

B. Jacobs

Globale Vergleiche, lokale Vergleiche und Größenschätzungen in Liniendiagramm und Säulendiagramm unter Superposition und Juxtaposition.

Hinweis: Eine Kurzfassung des Artikels liegt als HTML- Dokument vor. Dort können Sie auch einige Beispiele für die experimentellen Graphen einsehen. Die URL-Adresse lautet:

<http://www.phil.uni-sb.de/FR/Medienzentrum/Grafikexperiment/drei/ueber.html>

Druckfehler, stilistische Ungereimtheiten und formale Unstimmigkeiten bitte ich zu entschuldigen. Das Medienzentrum verfügte weder über die Mittel noch über das Personal, vorliegenden Artikel Korrektur zu lesen.

Inhaltsverzeichnis

Inhaltsverzeichnis	2
Zusammenfassung.....	3
1.0 Einleitung und Einstieg in das Thema	4
2.0 Bisherige empirische Forschung	5
3.0 Die Fragestellungen der Untersuchung	5
4.0 Die experimentellen Bedingungen dieser Untersuchung	5
4.1 Die unabhängigen Variablen:	6
4.1.1 Variationen der Graphpositionierung unter Juxtaposition:	6
4.2 Abhängige Variablen	6
5.0 Hypothesen:.....	7
6.0 Die Konstruktion der experimentellen Bedingungen	7
6.1 Konstruktion der Daten:	7
6.2 Konstruktion der experimentellen Bedingungen:	8
6.2.1 Vergleichbarkeit der Anforderungen:	8
7.0 Versuchspersonen und Versuchsdurchführung	9
8.0 Die speziellen Fragestellungen:	9
8.1 Vergleiche von Summen (bzw. Mittelwerten) von Datenreihen	9
8.1.1 Konstruktion der Graphen:	10
8.1.2 Ergebnisse	11
8.1.3 Unterschiede der Positionierung der Graphiken unter Juxtaposition.....	13
8.2 Vergleich von Variabilitäten (Fluktuationen) von Datenreihen.	13
8.2.1 Fluktuation als Variabilitätsmaß	14
8.2.2 Konstruktion der experimentellen Bedingungen.....	15
8.2.3 Ergebnisse	15
8.2.4 Unterschiede der Positionierung der Graphiken unter Juxtaposition.....	17
8.3 Größenvergleiche bei einer Rubrik.....	17
8.3.1 Konstruktion der experimentellen Bedingungen.....	18
8.3.2 Ergebnisse:	18
8.3.3 Positionierung der Graphiken unter Juxtaposition	
8.4 Die Schätzung von Größenwerten (point reading)	20
8.4.1 Konstruktion der experimentellen Bedingungen:	20
8.4.2 Untersuchungsvorgehen:.....	21
8.4.3 Datenauswertung	22
8.4.4 Ergebnisse	22
8.4.4.1 Fragestellung: Wie groß ist der Größenwert einer Rubrik?.....	22

8.4.4.2 Fragestellung: Wie groß ist bei der Datenreihe x der Größenwert einer vorgegebenen Rubrik	23
9.0 Zusammenfassende Schlußfolgerungen:	24
9.1 Wahrnehmung von Summen bzw. Mittelwerten.	24
9.2 Wahrnehmung von Variabilität	25
9.3 Schlußfolgerung zu Summen -und Variabilitätsvergleichen.	25
9.4 Replikation des Experimentes von Schutz 1961b	26
Literatur	26
Glossar	28

Globale Vergleiche, lokale Vergleiche und Größenschätzungen in Liniendiagramm und Säulendiagramm unter Superposition und Juxtaposition.

Zusammenfassung:

In Fortführung des von Jacobs 1994 begonnenen Projektes standen in dieser Untersuchung weitere Fragestellungen, die mit Hilfe graphischer Präsentationen beantwortet werden sollten. Als Graphiktypen dienten Säulendiagramm und Liniendiagramm. Beide Graphiktypen wurden in der Graphanordnung Superposition und Juxtaposition präsentiert. Innerhalb von Juxtaposition wurde zum Teil horizontale gegen vertikale Graphanordnung getestet. Als weiterer Faktor wurde die Anzahl der Datenreihen in einer Präsentation variiert. Im Mittelpunkt der Untersuchung stand die Frage, welche dieser Graphikvarianten sich für globale Vergleiche, insbesondere zur Einschätzung von Summen bzw. Mittelwerten und Variabilitäten von Datenreihen am ehesten eignen. In Ergänzung dazu wurden Teile aus dem Experiment von Schutz (1961b) einem Replikationsversuch unterzogen und um den Graphiktyp Säulendiagramm ergänzt.

Ein komplexes Computerprogramm steuerte den Versuchsablauf. Das Programm generierte nach ganz bestimmten Regeln die Daten für die einzelnen experimentellen Bedingungen und konnte so eine Vielfalt unterschiedliche Datenkonstellationen zur Verfügung stellen. 25 studentische Versuchspersonen nahmen honoralfrei am Experiment teil.

Für den Vergleich von Summen bzw. Mittelwerten von Datenreihen ist das Säulendiagramm dem Liniendiagramm unter der Graphanordnung Juxtaposition durchgängig eindeutig überlegen, unter der Graphanordnung Superposition hingegen erzielt das Liniendiagramm mit zunehmender Anzahl der Datenreihen Vorteile gegenüber dem Säulendiagramm (=Interaktion zwischen Graphtyp und Graphanordnung). Beim Vergleich von Variabilitäten ist das Liniendiagramm dem Säulendiagramm klar überlegen. Das Experiment von Schutz konnte ziemlich gut repliziert werden. Wenn es um Größenvergleiche von Datenreihen innerhalb einer bestimmten Rubrik geht, spielt der Graphtyp keine Rolle, jedoch ist der Vorteil von Superposition gegenüber Juxtaposition eindeutig und erzielt eine klare praktische Bedeutsamkeit. In allen Graphikvarianten können die Größenwerte recht gut eingeschätzt werden. Die entsprechenden Unterschiede sind vernachlässigbar.

