden außerdem zwischen Seductive Text für interessante, aber irrelevante Texte und Seductive Illustrations für interessante, aber irrelevante Abbildungen.

Zwar gibt es Hinweise darauf, dass Seductive Details das Interesse von Lernenden steigern können (siehe z. B. Harp & Mayer, 1997; Lenzner, Schnotz, & Müller, 2013), zahlreiche Studien bescheinigen den verführerischen Zusatzinhalten jedoch einen schädlichen

Einfluss auf die Behaltens- und Transferleistung (z. B. Harp & Mayer, 1997; Lehman, Schraw, McCrudden, & Hartley, 2007; Rowland-Bryant et al., 2009). Typischerweise wurden Seductive Details im Zusammenhang von Lernleistung mit Texten untersucht (siehe z. B. Mayer, Griffith, Jurkowitz, & Rothman, 2008; Rey, 2011), während die Diagrammforschung sich vornehmlich dem Einfluss von Merkmalen gewidmet hat, die Teil der konventionellen Diagrammstruktur sind (z. B. Legenden, Beschriftungen und Datenpunkte; Cleveland & McGill, 1987; Kim & Lombardino, 2015; Kumar & Benbasat, 2004; Shah & Hoeffner, 2002).

Digitale Medien (z. B. Präsentationen, Websites und computerisiertes Lernmaterial) ermöglichen es Lehrenden, Diagramme in Lehr- und Testmaterialien ebenso leicht mit zusätzlichen Inhalten anzureichern wie Texte. So können in Diagramme beispielsweise Abbildungen oder zusätzliche Beschreibungen von Datenpunkten integriert werden, die für das Ablesen der Daten und die Bearbeitung einer zugehörigen Aufgabe keinerlei Bedeutung haben. Allerdings gibt es Grund zur Annahme, dass das Lesen und Verstehen von Diagrammen ebenfalls von Seductive Details beeinträchtigt werden kann: Bei Diagrammen handelt es sich um hochgradig konventionalisierte Strukturen, die räumliche Relationen als Analogie für nicht-räumliche Relationen nutzen (Schnotz, 2002; Winn, 1990). Aus diesem Grund können sie anfällig für überflüssige visuelle Elemente sein, die vom Leseprozess ablenken oder ihn unterbrechen.

Erklärungsansätze für den Seductive-Details-Effekt stehen in Zusammenhang mit der Annahme, dass das menschliche Arbeitsgedächtnis nur begrenzte Kapazitäten hat, sodass überflüssige Elemente zu einer hohen kognitiven Belastung führen können (*Cognitive Load Theory*; siehe z. B. Paas & Sweller, 2014). Harp und Mayer (1998) postulieren drei mögliche Mechanismen: *Distraction*, *Disruption* und *Diversion*. Die Distraction-Hypothese geht davon aus, dass Seductive Details die Aufmerksamkeit von Lernenden auf sich ziehe und damit vom relevanten Material ablenke. Demgegenüber liegt der Disruption-Hypothese die Annahme zugrunde, dass der Prozess einer mentalen Modellbildung (siehe z. B. Schnotz, 2002) durch Seductive Details unterbrochen werde, was zu einem inkohärenten mentalen Modell führe. In der Diversion-Hypothese werden die Effekte von Seductive Details hingegen auf die Aktivierungen von unpassendem Vorwissen attribuiert, durch welches zwar ein mentales Modell entstehe, dieses werde jedoch um die irrelevanten Informationen herum aufgebaut.

Sanchez und Wiley (2006) betonen die Rolle von Arbeitsgedächtniskapazität in Zusammenhang mit Seductive Details und verweisen damit insbesondere auf Distraction und Disruption als bedeutsame Erklärungsmodelle. In ihrer Studie zeigte sich, dass Personen mit

einer niedrigen Arbeitsgedächtniskapazität stärker vom Seductive-Details-Effekt betroffen waren als solche mit hoher Kapazität.

Um den Seductive-Details-Effekt im Kontext von Diagrammverstehen zu untersuchen, wurden zwei Arten von Seductive Details in typische Säulendiagramme integriert. In einem Within-Subject-Messwiederholungsdesign bearbeiteten $N = 68$ Studierende computerisierte Ableseaufgaben zu Diagrammen, die entweder (a) irrelevante Abbildungen (Seductive Illustrations), (b) irrelevante Texte (Seductive Text) oder (c) kein irrelevantes Material enthielten. Zum Lösen der Aufgaben mussten mehrere Datenpunkte verglichen und unter vier Antwortalternativen die korrekte ausgewählt werden. Es wurde erwartet, dass die Teilnehmenden mehr Fehler machen und länger für die Bearbeitung der Diagrammleseaufgaben benötigen, wenn Seductive Details in ein Diagramm integriert sind und dass diese Effekte größer sind bei Teilnehmenden mit einer niedrigen Arbeitsgedächtniskapazität als bei solchen mit hoher Kapazität. Mithilfe der Blickbewegungsmaße wurden zwei potenzielle Erklärungen für den Seductive-Details-Effekt im Kontext von Diagrammen untersucht: Handelt es sich bei dem zugrundeliegenden Mechanismus um Distraction, sollte zusätzliche Bearbeitungszeit auf die Inspektion der Seductive Details entfallen, jedoch nicht auf eine längere Betrachtung anderer Regionen. Geht der Effekt stattdessen auf Disruption zurück, sollte zusätzliche Bearbeitungszeit sowohl auf die Betrachtung der Seductive Details als auch auf die übrigen Diagrammregionen entfallen, da eine Unterbrechung der mentalen Modellbildung dazu führt, dass zuvor inspizierte Regionen erneut betrachtet werden müssen.

Die Analyse der Daten erfolgte in kreuzklassifiziertem Mehrebenenmodellen mit Random Intercepts für Personen und Stimuli (siehe z. B. Hoffman & Rovine, 2007; Snijders & Bosker, 2012). Zudem wurden Bearbeitungszeit je Aufgabe, Lösungshäufigkeit sowie die Fixationszeiten auf funktionale Diagrammbereiche (Aufgabenmaterial, Achsen und Label, Daten und Seductive Details) als abhängige Maße erhoben. Darüber hinaus wurde ein computerisierter Test für räumliche Arbeitsgedächtniskapazität von Oberauer, Süß, Schulze, Wilhelm und Wittmann (2000) adaptiert und eingesetzt. Mittels Fragebogen haben die Teilnehmenden zudem eine Einschätzung ihrer Fähigkeiten im Diagrammlesen abgegeben. Neben Likelihood-Ratio-Tests und R^2 -Maßen wurden für Modellvergleiche auch Bayes Factors berechnet, da sie ein kontinuierliches Maß für „Evidenz“ für eines von zwei verglichenen Modellen bieten (Kass & Raftery, 1995; Raftery, 1995).

Wie die Ergebnisse der Modellvergleiche mit hoher Übereinstimmung zeigen, gab es in dieser Untersuchung keinen Effekt von den Seductive Details auf die Leistung in den Diagrammleseaufgaben, allerdings war die Bearbeitungszeit in beiden Seductive-Details-Bedingungen signifikant höher als in der Kontrollbedingung. Die Inspektion der Fixationszeiten auf die verschiedenen Regionen zeigte keine bedeutsamen Unterschiede zwischen den beiden Experimentalbedingungen und der Kontrollbedingung, mit einer Ausnahme: zusätzliche Fixationszeit entfiel jeweils auf die Seductive-Details. Dieses Muster in den Blickbewegungsdaten lässt sich besser mit der Distraction-Hypothese erklären als mit der Disruption-Hypothese. Offenbar waren die Studierenden durch die Seductive Details kurz abgelenkt, im Übrigen jedoch unbeeinträchtigt.