Schlagworte: Präsentationsmodi, Diagramme, Charts, Graphs, Säulendiagramm, Liniendiagramm, bar chart, line graph, graphical displays, graphical perception

1.0 Einleitung und Einstieg in das Thema

Ein wesentlicher Vorteil der graphischen Präsentation von Daten gegenüber einer Tabelle liegt im Vermögen der Wahrnehmung mehrere Daten sehr schnell zu einer Einheit zusammenzufassen und auf dieser Basis wesentlich leichter Vergleiche zwischen Datengruppen durchzuführen (Jacobs 1990). Neben den einzelnen Größenwerten bekommt man auf einen Blick einen ungefähren Eindruck vom Durchschnitt und der Variabilität der Daten. So kann man mit Hilfe der Graphik aus den Einzeldaten angenäherte statistische Kennwerte z.B. Mittelwerte oder Streuungen miteinander vergleichen. Derartige Abschätzungen sind z.B. relevant bei der schnellen Überprüfung der Fragen, ob sich die Stufen eines Faktors in einer mehrfaktoriellen VA bedeutsam voneinander unterscheiden oder ob klare Schwankungsunterschiede im Aktienkursverlauf an verschiedenen Börsenplätzen auszumachen sind.

Nach Carswell (1992) sind dies Fragen nach globalen Vergleichen, nach Bertin (1974) Fragen auf der mittleren oder gar höchsten Stufe des Erfassens. Der Klassifikation von Wainer (1992) bzw. Schnotz (1994) zufolge könnte man die Fragestellungen sicherlich auf mindestens der 2., vielleicht aber eher noch auf der 3. Ordnungsstufe der Gewinnung von Informationen einordnen. Theoretisch interessant ist das Problem, wie Einzeldaten so dargestellt werden können, daß einerseits die Wahrnehmung der Zugehörigkeit zu einer bestimmten Gruppe möglichst einfach, andererseits die Abschätzung der interessierenden Gruppeneigenschaft besonders gut gelingt, so daß die Gruppenunterschiede direkt ins Auge springen. Segmentierte Graphen (z.B.: divided bar chart), bei denen die Einzelwerte aufeinander gestapelt werden und die Gruppen etwa nebeneinander gestellt werden könnten, scheinen zumindest für die Abschätzung von "größte Summe ?" Vorteile zu versprechen, erschweren so aber den Vergleich der Einzelwerte. Deshalb werden diese Graphtypen hier nicht verwendet. Es geht vielmehr darum, festzustellen, wie sich derartige Kennwerte bei Präsentationen wahrnehmen lassen, die zugleich auch eine sehr gute Abschätzung der Einzeldaten erlauben.

Vorliegende Untersuchung ergänzt die Studien von Jacobs (1994) und (1995) um weitere Fragestellungen. Wie dort werden auch hier als Graphiktypen Säulendiagramm und Liniendiagramm in Superposition und Juxtaposition analysiert. Um Wiederholungen zu vermeiden, wird gelegentlich auf Jacobs (1994 bzw. 1995) verwiesen.

Die Fragestellungen erfordern nicht immer, wie bei den vorherigen Studien, eine mindestens ordinal skalierte x-Achse, sind aber auch für Daten relevant, die Verlaufsinformationen darstellen sollen. Die Vorteile einer speziellen graphischen Präsentation sollen dabei primär durch günstigere Wahrnehmung zustande kommen. Die kognitiven Grundvoraussetzungen beim Lesen einer Graphik (die extrinsische und intrinsische Identifizierung nach Bertin (1974, S.17)) werden als gegeben vorausgesetzt. Es ist darüber hinaus auch gänzlich irrelevant, ob ein naiver Leser bei einem Liniendiagramm direkt einen Verlauf antizipiert. Würde der Leser spontan ein Verlaufsschema aktivieren, dann könnte man ihm kaum ein Liniendiagramm zumuten, wenn keine Verlaufsinformationen zugrunde lägen. Es geht hier aber nicht um die Erfassung bestehender Schemata, sondern um Wahrnehmungsleistungen auf der Basis eines vorher explizierten Schemas. Es könnte ja sein, daß man in einem Liniendiagramm bestimmte Phänomene besser erkennt als in einem Säulendiagramm, obwohl es sich dabei nicht um einen Verlauf handelt. Möglicherweise zeigen sich hier aber auch Vorteile für das Säulendiagramm, das in den früheren Experimenten zu speziellen Verlaufsinformationen dem Liniendiagramm unterlegen war.

2.0 Bisherige empirische Forschung

Carswell (1992, S. 543) listet mehrere Studien auf, in denen verschiedene graphische Präsentationen im Hinblick auf globale Vergleiche getestet wurden. Ein direkter Vergleich Säulendiagramm vs. Liniendiagramm wurde dabei jedoch sehr selten durchgeführt. Casali und Gaylin (1988) fanden keine Unterschiede zwischen Liniendiagramm und Säulendiagramm bei einer Fragestellung, die den Vergleich zweier oder mehrerer Größenwerte im Hinblick auf die Relation größer oder kleiner zum Gegenstand hatte (point comparison nach Casali & Gaylin). Mir ist keine Studie bekannt, bei der ganze Datenreihen miteinander verglichen werden mußten, oder welche die Frage nach Variabilitätsunterschieden in einer Graphikpräsentation überprüft hätte.