Wie eine aktuelle Studie zeigt, werden Personen von Seductive Details weniger stark beeinträchtigt, wenn ihnen bewusst ist, dass ihnen neben Aufgaben-relevanten Inhalten auch irrelevante Informationen gezeigt werden (Peshkam, Mensink, Putman & Rapp, 2011). Eine mögliche Erklärung für die vorliegenden Befunde kann deshalb sein, dass die Studierenden schnell in der Lage waren, die Seductive Details als irrelevant zu identifizieren. Diese Annahme wird unter anderem dadurch gestützt, dass die Studierenden relativ hohe Fähigkeiten im Diagrammlesen berichteten. Das heißt, sie waren mit den Säulendiagrammen vertraut genug, um unmittelbar zu erkennen, dass die Seductive Details nicht zu den funktionalen Elementen des Diagramms gehörten. So ist denkbar, dass die hochgradig konventionalisierte Struktur von Diagrammen einen abschirmenden Effekt gegenüber Seductive Details entfalten kann, jedenfalls wenn Personen mindestens moderate Fähigkeiten im Lesen von Diagrammen mitbringen. Die Arbeitsgedächtniskapazität hat das Auftreten oder die Stärke des Seductive-Details-Effekts in dieser Studie nicht beeinflusst, eine hohe Kapazität ging jedoch mit einer niedrigeren Fehlerquote einher.

Zusammengenommen haben Seductive Details beim Bearbeiten von Diagrammleseaufgaben nicht all ihre negativen Effekte entfaltet. Während sich kein Effekt auf die Leistung zeigte, erhöhte sich allerdings die Bearbeitungszeit. Dieser Effekt ließ sich am besten durch Distraction erklären (d.h. die Studierenden waren durch die Seductive Details kurz abgelenkt, andernfalls war der Bearbeitungsprozess jedoch unverändert). Die Arbeitsgedächtniskapazität der Studierenden zeigte keine signifikante Interaktion mit dem Seductive-Details-Effekt, eine hohe Kapazität war aber mit besserer Leistung in den Diagrammleseaufgaben assoziiert. Diese Befunde legen nahe, dass Seductive Details in Diagrammen keine große Gefahr für Personen darstellen, die mindestens moderate Fähigkeiten im Diagrammlesen haben. Die Ergebnisse verweisen zudem auf die Bedeutung

von Individualfaktoren (wie Arbeitsgedächtniskapazität) für das Lesen und Verstehen von Diagrammen.

5.4 Diskussion und Forschungsausblick

Die vorliegende Dissertation hat sich dem Einfluss visueller Gestaltungsmerkmale auf kognitive Prozesse der Bearbeitung von Diagrammleseaufgaben gewidmet. Die erste Studie untersuchte die Fähigkeit von Personen, für gegebene Aufgaben dasjenige Diagramm zu verwenden, das für die Bearbeitung der jeweiligen Aufgabe von Vorteil war. Wie die Ergebnisse zeigten, präferierten die Teilnehmenden tatsächlich die Darstellungsformen, die auch theoriegemäß computationale Vorteile für die gestellten Aufgaben bieten. Dies war jedoch deutlicher bei der Bearbeitung von Trendaufgaben (mit Liniendiagrammen) als bei der Bearbeitung von Differenzaufgaben (mit Säulendiagrammen). Die zweite Studie nahm den Einfluss aufgaben-irrelevanter Datenpunkte auf die Bearbeitung von Diagrammleseaufgaben in den Fokus. Dabei kam heraus, dass insbesondere irrelevante Datenreihen zu einer höheren kognitiven Belastung, längeren Bearbeitungszeiten und schlechterer Leistung führten. Die Teilnehmenden zeigten außerdem ein verändertes Bearbeitungsverhalten, welches nahelegte, dass die irrelevanten Informationen eine erneute Prüfung des Aufgabentextes nach sich zogen. Schließlich ging es in der dritten Studie um die Frage, ob und in welchem Ausmaß der aus der Multimedia-Forschung bekannte Seductive-Details-Effekt auch beim Bearbeiten von Diagrammleseaufgaben auftritt. Die verführerischen Zusatzinhalte in den Diagrammen wirkten sich nicht auf die Leistung in den Aufgaben aus, verlängerten jedoch die Bearbeitungszeit. Anders als in der zweiten Studie zeigten sich keine fundamentalen Änderungen im Blickbewegungsverhalten, wenn irrelevante Text- oder Bildinformationen zu sehen waren. Stattdessen ließ sich die zusätzliche Bearbeitungszeit am besten mit der Distraction-Hypothese erklären, die besagt, dass der Seductive-Details-Effekt auf eine Ablenkung der Aufmerksamkeit zurückgeht.

Mit der Zusammenschau der drei Dissertationsstudien folgt in diesem Abschnitt eine integrative Betrachtung und gemeinsame Diskussion. Zunächst sind die Ergebnisse der Studien im Einklang mit bisherigen Befunden, dass die Bearbeitung von Aufgaben mit Diagrammen sowohl von der Art der Aufgabe als auch von Merkmalen der Gestaltung abhängt (z. B. Kim & Lombardino, 2015; Peebles & Cheng, 2003; Shah & Hoeffner, 2002; Shah, Mayer, & Hegarty, 1999; Simkin & Hastie, 1987). Studie 1 stützte darüber hinaus die Annahme, dass Personen in der Lage sind, für Aufgaben das jeweils geeignetere Diagramm

auszuwählen. Die gleichzeitige Vorgabe von zwei Diagrammen führte allerdings auch dazu, dass die Teilnehmenden mitunter beide Darstellungsformen integrativ zur Lösung der Aufgaben verwendeten. In Abschnitt 5.4.1 wird auf dieser Grundlage diskutiert, welche Implikationen die Verfügbarkeit multipler, informational äquivalenter Diagramme für die Bearbeitung von Diagrammleseaufgaben mit sich bringt.

Der Vergleich der Studien 2 und 3 weist darauf hin, dass aufgaben-irrelevante Merkmale in Art und Gestaltung derart verschieden sein können, dass sich ihr Einfluss auf die Aufgabenbearbeitung drastisch unterscheiden kann. Diese Beobachtung wird in Abschnitt 5.4.2 vertiefend diskutiert.

Wenngleich Gestaltungsmerkmale im Mittelpunkt dieser Arbeit stehen, zeigte sich in den Ergebnissen, dass Personenmerkmale ein wichtiges Komplement zu diesem Fokus darstellen. Die Rolle individueller Einflussfaktoren wird in Abschnitt 5.4.3 vor dem Hintergrund der vorliegenden Studien diskutiert. Anschließend werden Implikationen für die Gestaltung von Diagrammen als Lehr- und Testmaterial aus den Studienergebnissen abgeleitet (Abschnitt 5.4.4), die Limitationen der Studien zusammenfassend dargelegt (Abschnitt 5.4.5) und mit einer Conclusio geschlossen (Abschnitt 5.4.6).

5.4.1 Multiple informational äquivalente Diagramme

Der Überblick über die Literatur zum Diagrammverstehen zeichnet ein klares Bild, dass der Erfolg beim Lesen von Diagrammen sowohl von der Art der Aufgabe als auch von Gestaltungsmerkmalen des Diagramms abhängt (z. B. Kim & Lombardino, 2015; Peebles & Cheng, 2003; Simkin & Hastie, 1987; Shah & Hoeffner, 2002; Shah et al., 1999). Keine Darstellungsform besitzt eine Universaleignung für alle denkbaren Fragen, die man an sie richten kann (vgl. Bertin, 1983; Schnotz, 1994; Wainer, 1992; siehe auch Abschnitt 1.2.1). Die Vorzüge der von Diagrammen genutzten Verräumlichung beziehen sich viel eher auf jeweils bestimmte Verarbeitungsprozesse. So kann ein Diagramm etwa dann effizient verarbeitet werden, wenn die perzeptuellen Einheiten, in welche die Abbildung beim Wahrnehmungsprozess zerlegt wird, denen von einer Aufgabe geforderten Sinneinheiten entsprechen oder wenn die von der Aufgabe geforderten Operationen bereits auf perzeptueller Ebene durchgeführt werden können (Freedman & Shah, 2002; Pinker, 1990). Dies ist beispielsweise dann der Fall, wenn Trends in Liniendiagrammen identifiziert werden sollen. Zum einen entspricht ein konzeptueller Trend genau einer Linie und die Ausprägung des Trends der Steigung dieser Linie (siehe auch Abschnitt 1.3.3).