Schutz (1961b) überprüfte Zeit und Genauigkeitsunterschiede zwischen Superposition und Juxtaposition im Liniendiagramm bei 2 Fragestellungen. Bei der Schätzung von Größenwerten (point reading) wurden keine Unterschiede zwischen beiden Graphanordnungen festgestellt. Superlinie erbrachte jedoch klare Zeitvorteile gegenüber Juxtaline, wenn entschieden werden sollte, welche Datenreihe bei einem bestimmten Meßzeitpunkt den höchsten Wert erzielte (point comparison). Beide Fragestellungen sollen hier einem Replikationsversuch unterzogen werden und darüber hinaus noch überprüft werden, ob das Säulendiagramm für derartige Fragestellungen mehr oder weniger geeigneter ist.

3.0 Die Fragestellungen der Untersuchung

Wie bereits oben erwähnt, standen im Mittelpunkt der Untersuchung Fragestellungen zur Wahrnehmung eines Vergleiches von Summen bzw. Variabilitäten von Datenreihen. Dazu mußten bestimmte Fragen beantwortet bzw. Aufforderungen nachgekommen werden, die nachfolgend in Form einer Übersicht entsprechend der für alle VPn konstanten Testungsabfolge im Experiment genannt werden:

- Welche Datenreihe (von 2, 4 oder 8 Datenreihen) weist die größte Summe bzw. den größten Mittelwert auf?
- Welche Datenreihe (von 2 oder 4 Datenreihen) weist die größte Variabilität auf ?

Zur Replikation und Erweiterung des Experimentes von Schutz (1961 b):

- Welche Datenreihe erzielt bei einer bestimmten Rubrik den höchsten Größenwert ?
- Schätze den Größenwert einer bestimmten Rubrik (bei 1 und bei 4 Datenreihen)

4.0 Die experimentellen Bedingungen dieser Untersuchung

Die experimentellen Graphen entsprechen in der Grundstruktur weitgehend den Graphen bei Jacobs (1994), aus denen man weitere Details und Bemerkungen dazu entnehmen kann. Deshalb folgt hier nur eine Kurzfassung:

4.1 Die unabhängigen Variablen:

Experimentell variiert wurden in jedem Fall der Graphtyp. Je nach Fragestellung wurde dann die Graphanordnung und in unterschiedlichem Ausmaß die Anzahl der Datenreihen manipuliert.:

Unabhängige Variable: Variationsmöglichkeiten

Graphtyp: Liniendiagramm, Säulendiagramm

Graphanordnungen: Superposition, Juxtaposition

Anzahl der Datenreihen: 1,2,4,8

Da Juxtaposition mindestens 2 Datenreihen voraussetzt, liegen im Falle einer Datenreihe stets konventionelles Linien- bzw. Säulendiagramme zugrunde.

4.1.1 Variationen der Graphpositionierung unter Juxtaposition:

Als geeignete Anordnungen der Graphiken unter Juxtaposition bieten sich an:

- a) horizontale Positionierung (Graphiken nebeneinander)
- b) vertikale Positionierung (Graphiken untereinander).

Diese Anordnungsvarianten scheinen willkürlicher oder sonstwie regelmäßiger Anordnung (wie. z.B. kreisförmig oder diagonal) überlegen zu sein. Bei manchen Fragestellungen wurden daher diese Varianten gegeneinander getestet, jedoch bis zu höchstens 4 Datenreihen. Beim Vergleich von Juxtaposition gegen Superposition werden die Daten aus beiden Positionierungsarten unter Juxtaposition zur Testung verwendet.

Bei 8 Datenreihen lagen die Diagramme in 2 Viererzeilen untereinander. Die Größe einer Graphikvariante umfaßte beinahe den gesamten Bildschirm. Die superpositionierten Diagramme waren deutlich größer als die Einzeldiagramme unter Juxtaposition. Die Achsenverhältnisse aller Diagramme wurden jedoch gleich gehalten.

Jede Datenreihe umfaßte 10 oder 12 Datenelemente. Um die einzelnen Datenreihen voneinander abzugrenzen wurde die Farbe gewählt, die in der Literatur ziemlich übereinstimmend als eine der besten Codierungsarten für mehrere Datenreihen gilt (siehe dazu Jacobs (1995)).

4.2 Abhängige Variablen

Die Messung der abhängigen Variablen:

- a) Zeit bis zur Beantwortung der Frage (Zeit)
- b) Prozentsatz der richtigen Antworten (Genauigkeit)

Entscheidende abhängige Variable ist die Zeit von der Darbietung der Präsentation bis zur Entscheidung der VP, die Frage als beantwortet zu betrachten. Die VP sollte dabei so schnell wie möglich, aber dennoch genau, die Frage beantworten. Der gesamte Ablauf wird von der VP selbst kontrolliert:

Sie liest die spezielle Fragestellung ohne Zeitbeschränkung durch, bestimmt durch Tastendruck (Leertaste = Zeitanfang) den Beginn der Graphikpräsentation und beendet durch erneuten Tastendruck (Leertaste = Zeitende) die Reizdarbietung. Anschließend gibt sie die Antwort ein. Dies geschieht

entweder in Form vom Multiple Choice durch Mausklick auf eine der Alternativen oder bei den Größenschätzungen durch Eingabe einer Zahl. Der erste Tastendruck initialisiert die Zeitmessung, der zweite Tastendruck beendet die Zeitmessung.

Als Genauigkeit wird der Prozentsatz der korrekten Antworten erfaßt und verrechnet. Eine Vergleichbarkeit über verschiedene Fragestellungen ist nur eingeschränkt möglich, da die Alternativenanzahl der MC-Antworten schwankt. Die experimentellen Bedingungen umfassen in der Regel mehrere Graphikpräsentationen, die dann zur Messung der Bedingungen gemittelt wurden.