Die erste Studie dieser Arbeit konnte über die etablierten Befunde hinaus zeigen, dass Personen bei Vorgabe multipler Repräsentationen diejenige bevorzugt verwenden, welche die Lösung der gestellten Aufgabe erleichtert. So präferierten die Teilnehmerinnen und Teilnehmer Liniendiagramme zur Lösung von Trendaufgaben, tendierten aber zur Nutzung von Säulendiagrammen, wenn sie Vergleichsaufgaben erhielten. Im Verlauf der Untersuchung zeigte sich allerdings, dass Personen bei Vergleichsaufgaben zunehmend dazu neigten, beide Diagramme gleichermaßen zu nutzen. Dieser Befund kann nicht nur als Hinweis dafür gedeutet werden, dass sich die Spezifität der Eignung für die beiden Aufgabentypen möglicherweise unterscheiden könnte (sie scheint bei Liniendiagrammen und Trendaufgaben größer zu sein) oder dass Vergleichsaufgaben inhärent schwieriger sind als Trendaufgaben. Die Ergebnisse zeigten darüber hinaus, dass die Darbietung multipler, informational äquivalenter Repräsentationen eine neue Bearbeitungsstrategie ermöglicht, nämlich die integrative Nutzung beider Darstellungsformen.

Bislang ist das Lesen und Verstehen multipler (und informational äquivalenter) Diagramme wenig untersucht. Allerdings werfen die Resultate der ersten Studie eine Reihe weiterer Fragen auf, die zum Gegenstand künftiger Untersuchungen werden sollten. Dazu zählt zunächst die Frage, unter welchen Umständen multiple Diagramme von Lesenden genutzt werden und ob diese Strategie zu höherem Erfolg bei der Lösung einer Aufgabe führt, auch wenn die Abbildungen, wie in Studie 1, inhaltsgleich sind und sich nur in Merkmalen der Darstellung unterscheiden. Anknüpfungspunkte für diese Forschungsfrage bieten Untersuchungen zu repräsentationalen Bildern und multiplen Repräsentationen (z. B. Ainsworth, 1999; Vekiri, 2002). Hierbei werden typischerweise inhaltsgleiche Abbildungen und Texte betrachtet. Beispielsweise konnten Lindner, Ihme, Saß und Köller (2016) zeigen, dass repräsentationale Bilder die Itemschwierigkeiten von naturwissenschaftlichen Aufgaben reduzierten, wenn sie Teile des Textes visualisierten, die zur Lösung der Aufgabe relevant waren. Koć-Januchta, Höffler, Thoma, Prechtl und Leutner (2017) fanden zudem heraus, dass Personen, die Informationen bevorzugt visuell verarbeiten, in Lernmaterial mit Bildern und Texten bevorzugt Bilder betrachteten und anschließend eine bessere Behaltensleistung zeigten als Personen, die bevorzugt die Texte verarbeitet haben. Multimedia-Theorien attribuieren die positiven Effekte multipler Repräsentationen im Besonderen auf die unterschiedliche Modalität der Darbietung (*Cognitive Theory of Multimedia Learning*, Mayer, 2005; *Integrated Model of Text and Picture Comprehension*, Schnotz & Bannert, 2003). Die duale Kodierung von Informationen in einem verbalen (Text) und einem visuellen Kanal (Bild) kann demnach die Geschwindigkeit und Genauigkeit verbessern, mit der Personen mentale

Modelle von einem Sachverhalt anlegen (Eitel, Scheiter, Schüler, Nyström, & Holmqvist, 2013; Schnotz & Bannert, 2003). Es stellt sich daher die Frage, ob die Darbietung multipler Diagramme ähnliche Effekte erzielen kann, auch wenn sie vorwiegend nur eine Modalität bedienen (den visuellen Kanal). Weiter bleibt zu klären, ob und inwiefern Personen sich in ihrem Verarbeitungsstil von multiplen Diagrammen unterscheiden (vorwiegende Nutzung eines oder des anderen Diagramms gegenüber einer ausgewogenen Nutzung). Dieses Nutzungsverhalten kann darüber hinaus mit weiteren Personenmerkmalen zusammenhängen, beispielsweise den individuellen Fähigkeiten im Diagrammlernen oder Vorwissen zu den dargestellten Inhalten (siehe z. B. Freedman & Shah, 2002; Shah & Hoeffner, 2002). Generell stellt der Einfluss von Personenmerkmalen einen weiteren, elementaren Anknüpfungspunkt für diese Forschungsarbeit dar und wird in Abschnitt 5.4.3 weiter diskutiert.

5.4.2 Systeminhärente und systemfremde Gestaltungsmerkmale

Die zweite und dritte Studie dieser Arbeit bilden, wie die Ergebnisse aufzeigen, einen bemerkenswerten Kontrast. In Studie 2 wurde die Datenkomplexität in Diagrammen variiert, indem die Abbildungen um Aufgaben-irrelevante Datenpunkte oder Datenreihen erweitert wurden. Die Ergebnisse zeigten teils große Effekte durch die experimentelle Manipulation auf die Performanz wie auch auf das Blickbewegungsverhalten. In dieser Studie zeigte sich eine qualitative Veränderung im Bearbeitungsprozess, wenn eine zusätzliche Datenreihe in den Diagrammen zu sehen war: Personen beschäftigten sich nicht nur mit den irrelevanten Daten, sondern auch wiederholt mit dem Aufgabentext. Eine erneute Prüfung des Aufgabentextes wurde womöglich dazu genutzt, sich darüber zu versichern, welche Datenpunkte zur Lösung der Aufgabe relevant sind.

In Studie 3 wurden Seductive Details in typische Diagramme eingefügt, sodass sie entweder zusätzliche Fakten oder ergänzende Abbildungen enthielten, die zum Lösen der Aufgaben nicht relevant waren. Diese Manipulation zeigte jedoch keinen Effekt auf die Leistung der Teilnehmenden. Wenngleich die Bearbeitungszeit bei Aufgaben mit Seductive Details etwas höher war, blieb das Blickbewegungsverhalten in dieser dritten Studie überwiegend unverändert (es entfielen jeweils nur kurze Fixationszeiten auf die Betrachtung der Seductive Details). Dieser Befund war zunächst überraschend, da Seductive Details im Kontext von Lernen und Textverstehen wiederholt negative Effekte auf die Behaltens- und Transferleistung gezeigt haben (z. B. Doolittle & Altstadter, 2009; Garner et al., 1989; Harp & Mayer, 1997; Lehman et al., 2007). Dass sie ausgerechnet bei der Bearbeitung von Aufgaben in Diagrammen keinen starken Einfluss zu entfalten scheinen, ist womöglich kein

Zufall, sondern könnte mit dem stark konventionalisierten, räumlichen Aufbau von Diagrammen zusammenhängen. Während Texte viele Freiheitsgrade in Bezug auf die Gestaltung besitzen (Wahl von Ausdruck und Begriffen, Satz- und Absatzstruktur und viele mehr), handelt es sich bei Diagrammen um hochgradig konventionalisierte, räumliche Systeme, in denen es vergleichsweise wenig Freistellen gibt, die aufgefüllt werden können. Da den meisten räumlichen Parametern eine essenzielle Funktion in der Darstellung zukommt (z. B. Position und Höhe von Säulen für die Ausprägung zweier Variablen; siehe Pinker, 1990), können sie nicht nach Belieben variiert werden. Texte haben vergleichsweise weniger starke Beschränkungen.