Beide abhängige Variablen sollen Vergleichbares in dem Sinne messen, daß mit wachsendem Zeitbedarf auch die Ungenauigkeit anwächst. Um diese erwartete Gleichsinnigkeit beider Meßvariablen zu veranschaulichen, wird deshalb bei der graphischen Darstellung der Ergebnisse die Genauigkeit als Prozentsatz der Fehler angegeben. Eine Graphikvariante ist einer anderen nur dann überlegen, wenn Sie schnellere Beantwortung bei mindestens gleicher Genauigkeit unter Beweis stellt.

Bei der Datenauswertung wurden Ausreißer (extrem lange Zeitspannen) auf Itemniveau, ähnlich wie bei Jacobs (1994), aus der Verrechnung entfernt. Auf diese Weise wurden je nach Fragestellung ca. 2 bis 5 % der Daten eliminiert.

5.0 Hypothesen:

Hinsichtlich der zentralen Untersuchungsfragen zum Vergleich vom Mittelwerten und Variabilitäten von Datenreihen lagen keine klar begründbaren gerichteten Hypothesen vor, wodurch die Testung einen explorativen Charakter annimmt. Beim Replikationsversuch der Studie von Schutz wurde eine erfolgreiche Replikation seiner Ergebnisse für das Liniendiagramm erwartet, aber bestimmte Vorteile für das Säulendiagramm vermutet, die später genauer begründet werden.

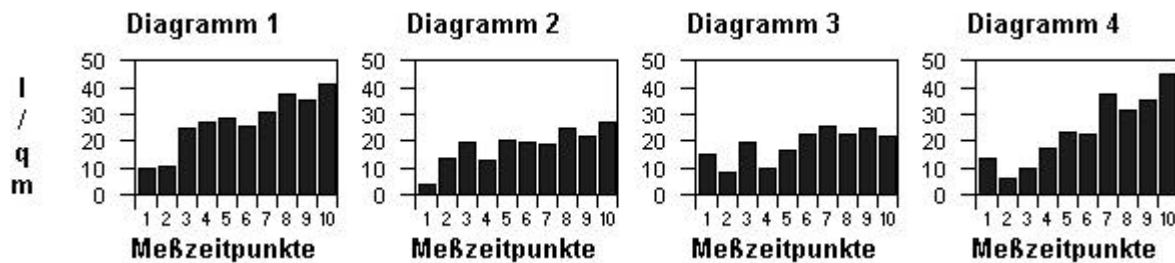
6.0 Die Konstruktion der experimentellen Bedingungen

6.1 Konstruktion der Daten:

Jacobs (1989) und (1990) wählte als Daten für seine Graphiken bei vielen Fragestellungen reine Zufallswerte. Dieses programmtechnisch einfache Konstruktionsprinzip basiert auf einer sehr pessimistischen Einschätzung des wissenschaftlichen Wertes einer Datenerhebung, da so letztlich nur Chaos erfaßt wird, und selten eine wie auch immer geartete vernünftige Gesamtstruktur zu sehen ist. Hier wurde daher eine anderes Verfahren zur Erzeugung künstlicher Daten gewählt:

Eine Datenreihe setzt sich wie folgt zusammen: Zunächst wurden die Elemente einer Datenreihe mittels einer mathematischen Funktion bestimmt. Somit entstanden ideale Funktionsverläufe, ähnlich wie sie bei Jacobs (1994,1995) verwandt wurden. Da diese Funktionen aber bereits bestimmten Variationen z.B. hinsichtlich Steigung und y-Achsenabschnitt aufwiesen, führte dieser erste Datengenerierungsprozeß bei derselben Funktion bereits zu verschiedenen Daten. In einem zweiten Schritt wurde nun jeder Funktionswert (das ideale Datum) durch eine weitere Zufallsprozedur in Form einer variablen Abweichung nach oben oder unten modifiziert. Auf diese Weise entstehen Datenreihen, denen man irgendeine Art von Verlauf ansehen könnte, wenn man sich die Zufallsstreuung als herauspartialisiert denkt. Diese Datenreihen sind quasi konstruktive wissenschaftliche Graphen, denen angenäherte Strukturen zugrunde liegen, die lediglich "infolge der Unbilden der Empirie" entsprechend verwässert sind.

Abbildung 1: Beispiel für die Datengenerierung basierend auf einer linearen positiven Funktion:



Für jede Graphik wurde dieser Datengenerierungsprozess aktiviert. Bereits bei einer bestimmten zugrunde liegenden Funktion erhält man unterschiedliche Daten (siehe Abbildung 1). Diese sind aber im Durchschnitt von der Struktur her zum Teil ähnlich. So führt die zugrunde liegende lineare positive Funktion dazu, daß die Datenreihe in der Regel im wesentlichen ansteigt. Um grundsätzlich unterschiedliche Datenkonstellationen zu erhalten, wurden verschiedene Funktionen zugrunde gelegt. Zugrunde liegende Sinus - Funktionen bringen irgendeine Art von Wellenbewegung in die Graphik hinein und unterscheiden sich somit im Durchschnitt von zugrunde liegenden linear-positiven Funktionen. Die Auswahl der Funktionen zur Datenerzeugung ist relevant für die Parallelisierung von Aufgaben und die Repräsentativität der Testung.

6.2 Konstruktion der experimentellen Bedingungen:

Das Verfahren entspricht im Prinzip weitgehend dem von Jacobs (1995). Zum genaueren Verständnis wird eigens darauf verwiesen.