Die räumliche Struktur von Diagrammen ist mitunter ein Grund dafür, warum sie eine effizientere kognitive Verarbeitung ermöglichen. Um die Zusammenhänge zwischen zwei Variablen aus einem Text zu gewinnen, müssen Lesende mehrere aufeinander folgende Sätze lesen und zahlreiche Variablen im Arbeitsgedächtnis halten. Erst wenn Personen eine entsprechende Textpassage zu Ende gelesen haben, können sie beispielsweise bestimmen, an welchem Punkt eine Variable die größte Ausprägung besitzt. In einem Diagramm lassen sich viele derartige Prozesse bereits auf perzeptueller Ebene lösen, etwa indem man die längste Säule in einem Säulendiagramm identifiziert und die Bezeichnung an der richtigen Position abliest (Pinker, 1990; siehe auch Abschnitt 1.2.2.2). Diese Eigenschaften machen Diagramme womöglich auch resistenter gegenüber dem Einfluss zusätzlicher Informationen, die in anderen Kontexten zu kognitiven Belastungen führen. Sind den Lesenden die Darstellungskonventionen bekannt (wie es in der vorliegenden Studie mit Studierenden der Fall war), so fällt es ihnen womöglich leichter, diejenigen Elemente zu identifizieren, die zum Lösen der Aufgabe nicht relevant sind. Diese Annahme findet Unterstützung in einer Studie von Peshkam et al. (2011). Die Untersuchung zeigte, dass Personen von Seductive Details weniger stark beeinträchtigt wurden, wenn ihnen bewusst war, dass ihnen irrelevante Informationen gezeigt werden. In vergleichbarer Weise zeigte sich in der dritten Studie, dass Teilnehmende nur kurze Zeit mit der Betrachtung von Seductive Details verbrachten und auch sonst kaum Veränderungen bei der Betrachtung der Diagramme aufwiesen. Demgegenüber benötigten Teilnehmende der zweiten Studie signifikant mehr Zeit für die Bearbeitung einer Aufgabe, wenn zusätzliche, irrelevante Datenpunkte zu sehen waren. Bei Datenpunkten handelt es sich nämlich um systeminhärente Merkmale, also solche, die in der Diagrammkonvention enthalten sind.

Die Zusammenschau der beiden Studien führt zu einer wichtigen Conclusio. Als hochgradig konventionalisierte, räumliche Systeme sind Diagramme womöglich resistenter

gegenüber Einflüssen externaler Gestaltungsmerkmale. Aufgaben-irrelevante Inhalte können von Lesenden leichter identifiziert und bei der Bearbeitung ausgesondert werden als systeminhärente Merkmale, bei denen genauer geprüft werden muss, ob sie relevant sind oder nicht. Es wird Aufgabe zukünftiger Untersuchungen sein, die genauen Bedingungen zu identifizieren, unter denen systemfremde Gestaltungsmerkmale (wie Seductive Details) Einflüsse nehmen können oder nicht. So könnte es beispielsweise von den individuellen Fähigkeiten im Diagrammlernen abhängen, ob Personen hinreichend mit den Konventionen vertraut sind, um zu erkennen, dass die zusätzlichen Informationen nicht zum eigentlichen Diagramm gehören. Da die Stichproben der vorliegenden Studien hohe Ausprägungen in den selbstberichteten Diagrammlesefähigkeiten aufwiesen, eigneten sie sich nicht zur Untersuchung dieses Merkmals. Die Rolle von individuellen Merkmalen für künftige Untersuchungen wird im folgenden Abschnitt ausführlich diskutiert.

5.4.3 Der Einfluss von Individualfaktoren

Während der Fokus dieser Forschungsarbeit auf der Untersuchung von Aufgabenmerkmalen lag, insbesondere auf visuellen Gestaltungsmerkmalen von Diagrammen, legen die Ergebnisse der Studien einhellig nahe, dass individuelle Faktoren für das Lösen von Aufgaben mit Diagrammen darüber hinaus eine wichtige Rolle spielen. So fanden sich in den Zufallseffekten der Mehrebenenmodelle in allen drei Studien große Varianzkomponenten auf Personenebene. Dies galt sowohl in den Modellen für Fehlerquoten und Bearbeitungszeiten bei den Aufgaben als auch für Fixationszeiten, die durch Eyetracking ermittelt wurden. Das bedeutet, ein beträchtlicher Teil der Variabilität, die in der Leistung und im Blickverhalten der Teilnehmenden zu beobachten war, kann auf den Zufallsfaktor Person attribuiert werden. Der Einfluss individueller Faktoren stellt dementsprechend ein wichtiges Forschungsdesiderat dar.

Das Interaktionsmodell des Diagrammverstehens von Freedman und Shah (2002) bietet einige Anhaltspunkte, welche Personenmerkmale beim Lesen von Diagrammen eine Rolle spielen können: (1) domänenspezifisches Wissen, (2) Fähigkeiten im Diagrammlernen und (3) Fähigkeiten im Erklären.

In den Studien, die in dieser Arbeit zusammengefasst sind, wurde der Versuch unternommen, den Einfluss von domänenspezifischem Wissen so gering wie möglich zu halten. So waren die verwendeten Aufgaben stets derart gestaltet, dass zur Lösung nur Ablesefähigkeiten gefordert waren, jedoch kein spezifisches Vorwissen zu den dargestellten Inhalten. Allerdings konnten bereits Lord, Ross und Lepper (1979) zeigen, dass die Erwartung

von Personen über Zusammenhänge zwischen Variablen ihre Interpretation beeinflussen kann. Das gilt insbesondere dann, wenn gegebene Daten nicht mit den vorherigen Erwartungen konsistent sind (Shah & Hoeffner, 2002). Es ist deshalb nicht auszuschließen, dass Erwartungen über die präsentierten Inhaltsbereiche auf individueller Ebene einen Einfluss auf die Aufgabenbearbeitung genommen haben. Dies könnte in zukünftigen Studien kontrolliert werden, indem Inhalte konstant gehalten und die Erwartungen von Personen dazu explizit erhoben werden.

Die berichteten Fähigkeiten im Diagrammlesen zeigten in den Studien dieser Forschungsarbeit keinen Einfluss auf die Leistung und Blickbewegungsmaße und wurden deshalb nicht in den finalen Mehrebenenmodellen berücksichtigt. Bereits in den deskriptiven Statistiken war in den Stichproben ein Deckeneffekt in den berichteten Fähigkeiten erkennbar, sodass auch nicht zu erwarten ist, dass diese Variable in den vorliegenden Studien weitere Varianz aufklärt. Wie zahlreiche andere Studien nahelegen, scheint der Einfluss von Diagrammlesefähigkeiten jedoch vor allem bei jüngeren Schülerinnen und Schülern eine bedeutsame Rolle zu spielen (z. B. Baker, Corbett, & Koedinger, 2001, Bell & Janvier, 1981; Gattis & Holyoak, 1996; Schiano & Tversky, 1992; Tversky & Schiano, 1989). Für zukünftige Studien wird die allgemeine Kompetenz, Daten aus Diagrammen korrekt abzulesen, weiterhin ein wichtiges Personenmerkmal darstellen, insbesondere bei der Untersuchung von jüngeren Schülerinnen und Schülern, anderen weniger leistungsstarken Personengruppen bzw. in Stichproben, die eine höhere Heterogenität in der Erfahrung mit Diagrammen erwarten lassen.

Fähigkeiten im Erklären von Diagrammen wie sie bei Freedman und Shah (2002) genannt werden, wurden in den Studien dieser Arbeit nicht überprüft. In einer Studie der Autoren zeigte sich, dass Experten häufiger in der Lage waren, spontane Erklärungen für ein Diagramm zu generieren als Novizen (Freedman & Shah, 2001). Laut den Autoren könne die Interpretation von Diagrammen durch die Fähigkeiten im Erklären auf ähnliche Weise beeinflusst werden wie durch Vorwissen. Da die Experten allerdings auch ein höheres Fachwissen besitzen und vertrauter im Umgang mit Diagrammen sind, ist nicht klar, ob die Fähigkeiten im Erklären von diesen Merkmalen getrennt werden können. Dennoch bietet das Wissen um diese Einflüsse einen wichtigen Anknüpfungspunkt für zukünftige Untersuchungen.

Über die von Freedman und Shah (2002) zusammengefassten, individuellen Einflussfaktoren hinaus sollte auch die Arbeitsgedächtniskapazität als potenziell relevantes Personenmerkmal in Betracht gezogen werden. Wie sich in der zweiten Studie zeigte, können

Diagrammmerkmale wie irrelevante Daten zu einer zusätzlichen kognitiven Belastung führen (siehe auch Studie 2). Entsprechend kann die individuelle Arbeitsgedächtniskapazität einen Einfluss darauf haben, ob beziehungsweise wie stark von Gestaltungsmerkmalen ausgehende Belastungen tatsächlich zu Leistungseinbußen führen. In der dritten Studie fand sich zwar kein Einfluss von der individuellen Arbeitsgedächtniskapazität auf das Vorhandensein oder die Stärke des Seductive-Details-Effekts (cf. Sanchez & Wiley, 2006), jedoch war eine höhere Kapazität mit einer besseren Leistung in den Diagrammleseaufgaben assoziiert (siehe auch Studie 3). Bei dieser Untersuchung kam ein Maß für räumliches Arbeitsgedächtnis zum Einsatz, das von Oberauer et al. (2000) adaptiert wurde. Die Autorinnen und Autoren geben an, dass simultane Speicherung und Transformation sowie das Erinnern räumlicher Strukturen die wesentlichen Anforderungen des Tests darstellen. Dieses Anforderungsprofil weist große Überschneidungen mit den Anforderungen von typischen Ableseaufgaben auf, bei denen Personen ebenfalls mehrere räumlich angeordnete Elemente im Gedächtnis halten oder mental transformieren müssen. Wie die dritte Dissertationsstudie nahelegt, kann sich die Erfassung von räumlicher Arbeitsgedächtniskapazität als vielversprechendes Werkzeug für die Diagrammforschung erweisen, beispielsweise zur Aufklärung interindividueller Leistungsunterschiede im Diagrammlesen oder für Belege konvergenter Validität bei der Entwicklung von Tests für Diagrammlesefähigkeiten.

5.4.4 Implikationen für Lehren und Testen

Aus den Studienergebnissen dieser Arbeit lassen sich einige Implikationen für die Gestaltung von Diagrammen als Teil von Lehrmaterialien und Tests ableiten. Da sich die Implikationen für die Kontexte des Lernens und Testens zum Teil unterscheiden, werden sie in den folgenden Abschnitten getrennt voneinander diskutiert.

5.4.4.1 Implikationen für die Gestaltung von Diagrammen in Lehrmaterialien

Wie die Studien 2 und 3 dieser Arbeit nahelegen, wirken sich irrelevante Daten und andere interessante, aber irrelevante Elemente (Seductive Details) unterschiedlich auf die Bearbeitung von Aufgaben mit Diagrammen aus. Die zweite Studie zeigte signifikante, aber kleine Unterschiede in der Leistung und der berichteten kognitiven Belastung, wenn Aufgaben-irrelevante Datenpunkte im Diagramm zu sehen waren. Wurden allerdings ganze Aufgaben-irrelevante Datenreihen in Diagramme integriert, erhöhte das die Bearbeitungszeit, die Fehlerquote und die kognitive Belastung deutlich. Blickbewegungsmuster zeigten, dass Personen wiederholt den Aufgabentext studieren mussten, um zu prüfen, welche Datenreihe

tatsächlich relevant für die gestellte Aufgabe ist. In der dritten Studie zeigte sich demgegenüber, dass die Präsentation von Seductive Details kaum einen Einfluss auf die Aufgabenbearbeitung ausübte. Weder die Analyse der Leistung noch die Auswertung des Blickbewegungsverhaltens wiesen bedeutsame Unterschiede auf, ganz gleich ob zusätzliche Abbildungen oder Kurztexte eingefügt wurden.

Da Lehrende häufig auf vorgefertigtes Material zurückgreifen, kann sich mitunter die Frage stellen, ob es sinnvoll ist, die Abbildung vor dem Einsatz anzupassen. Die berichteten Studienergebnisse deuten darauf hin, dass vor allem irrelevante Datenreihen zu Verwirrung führen und die Bearbeitung dadurch erschweren können, sogar wenn Personen, wie in Studie 2, hohe Fähigkeiten im Lesen von Diagrammen berichten (siehe Studie 2). Sind Datenreihen oder ganze Variablen für ein Lernziel nicht relevant, so scheint es empfehlenswert, Diagramme derart zu überarbeiten, dass sie nur noch die benötigten Informationen darstellen. Andernfalls kann das Material die Erwartung erzeugen, dass die zusätzlichen Informationen einen bestimmten Zweck erfüllen. Personen können in der Folge verunsichert darüber sein, ob und inwiefern sie die Informationen berücksichtigen sollen (Grice, 1975, 1978). Einzelne irrelevante Datenpunkte in einer bestehenden Datenreihe stellten zumindest für die fähigen Personen in der zweiten Studie kein großes Problem dar, sodass es vertretbar erscheint, einzelne, irrelevante Datenpunkte in der Abbildung zu belassen. Dasselbe gilt für Seductive Details in Form kleiner Abbildungen und zusätzlicher Fakten, die in keinem direkten Bezug zu der zu bearbeitenden Aufgabe stehen. Möglicherweise haben Seductive Details deshalb keinen starken Einfluss auf die Bearbeitung entfaltet, weil sie aufgrund der hochgradig konventionalisierten Diagrammstruktur leicht als externale (und irrelevante) Elemente erkannt wurden und somit mental im Lösungsprozess ausgeblendet werden können. Bei Personen mit guten Kenntnissen über Diagramme, scheinen Seductive Details in Diagrammen eine vertretbare Ergänzung sein. Ob der Einsatz in Lehrmaterialien darüber hinaus einen Zugewinn darstellen kann, sollte in zukünftigen Untersuchungen aufgeklärt werden.

Bei der Interpretation dieser Ergebnisse müssen die jeweils hohen Ratings für die eigenen Fähigkeiten im Diagrammlesen beachtet werden. Die Befunde gelten demnach zunächst nur für Personen mit mindestens moderaten Fähigkeiten im Lesen von Diagrammen. Zukünftige Untersuchungen sollten die Resultate an Stichproben mit jüngeren oder weniger erfahrenen Personen überprüfen oder verschiedene Alters- und Fähigkeitsgruppen miteinander vergleichen. So wäre beispielsweise denkbar, dass Seductive Details in Diagrammen einen stärkeren Einfluss entfalten, wenn Personen geringere Fähigkeiten im Diagrammlesen besitzen (siehe auch Abschnitt 5.4.3). Darüber hinaus wurden Seductive

Details in der dritten Studie derart in die Diagramme integriert, dass sie Ablese-Prozesse nicht durch Verdeckung relevanter Bereiche behindern. Es ist davon auszugehen, dass die Okklusion funktionaler Diagrammbereiche Ableseprozesse erheblich erschwert oder vollständig unmöglich macht (beispielsweise wenn ein Wert nicht mehr abgelesen oder die Höhe einer Säule nicht mehr genau bestimmt werden kann). Aus diesem Grund ist von einer derartigen Integration in Lehrmaterialien bereits a priori abzuraten. Um die Relevanz derartiger Einflüsse abzuschätzen, sollten sie in zukünftigen Studien allerdings genauer untersucht werden.