Nach bestimmten Regeln wurde zunächst festgelegt, welche Teilmenge von zugrunde liegenden Funktionen für alle experimentellen Bedingungen ausgewählt werden sollen. Im Gegensatz zu üblichen Gruppenexperimenten erhielten die VPn für identische experimentelle Bedingungen nicht dasselbe Reizmaterial bzgl. der Daten in den Graphiken. Für die einzelnen experimentellen Bedingungen wurden parallele Aufgaben entwickelt. So wurde z.B. sichergestellt, daß alle experimentellen Graphen vergleichbaren zugrunde liegenden Funktionen unterlagen. Die konkreten Daten dieser Funktionen konnten aber wie oben beschrieben variieren. Zudem wurden weitere variable Teile eingebaut, um Erinnerungseffekte weitgehend auszuschalten. So kann etwa die Farbe der Datenreihe oder die Nummer des Diagramms, das es zu entdecken galt, oder die einzuschätzende Rubrik usw. für die parallelen Aufgaben variieren.

Die Konstruktion von parallelen Aufgaben bezieht sich jeweils nur auf eine VP, so daß innerhalb einer VP Vergleichbarkeit der experimentellen Bedingungen durch parallele Aufgaben gewährleistet wurde. VP 2 bearbeitete aber teilweise strukturell andere Datenkonstellationen als VP 1 bei derselben experimentellen Bedingung (z.B. andere Funktionstypen). Auf diese Weise sollte die externe Validität deutlich gesteigert werden, weil im Gesamtexperiment so sehr viele variable Datenkonstellationen getestet werden konnten.

6.2.1 Vergleichbarkeit der Anforderungen:

Alle Fragestellungen dieser Untersuchung erfordern bestimmte Schätzleistungen, die aus experimentellen Gründen in gewisser Weise standardiert wurden, um einerseits die Anforderungen über die experimentellen Bedingungen annähernd gleich zu halten und um andererseits zu triviale Einschätzungen zu verhindern. So wurde etwa festgelegt, daß die Datenreihe mit dem höchsten Mittelwert einen um genau

12,5% höheren Mittelwert aufweisen mußte als der Mittelwert der Datenreihe mit dem zweithöchsten Mittelwert. Die in diesem Zusammenhang zu nennenden Prozentsätze beziehen sich auf jeweils zugrunde liegende Ordinate. Die gesamte Ordinate entspricht 100 %. Um Unterschiede zwischen den einzelnen Graphikvarianten nachweisen zu können, dürfen die Aufgaben nicht zu leicht ausfallen. Andererseits sollten die Aufgaben dennoch eindeutig lösbar sein bzw. die Wahrnehmung nicht über Gebühr strapazieren. Dann sollte auch noch eine große Palette möglicher Aufgaben mit unterschiedlichen Datenkonstellationen getestet werden. 12,5 % Unterschiedlichkeit kann dabei je nach Datenkonstellation mehr oder weniger schwierig sein, was durch Randomisierung jedoch ausgeglichen werden kann. Die kritischen Unterschiede wurden durch Selbstversuche zu ermitteln versucht. Es ist klar, daß ein derartiges Verfahren keine optimalen Resultate liefern kann und gewisse Fehler nicht auszuschließen sind gerade dann, wenn man das Ziel "teilweise anspruchsvoll und zugleich eindeutig" anstrebt. Wie gut die Prozedur letztlich gelang, läßt sich zum Teil aus den Genauigkeitsdaten schätzen. Dieses Konstruktionsprinzip ist bei der Interpretation der Befunde zu berücksichtigen. Die Ergebnisse lassen sich nicht bedenkenlos generalisieren.

7.0 Versuchspersonen und Versuchsdurchführung

An der Untersuchung nahmen 25 VPn unentgeltlich teil. Die meisten davon waren Student(inn)en der Erziehungswissenschaft, die im SS 1994 ein Seminar in Statistik 2 belegten. Die übrigen VPn rekrutierten sich aus sonstigen, vom Verfasser angesprochenen Studenten und Mitarbeiter der Universität. Die Statistikteilnehmer absolvierten das Experiment im CIP-Raum der Philosophischen Fakultät in Gruppen von 3 bis 8, jeder für sich an einem eigenen Computer. Die restlichen VPn bearbeiteten das Experiment allein am eigenen Computer oder am Computer des Verfassers. Bei einer Testserie zu Summenvergleichen hängte sich in einigen Fällen das Programm aus bisher ungeklärten Gründen auf, (da half nur noch Warmstart) was letztlich zu einem VPn-Schwund bei dieser Fragestellung führte.

Vor der eigentlichen Untersuchung hatten die VPn ein kleines Informationsprogramm durchzuarbeiten. Dadurch sollte sichergestellt werden, daß die VPn Säulen- und Liniendiagramme in Superposition und Juxtaposition verstehen und interpretieren sowie den Versuchsinstruktionen hinreichend folgen können.

Das Gesamtexperiment gliedert sich in mehrere experimentelle Einheiten, die nacheinander getestet wurden. Eine experimentelle Einheit (=Testserie) umfaßt folgende Bestandteile: genaue Instruktion der Fragestellung, Beispiellösung(en) für die Fragestellung, mindestens 2 konkrete Beispielaufgaben, die experimentellen Aufgaben. Wenn z.B. mehrere Datenreihen getestet werden, so bildet jede Stufe des Faktors "Anzahl der Datenreihen" eine experimentelle Einheit. Eine Fragestellung wie z.B. "Finde die Datenreihe mit der größten Summe aus all ihren Größenwerten" umfaßt 3 Testserien, weil sie für 2, 4 und 8 Datenreihen überprüft wurde. Innerhalb jeder Testserie wurden alle Präsentationen in zufälliger Anordnung dargeboten.