5.4.4.2 Implikationen für die Gestaltung von Diagrammen in Leistungstests

Die vorgestellten Befunde sind grundsätzlich auch auf die Gestaltung von Diagrammleseaufgaben in Leistungstests anwendbar, wie etwa bei naturwissenschaftlichen Aufgaben in PISA (OECD, 2016). Die Implikationen unterscheiden sich jedoch im Testkontext von denen des Lernkontextes. Das Vorhandensein irrelevanter Daten oder zusätzlicher Details im Diagramm lassen sich grundsätzlich als schwierigkeitsgenerierende Eigenschaften der Aufgabe verstehen. Wie die Studien dieser Arbeit zeigen, kann die Leistung in einer Aufgabe mitunter von diesen Eigenschaften beeinflusst werden. Anforderungen einer Aufgabe, die nicht dem zu messenden Konstrukt zuzurechnen sind, können einen Test unnötig schwieriger machen und werden als Konstrukt-irrelevante Varianz bezeichnet (Messick, 1989). Somit sollten Gestaltungsmerkmale von Diagrammen auch beim Design von Testaufgaben berücksichtigt werden. Es ist insbesondere darauf zu achten, ob die aus Gestaltungsmerkmalen resultierenden Anforderungen in der Definition des jeweiligen Konstruktes enthalten sind. Wenn ein Test beispielsweise die Fähigkeiten im Diagrammlesen abbilden soll, ist das Trennen von relevanten und irrelevanten Daten wahrscheinlich als Teil des Konstruktes zu verstehen. Andersherum kann es Ziel eines Tests sein, Schlussfolgerndes Denken auf Basis gegebener Daten zu messen, sodass Schwierigkeiten, die aus Gestaltungsmerkmalen resultieren, eher der Konstrukt-irrelevanten Varianz zuzuordnen sind. Dasselbe gilt für eine Darstellung von Daten in Formaten, die zur Beantwortung der Frage ungeeignet sind, beispielsweise wenn Trends aus Balkendiagrammen abgelesen werden müssten, anstatt aus Liniendiagrammen.

Aus den dargestellten Gründen lassen sich für den Einsatz von Diagrammen in Leistungstests keine Gestaltungsprinzipien mit uneingeschränkter Gültigkeit ableiten. Stattdessen scheint es empfehlenswert, mit dem Wissen um den Einfluss einzelner Merkmale

die Gestaltung derart zu wählen, dass sie eine valide Messung im Sinne des angestrebten Konstrukts ermöglicht.

5.4.5 Limitationen

Um die Grenzen dieser Arbeit transparent abzustecken, werden in diesem Abschnitt zusammenfassend die zentralen Einschränkungen der durchgeführten Studien berichtet. In diesem Zuge wird auch auf zugehörige Forschungsdesiderate verwiesen.

In dieser Arbeit habe ich verschiedene Aufgabentypen (Abschnitt 1.2.1) und eine Vielzahl von Repräsentationsformen (Abschnitt 1.1.3) vorgestellt. Die berichteten Studien haben sich aus Gründen der Ökonomie und Vergleichbarkeit jedoch auf Vergleichsaufgaben (Studien 2 und 3) sowie den Vergleich dieser mit Trendaufgaben (Studie 1) beschränkt. Gegenstand der Untersuchung waren drüber hinaus vorwiegend (gruppierte) Säulendiagramme (Studien 2 und 3) sowie Liniendiagramme (Studie 1). Es existieren allerdings weitere Aufgabentypen (siehe Bertin, 1983; Schnotz, 1994; Wainer, 1992) und eine große Menge von Visualisierungen, die gleichermaßen gebräuchlich sind (z. B. Kreisdiagramme; siehe Bertin, 1983; Kosslyn, 1989; Lohse, Biolsi, Walker, & Rueter, 1994). Das ist insbesondere deshalb wichtig zu bemerken, weil die Leistung und Effizienz bei der Bearbeitung von Diagrammleseaufgaben von einer Interaktion zwischen Aufgabentyp und Darstellungsform abhängt (z. B. Pinker, 1990; Shah & Hoeffener, 2002). So ist beispielsweise denkbar, dass die Identifikation einzelner Datenpunkte oder Trends anders durch die Datenkomplexität beeinflusst wird als der Vergleich von Datenpunkten oder dass Datenkomplexität eine andere Bedeutung für die Bearbeitung hat, wenn andere Repräsentationsformen betrachtet werden oder sogar multiple Repräsentationen zur Verfügung stehen (siehe auch Abschnitt 5.4.1). Um den Gültigkeitsraum der vorliegenden Befunde abzustecken, sollten zukünftige Studien weitere Kombinationen von Aufgabentypen und Diagrammen in den Fokus der Untersuchung rücken.

Dies gilt im weiteren Sinne auch für die gewählten Operationalisierungen. So ist beispielsweise noch offen, ob und wie sich die Einflüsse von Datenkomplexität und Seductive Details mit der Anzahl zusätzlicher Elemente (Datenpunkte, Details) verändert. Beispielsweise fanden Spence und Lewandowsky (1991) heraus, dass sich die Leistung ihrer Teilnehmer für fünf und sieben Datenpunkte nicht unterschied. In beiden Bedingungen war die Leistung jedoch schlechter als bei vier Datenpunkten. In den Studien dieser Arbeit wurden sowohl irrelevante Datenpunkte als auch Seductive Details jeweils in einer fixen Anzahl in

Diagramme eingefügt. Zukünftige Untersuchungen können mit einer breiteren Variation dieser Manipulation an die bisherigen Ergebnisse anschließen.

Schließlich müssen die Stichproben der vorliegenden Studien vor dem Hintergrund möglicher Varianzeinschränkungen betrachtet werden. Zunächst handelt es sich aufgrund der aufwändigen Datenerhebungen mit Eyetracking-Methoden um vergleichsweise kleine Stichprobengrößen (Spannweite = 32–68). Auch wenn diese Schwäche durch Within-Subject-Messwiederholungsdesigns und die Verwendung von Mehrebenenmodellen gut kompensiert werden konnte, können Daten von größeren Stichproben zu einer zuverlässigeren Einschätzung beitragen. Weiterhin setzen sich die Stichproben der Studien ausschließlich aus Universitätsstudierenden zusammen, die jeweils hohe Fähigkeiten im Diagrammlesen berichteten. Es ist beispielsweise denkbar, dass ein hohes Niveau der Diagrammlesefähigkeiten die Einflüsse von Gestaltungsmerkmalen (teilweise) kompensieren kann. Dadurch ist insbesondere die Generalisierbarkeit der Befunde auf jüngere und leistungsschwächere Personen eingeschränkt, die vermutlich nicht über dieselben Kenntnisse im Diagrammlesen verfügen. Zukünftige Studien sollten diese Einschränkung unbedingt berücksichtigen und könnten Lernende unterschiedlicher Altersgruppen und verschiedener Fähigkeitsniveaus miteinander vergleichen, um beispielsweise besonders vulnerable Personengruppen zu identifizieren. In diesem Zusammenhang erscheint es essenziell, einen größeren Fokus auf Personenmerkmale zu richten und diese gemeinsam mit Kontextmerkmalen wie der visuellen Gestaltung von Diagrammen zu betrachten.

5.4.6 Conclusio

Die vorliegende Arbeit leistet einen Beitrag zur Erforschung von Gestaltungsmerkmalen in Diagrammen und ihren Einflüssen auf die kognitive Verarbeitung bei der Aufgabenbearbeitung. Zum einen zeigte sich, dass Personen unter Vorgabe multipler Repräsentationen bevorzugt diejenige Diagrammform verwenden, die für eine gestellte Aufgabe eine effizientere Verarbeitung ermöglicht. Personen sind demnach in der Lage, computationale Vorteile, die sich aus der räumlichen Darstellung in Diagrammen ergeben, bei der Aufgabenbearbeitung zu nutzen. Zum anderen demonstrieren die Studienergebnisse, dass die computationale Effizienz von Diagrammen von verschiedenen Gestaltungsmerkmalen auch unterschiedlich stark beeinflusst wird. So können aufgaben-irrelevante Merkmale insbesondere dann den Bearbeitungsprozess beeinträchtigen, wenn sie als Teil der Diagrammkonvention nicht unmittelbar als irrelevant erkannt werden können. Demgegenüber scheinen externale, irrelevante Merkmale, wie zusätzliche Bilder und Kurztexte, nicht

dieselben negativen Effekte zu entfalten, da für Personen offenbar leicht ersichtlich ist, dass sie nicht zur Diagrammkonvention gehören und deshalb zur Lösung der Diagrammleseaufgaben nicht beitragen.

Über diese Erkenntnisse hinaus unterstreicht die Arbeit den besonderen Nutzen von Eyetracking bei der Untersuchung von Diagrammen. In der ersten Studie ist es gelungen, Fixationszeiten auf unterschiedliche Diagrammtypen als Präferenzmaß zu nutzen. In den Folgestudien gewährten Blickbewegungsmaße genaueren Einblick in die raumzeitliche Aufmerksamkeitsverteilung bei der Aufgabenbearbeitung und konnten damit einen Beitrag zur Erklärung von Performanzunterschieden leisten. Aufgrund der hochgradig konventionalisierten, räumlichen Diagrammstruktur werden Blickbewegungsanalysen auch zukünftig einen fruchtbaren Ansatz in der Diagrammforschung darstellen.

Der Fokus auf den Einfluss von Gestaltungsmerkmalen auf die Bearbeitung von Diagrammleseaufgaben offenbarte zugleich die Bedeutung einer zweiten Seite derselben Medaille, nämlich den Einfluss individueller Merkmale. Große Varianzkomponenten in den Zufallseffekten der Mehrebenenmodelle wiesen auch statistisch darauf hin, dass Personenmerkmalen eine ähnlich große Rolle zukommt wie den Gestaltungsmerkmalen von Diagrammen. Eine integrative Betrachtung beider Seiten stellt womöglich die wichtigste Aufgabe der zukünftigen Diagrammforschung dar.

5.5 Literatur

- Ainsworth, S. (1999). The functions of multiple representations. *Computers & Education*, 33, 131–152. doi:10.1016/S0360-1315(99)00029-9
- Baker, R. S., Corbett, A. T., & Koedinger, K. R. (2001). Toward a model of learning data representations. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society* (pp. 45–50). Mahwah, NJ: Erlbaum.
- Bell, A., & Janvier, C. (1981). The interpretation of graphs representing situations. *For the learning of mathematics*, 2(1), 34–42.
- Bertin, J. (1983). *Semiology of graphics: Diagrams, networks, maps*. Madison, WI: University of Wisconsin Press.
- Bybee, R. W., & Fuchs, B. (2006). Preparing the 21st century workforce: A new reform in science and technology education. *Journal of Research in Science Teaching*, 43, 349–352. doi:10.1002/tea.20147
- Cleveland, W., & McGill, R. (1987). Graphical perception: The visual decoding of quantitative information on graphical displays of data. *Journal of the Royal Statistical Society. Series A (General)*, 150(3), 192–229. doi:10.2307/2981473
- Demetriadis, S., & Cadoz, C. (2005). A conceptual framework for the integrative design of adaptable representations for learning. In A. Méndez-Vilas, B. González-Pereira, J. Mesa González, & J. A. Mesa González (Eds.), *Recent research developments in learning technologies* (pp. 1250–1255). Badajoz, Spain: Formatex.
- Doolittle, P. E., & Altstaedter, L. L. (2009). The effect of working memory capacity on multimedia learning: Does attentional control result in improved performance? *Journal of Research in Innovative Teaching*, 2(1), 7–23.
- Eitel, A., Scheiter, K., Schüler, A., Nyström, M., & Holmqvist, K. (2013). How a picture facilitates the process of learning from text: Evidence for scaffolding. *Learning and Instruction*, 28, 48–63. doi:10.1016/j.learninstruc.2013.05.002
- Freedman, E. G., & Shah, P. (2001, November). *Individual differences in domain knowledge, graph reading skills, and explanatory skills during graph comprehension*. Paper presented at the 42nd Annual Meeting of the Psychonomic Society, Orlando, FL.

- Freedman, E. G., & Shah, P. (2002). Toward a model of knowledge-based graph comprehension. In M. Hegarty, B. Meyer, & N. H. Narayanan (Eds.), *Diagrammatic representation and inference* (pp. 8–31). Berlin, Germany: Springer.
- Garner, R., Brown, R., Sanders, S., & Menke, D. J. (1992). “Seductive details” and learning from text. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The Role of Interest in Learning and Development* (pp. 239–254). Hillsdale, NJ: Erlbaum.
- Garner, R., Gillingham, M. G., & White, C. S. (1989). Effects of “seductive details” on macroprocessing and microprocessing in adults and children. *Cognition and Instruction*, 6(1), 41–57. doi:10.1207/s1532690xci0601_2
- Gattis, M., & Holyoak, K. J. (1996). Mapping conceptual to spatial relations in visual reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(1), 231–239.
- Glazer, N. (2011). Challenges with graph interpretation: A review of the literature. *Studies in Science Education*, 47, 183–210. doi:10.1080/03057267.2011.605307
- Goldberg, J., & Helfman, J. (2011). Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. *Information Visualization*, 10, 182–195. doi:10.1177/1473871611406623
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York, NY: Academic Press.
- Grice, H. P. (1978). Some further notes on logic and conversation. In P. Cole (Ed.), *Syntax and semantics 9: Pragmatics* (pp. 113–128). New York, NY: Academic Press.
- Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: On the distinction between emotional interest and cognitive interest. *Journal of Educational Psychology*, 89(1), 92–102. doi:10.1037/0022-0663.89.1.92
- Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, 90(3), 414–434. doi:10.1037/0022-0663.90.3.414
- Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods*, 39, 101–117. doi:10.3758/BF03192848

- Huang, W., Eades, P., & Hong, S. H. (2009). Measuring effectiveness of graph visualizations: A cognitive load perspective. *Information Visualization*, 8(3), 139–152.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795.
- Kim, S., & Lombardino, L. J. (2015). Comparing graphs and text: Effects of complexity and task. *Journal of Eye Movement Research*, 8(3), 1–17. doi:10.16910/jemr.8.3.2
- Koć-Januchta, M., Höffler, T., Thoma, G. B., Prechtel, H., & Leutner, D. (2017). Visualizers versus verbalizers: Effects of cognitive style on learning with texts and pictures – An eye-tracking study. *Computers in Human Behavior*, 68, 170–179.
- Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, 3, 185–225. doi:10.1002/acp.2350030302
- Kumar, N., & Benbasat, I. (2004). The effect of relationship encoding, task type and complexity on information representation: An empirical investigation of 2D and 3D graphs. *MIS Quarterly*, 28, 255–281.
- Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11, 65–100. doi:10.1111/j.1551-6708.1987.tb00863.x
- Lehman, S., Schraw, G., McCrudden, M. T., & Hartley, K. (2007). Processing and recall of seductive details in scientific text. *Contemporary Educational Psychology*, 32(4), 569–587. doi:10.1016/j.cedpsych.2006.07.002
- Lenzner, A., Schnotz, W., & Müller, A. (2013). The role of decorative pictures in learning. *Instructional Science*, 41(5), 811–831. doi:10.1007/s11251-012-9256-z
- Lindner, M. A., Ihme, J. M., Saß, S., & Köller, O. (2016). How representational pictures enhance students' performance and test-taking pleasure in low-stakes assessment. *European Journal of Psychological Assessment*. Advance online publication. doi:10.1027/1015-5759/a000351
- Lohse, G. L., Biolsi, K., Walker, N., & Rueter, H. H. (1994). A classification of visual representations. *Communications of the ACM*, 37, 36–49. doi:10.1145/198366.198376
- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of personality and social psychology*, 37(11), 2098–2109. doi:10.1037/0022-3514.37.11.2098