8.0 Die speziellen Fragestellungen:

8.1 Vergleiche von Summen (bzw. Mittelwerten) von Datenreihen

Neben dem schnellen Erkennen von Verlaufsinformationen ist die Wahrnehmung von Datengruppen ein Vorteil graphischer Präsentationen. In der Untersuchung von Jacobs (1990) konnte klar belegt werden, daß der Vergleich von 2 Gruppen, die jeweils aus mehreren Rubriken bestehen, in einem Säulendiagramm bei vergleichbarer Genauigkeit schneller vollzogen wird als in einer Tabelle. Der Präsentationsunterschied wuchs mit steigender Anzahl der zu einer Gruppe gehörenden Elemente. Zur Wahrnehmung einer Summe oder eines Mittelwerts einer Datenreihe müssen alle Einzeldaten visuell zusammengefaßt

werden. Beim Summenvergleich in einem Säulendiagramm wird dann ein flächenartiges Gebilde mehrerer Säulen der einen Gruppe gegen ein entsprechendes Gebilde der zweiten Gruppe abgeschätzt. Es ist evident, daß ein derartiger Vergleich bei geringen Unterschieden der zu vergleichenden Gruppen sehr schwierig ist, weil die Wahrnehmung zweidimensionale und ansonsten recht regellose Gebilde nicht sehr gut nach ihrer "Flächengröße" differenzieren kann. Schon Croxton und Stein (1932) haben nachgewiesen, daß Quadrate (zweidimensional variiert) sich schlechter differenzieren lassen als Rechtecke (bars, die nur in der Höhe variieren). Quadrate haben aber immerhin noch eine regelmäßige Form.

Bei hinreichend großen Unterschieden zwischen den Gruppen ist jedoch davon auszugehen, daß der Größenvergleich zweier Datenreihen spontan sehr schnell gelingt. Mithin sind Graphiken insbesondere für die schnelle grobe Orientierung geeignet. Sieht man den Unterschied, hat man den Eindruck eines Unterschieds und dann scheint auch ein klarer Unterschied vorzuliegen, sieht man ihn nicht oder nur schwer, dann ist der Unterschied offenbar nicht groß und folglich unbedeutend. Ob man den Unterschied sieht oder nicht bzw. sehen soll oder nicht, hängt zu einem großen Teil auch von der Wahl der Ordinatenbegrenzungen ab. Hier liegt ein unerschöpfliches Tummelfeld, durch Stauchen oder Dehnen mit Graphiken zu lügen. Andererseits liegt hier ein echtes Problem. Bedeutsame Unterschiede muß man auch deutlich sehen können und gegebenenfalls muß man dann den Skalenbereich der Ordinate modifizieren (siehe z.B. Kosslyn (1994, S. 79)). Allerdings sollte dies in fairer Weise geschehen.

8.1.1 Konstruktion der Graphen:

Die Datenreihe mit dem größten Mittelwert bzw. der größten Summe wird hier Zieldatenreihe bzw. Zielverlauf genannt, die übrigen Datenreihen werden Distraktorverläufe genannt. Die Datenreihen in der Präsentation wurden so konstruiert, daß die Zieldatenreihe einen 12,5% höheren Mittelwert aufwies als die Datenreihe mit dem zweithöchsten Mittelwert. Der Unterschied von 12,5% wurde durch Selbstversuche des Verfassers ermittelt und als geeignet angesehen, da eine Entscheidung in den meisten Fällen anspruchsvoll aber zugleich noch weitgehend eindeutig erschien. Alle Datenreihen in Aufgabe 1 basierten auf ansteigenden Funktionen und in Aufgabe 2 auf abfallenden Funktionen. In Aufgabe 3 war die Hälfte der Datenreihen ansteigend, die andere Hälfte abfallend und in Aufgabe 4 wurde jede Datenreihe aus einer Menge komplexer (tritonier oder quadritonier) Verläufe nach Zufall gezogen. Es war angenommen worden, daß die Aufgaben unterschiedlich schwierig sind. Beispielsweise wurde davon ausgegangen, daß vier ansteigende Datenreihen leichter im Mittelwert verglichen werden konnten als vier komplexe Funktionen. Wie eine spätere Analyse der Zeiten jedoch überraschenderweise ergab, konnten keine überzeugenden, konsistenten und hinreichend großen Unterschiede zwischen den einzelnen Aufgaben statistisch nachgewiesen werden. Dies kann zum Teil auch daran liegen, daß die Variationen innerhalb eines Funktionstyps bereits sehr groß waren.

8.1.2 Ergebnisse

Tabelle 1: Dreifaktorielle Varianzanalyse mit Meßwiederholung auf allen Faktoren zur Fragestellung: **Finde die Datenreihe mit größter Summe (höchstem Mittelwert)!**

UV: Anzahl der Datenreihen (2,4,8)
 Graphanordnung (Superposition, Juxtaposition)
 Graphtyp (Liniendiagramm, Säulendiagramm)

AV: Genauigkeit (arcsin-transformierte Testwerte ;arsin(% richtig))

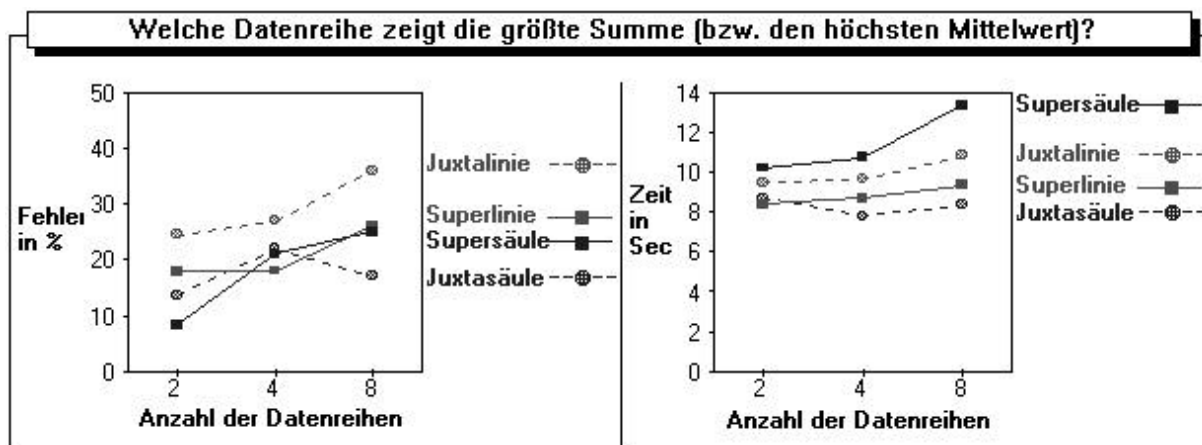
Quelle der Variation	MS	df	F	p für F	Sign-N.
Graphanordnung	1,83	1,20	10,32	,004	p<.01
Graphtyp	1,32	1,20	9,0	,007	p<.01