Kapitel 5 - Zusammenfassung und Gesamtdiskussion

- Mayer R. E. (Ed.). (2005). *The Cambridge handbook of multimedia learning*. Cambridge, United Kingdom: University Press. doi:10.1017/CBO9780511816819.005
- Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction, 20*(2), 167–171. doi:10.1016/j.learninstruc.2009.02.012
- Mayer, R. E., Griffith, E., Jurkowitz, I. T. N., & Rothman, D. (2008). Increased interestingness of extraneous details in a multimedia science presentation leads to decreased learning. *Journal of Experimental Psychology: Applied, 14*(4), 329–339. doi:10.1037/a0013835
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (pp. 13–103). New York, NY: Macmillan.
- Millar, R. (2006). Twenty first century science: Insights from the design and implementation of a scientific literacy approach in school science. *International Journal of Science Education, 28*, 1499–1521. doi:10.1080/09500690600718344
- National Council of Teachers and Mathematics (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers and Mathematics.
- Oberauer, K., Süß, H. M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity—facets of a cognitive ability construct. *Personality and individual differences, 29*(6), 1017–1045. doi:10.1016/S0191-8869(99)00251-2
- OECD. (2016). *PISA 2015 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: OECD.
- Paas, F., & Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbooks in psychology. The Cambridge handbook of multimedia learning* (pp. 27–42). doi:10.1017/CBO9781139547369.015
- Peebles, D., & Cheng, P. C. H. (2003). Modeling the effect of task and graphical representation on response latency in a graph reading task. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 45*, 28–46. doi:10.1518/hfes.45.1.28.27225
- Peshkam, A., Mensink, M. C., Putnam, A. L., & Rapp, D. N. (2011). Warning readers to avoid irrelevant information: When being vague might be valuable. *Contemporary Educational Psychology, 36*(3), 219–231. doi:10.1016/j.cedpsych.2010.10.006

- Pinker, S. (1990). A theory of graph comprehension. In R. O. Freedle (Ed.), *Artificial intelligence and the future of testing* (pp. 73–126). Hillsdale, NJ Lawrence Erlbaum Associates.
- Raftery, A. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25, 111–163. doi:10.2307/271063
- Rey, G. D. (2011). Seductive details in multimedia messages. *Journal of Educational Multimedia and Hypermedia*, 20(3), 283–314. Retrieved from <https://www.learntechlib.org/p/36221/>
- Roberts, D. A. (2007). Scientific literacy/science literacy. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 729–780). Mahwah, NJ: Lawrence Erlbaum Associates.
- Rowland-Bryant, E., Skinner, C. H., Skinner, A. L., Saudargas, R., Robinson, D. H., & Kirk, E. R. (2009). Investigating the interaction of graphic organizers and seductive details: Can a graphic organizer mitigate the seductive-details effect? *Research in the Schools*, 16(2), 29–40.
- Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory & Cognition*, 34(2), 344–355. doi:10.3758/BF03193412
- Schiano, D. J., & Tversky, B. (1992). Structure and strategy in encoding simplified graphs. *Memory & Cognition*, 20(1), 12–20. doi:10.3758/BF03208249
- Schnotz, W. (1994). Wissenserwerb mit logischen Bildern. In B. Weidenmann (Ed.), *Wissenserwerb mit Bildern* (pp. 95–148). Bern, Switzerland: Huber.
- Schnotz, W. (2002). Commentary: Towards an integrated view of learning from text and visual displays. *Educational Psychology Review*, 14, 101–120. doi:10.1023/A:1013136727916
- Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, 13, 141–156. doi:10.1016/S0959-4752(02)00017-8

- Sedig, K., & Parsons, P. (2013). Interaction design for complex cognitive activities with visual representations: A pattern-based approach. *AIS Transactions on Human-Computer Interaction*, 5(2), 84–133.
- Shah, P., & Carpenter, P. A. (1995). Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology*, 124, 43–61. doi:10.1037/0096-3445.124.1.43
- Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review*, 14, 47–69. doi:10.1023/A:1013180410169
- 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), 690–702. doi:10.1037/0022-0663.91.4.690
- Simkin, D., & Hastie, R. (1987). An information-processing analysis of graph perception. *Journal of the American Statistical Association*, 82, 454–465. doi:10.1080/01621459.1987.10478448
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Los Angeles, CA: SAGE.
- Spence, I., & Lewandowsky, S. (1991). Displaying proportions and percentages. *Applied Cognitive Psychology*, 5, 61–77.
- Tversky, B., & Schiano, D. J. (1989). Perceptual and conceptual factors in distortions in memory for graphs and maps. *Journal of Experimental Psychology: General*, 118(4), 387–398. doi:10.1037/0096-3445.118.4.387
- Vekiri, I. (2002). What is the value of graphical displays in learning? *Educational Psychology Review*, 14, 261–312. doi:10.1023/A:1016064429161
- Wainer, H. (1992). Understanding graphs and tables. *Educational Researcher*, 21, 14–23. doi:10.3102/0013189X021001014
- Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. II. *Psychologische Forschung*, 4(1), 301–350.
- Winn, W. D. (1990). A theoretical framework for research on learning from graphics. *International Journal of Educational Research*, 14, 553–564. doi:10.1016/0883-0355(90)90025-4

Kapitel 5 - Zusammenfassung und Gesamtdiskussion

Zacks, J., & Tversky, B. (1999). Bars and lines: A study of graphic communication. *Memory & Cognition*, 27, 1073–1079. doi:10.3758/BF03201236

LEBENS LAUF

Benjamin Strobel

PERSÖNLICHE DATEN

Geburtstag	06. Februar 1988
Geburtstort	Flensburg
Familienstand	ledig
Staatsangehörigkeit	deutsch

STUDIUM

10/2008 – 12/2013	Studium der Psychologie (Diplom) Christian-Albrechts-Universität (CAU) in Kiel
01/2014 – 05/2018	Promotionsstudium (Psychologie) Christian-Albrechts-Universität (CAU) in Kiel

BERUFLICHE TÄTIGKEITEN

06/2009 – 06/2013	Wissenschaftliche Hilfskraft Institut für Psychologie an der CAU in Kiel, Arbeitseinheit Allgemeine Psychologie II, Wahrnehmung und Kognition (Leitung: Prof. Dr. Rainer Mausfeld)
seit 01/2014	Wissenschaftlicher Mitarbeiter Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik (IPN) in Kiel, Abteilung Erziehungswissenschaften (Leitung: Prof. Dr. Olaf Köller)
seit 06/2018	Wissenschaftlicher Mitarbeiter Grimme-Institut, Gesellschaft für Medien, Bildung und Kultur mbH Grimme Medienbildung, Arbeitsbereich Games und Gesellschaft (Leitung: Wolfgang Zielinski)