AV: Zeit (logarithmierte Werte; Testwerte = ln(Zeit))

Quelle der Variation	MS	df	F	p für F	Sign-N.
Graphanordnung	0,47	1,20	9,19	,007	p<.01
Graphanordnung x Graphtyp	3,22	1,20	59,28	,000	p<.01

Die Tabelle 1 beschränkt sich auf Ergebnisexplikationen signifikanter Faktoren. Zunächst fällt der insignifikante Effekt für den Faktor "Anzahl der Datenreihen" bei Zeit und Genauigkeit auf. Neben Trainingseffekten ist dafür offensichtlich das Konstruktionsprinzip der Graphiken verantwortlich. Der Mittelwert des Zielverlaufs war stets 12,5% höher als der Mittelwert der zweithöchsten Datenreihe. Offensichtlich war der Größenvergleich der beiden ernsthaft in Frage kommenden Alternativen das schwierigste Problem und die Störungen und Verwirrungen durch weitere Datenreihen weniger beeinträchtigend. Der unbedeutende Effekt von "Anzahl der Datenreihen" rechtfertigt die Zusammenfassung (Mittelwertsbildung) der Daten über alle Stufen dieses Faktors.

Abbildung 2: Ergebnisse zur Fragestellung:



Die signifikante Interaktion zwischen Graphanordnung und Graphtyp bei der Zeit besagt, daß innerhalb des Säulendiagramms Juxtaposition und innerhalb des Liniendiagramms Superposition die günstigeren

Zeitwerte erzielen. Diese Interaktion deutet sich auch bei der Genauigkeit an, erreicht aber nicht das erforderliche Signifikanzniveau. Eine verständliche Interpretation der Effekte zur Genauigkeit geht davon aus, alle Graphvarianten bis auf die klar ungenauere Juxtalinie als annähernd gleich genau einzuschätzen.

Tabelle 2: Mittelwerte und (Standardabweichungen) aller experimentellen Bedingungen zur Fragestellung: Finde die Datenreihe mit der höchsten Summe (bzw. dem höchsten Mittelwert) ! (N=21)

Anzahl der Daten- reihen	Prozentsatz der korrekten Lösungen				Zeit in Sekunden			
	Juxta- linie	Super- linie	Juxta- säule	Super- säule	Juxta- linie	Super- linie	Juxta- säule	Super- säule
2	76 (16)	82 (23)	86 (10)	92 (14)	9.4 (4.6)	8.4 (4.6)	8.6 (4.4)	10.2 (5.9)
4	72 (16)	82 (24)	78 (16)	80 (21)	9.3 (5.7)	8.5 (4.5)	7.7 (3.7)	11.9 (5.8)
8	63 (22)	75 (25)	83 (18)	75 (26)	10.7 (5.6)	9.0 (5.3)	8.1 (3.7)	13.2 (6.9)
Im Mittel:	70	80	83	83	9.8	8.6	8.1	11.5
*	(11)	(19)	(13)	(8)	(4.1)	(3.8)	(3.4)	(5.1)

Im Mittel = Durchschnitt aus 2, 4 und 8 Datenreihen:

Aus der Abbildung 2 geht klar hervor, daß Juxtalinie höhere Fehlerprozentage aufweist als die übrigen Graphikvarianten. Dies wird auch bei der Zusammenfassung aller Genauigkeitswerte über alle Datenreihen (siehe 'Im Mittel' in Tabelle 2) deutlich. Um dies statistisch zu überprüfen, wurden der Einfachheit halber die entsprechenden Graphvariantenmittelwerte aus allen Datenreihen mittels t-Test jeweils gegen Juxtalinie getestet. Auf einem für die 3 durchgeführten Vergleiche Alpha-adjustierten Signifikanzniveau von 5 % schneidet Juxtalinie gegenüber allen anderen Graphikvarianten bedeutsam ungenauer ab.

Die Ergebnisse zur Entscheidungszeit zeigen ebenfalls Unterschiede zwischen den Graphvarianten. Einige interessante Vergleiche für die Mittelwerte aus allen Datenreihen wurden bei Alpha-adjustiertem 5% Niveau für die durchgeführten Vergleiche statistisch überprüft. Danach gibt es keine Unterschiede zwischen Superlinie und Juxtasäule ($t(19)=.95$). Supersäule erfordert statistisch signifikant mehr Zeit als Superlinie ($t(19)=4,07$) und mehr Zeit als Juxtasäule ($t(19)=5,8$). Juxtasäule benötigt signifikant weniger Zeit als Juxtalinie ($t(19)=4,17$).

Versucht man anhand beider abhängiger Variablen ein Gesamtresumee für die Fragestellung "Finde die Datenreihe mit der größten Summe oder dem höchsten Mittelwert", so sind Superlinie und Juxtasäule insgesamt die günstigsten Graphikvarianten und Juxtalinie der große Verlierer. Dies gilt für die hier durchgeführten Vergleiche. Allerdings sollte man bedenken, daß geringere Mittelwertsunterschiede in

manchen Fällen bei Superlinie vermutlich genauer und auch schneller erkannt werden können. Insbesondere ähnlich verlaufende Datenreihen, die sich nur geringfügig im Mittel unterscheiden, können wegen der räumlichen Nähe im Liniendiagramm noch sehr genau differenziert werden, überfordern hingegen die Unterschiedsdiskrimination bei Juxtaposition. Die sehr hohen Genauigkeitswerte von Supersäule bei 2 Datenreihen lassen diesen Graphiktyp für die untersuchte Fragestellung zumindest bei wenigen Datenreihen ebenfalls als echte Alternative einschätzen.

Würde es ausschließlich oder hauptsächlich um die Fragestellung gehen, welche von mehreren Datenreihen die höchste Summe erzielte, dann wäre offensichtlich ein segmentierter Graph die geeignete Präsentation. Bei 4 Datenreihen könnten man dann in einem divided bar chart 4 Säulen miteinander vergleichen und eine schnelle und genaue Entscheidung treffen. Im divided bar chart sind jedoch die Rubriken (hier die Meßzeitpunkte) dann aufeinander gestapelt und ein Vergleich der Rubriken erheblich erschwert, weil diese nur noch als Längen, aber nicht mehr als Positionen wahrgenommen werden könnten (Cleveland 1985). Verlaufsanalysen wären zudem quasi überhaupt nicht mehr durchzuführen. Wir haben beim segmentierten Graph den klassischen Fall, daß der Vorteil für eine Fragestellung zugleich einen gravierenden Nachteil für viele andere wichtige Fragestellungen nach sich zieht.

8.1.3 Unterschiede der Positionierung der Graphiken unter Juxtaposition.

Um zuverlässige Ergebnisse zum Effekt der Diagrammpositionierung innerhalb der Graphanordnung Juxtaposition zu erhalten, wurden die Werte für 2 und 4 Datenreihen zusammengefaßt. Die in Tabelle 3 dargestellten Mittelwerte basieren damit auf jeweils 8 Aufgaben.

Tabelle 3: Diagrammpositionierung innerhalb von Juxtasäule und Juxtalinie (Mittelwerte)

Graph- positionierung:	Fehler in Prozent		Zeit in Sekunden	
	horizontal	vertikal	horizontal	vertikal
Juxtasäule	17	20	8.5	7.8
Juxtalinie	20	32	8.3	10.4

Numerisch sehen die Befunde so aus, als ob die Positionierung der einzelnen Graphen für Juxtasäule gleichgültig ist und hauptsächlich bei Juxtalinie einen Effekt ausübt. Dies deutet sich auch statistisch in einer Wechselwirkung zwischen Graphtyp (Juxtasäule, Juxtalinie) und Graphpositionierung (untereinander, übereinander) teilweise an. Für die Genauigkeit fällt der entsprechende Prüfwert für die Interaktion mit $F(1,20) = 1,18$, $p = ,290$ klar insignifikant aus, bei der Zeit erreicht er mit $F(1,20) = 6,68$, $p = .018$ nicht das erforderliche Signifikanzniveau von 1%. Die Unterschiede der Graphpositionierung für Juxtalinie tendieren übereinstimmend in die Richtung "günstiger wenn die Diagramme nebeneinander stehen", der statistische Nachweis (jeweils t-Test mit transformierten Daten) ist aber mit $t(20) = 2,17$, $p = .043$ für die Genauigkeit und $t(20) = 2,02$, $p = .057$ für die Zeit recht dürftig und erreicht nicht das geforderte Signifikanzniveau von 1%.

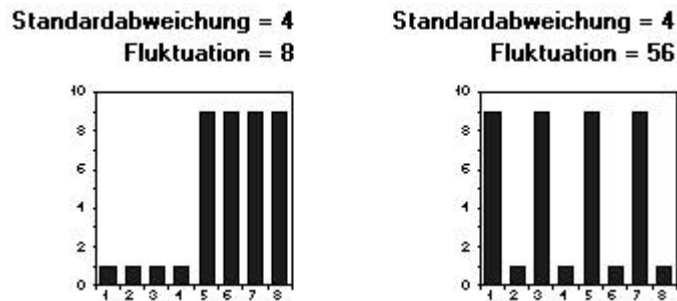
8.2 Vergleich von Variabilitäten (Fluktuationen) von Datenreihen.

Neben dem Mittelwert ist die Streuung ein wichtiges Maß für die zusammenfassende Beschreibung von Daten. Kann man so etwas wie die Variabilität von Werten in einer Graphik direkt sehen und Größenvergleiche hinsichtlich der Fluktuationen von Meßwerten vornehmen? Hierzu gehören z.B. ganz

typische Fragen wie "Gab es im Monat August in Deutschland mehr Temperaturschwankungen (über alle Tage hinweg) als in Schweden. Ein Hauptproblem bei der Überprüfung einer derartigen Fragestellung liegt im Finden eines geeigneten Maßes für Variabilität. Die Standardabweichung jedenfalls eignet sich dafür nicht, weil gleiche Streuung sich offenbar in ganz unterschiedlichen Strukturen niederschlagen kann, die kaum die Wahrnehmung von gleicher Veränderung suggerieren können, wie Abbildung 3 aufzeigt. Außerdem kann man von der Wahrnehmung nicht erwarten, daß sie Abweichungen vom Mittelwert quadratisch gewichtet, so daß als akzeptables Streuungsmaß S , wenn überhaupt höchstens

$S = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{X}|$ in Frage käme, welches aber die gleichen, wie in der Abbildung angedeuteten Probleme aufwirft.

Abbildung 3: Standardabweichung und Fluktuation anhand zweier Beispiele



8.2.1 Fluktuation als Variabilitätsmaß

Als Streuungsmaß sollte ein Kennwert gefunden werden, welcher die aufeinanderfolgenden Änderungen in einem Verlauf widerspiegelt. Wir nennen dieses Maß Fluktuation. **Die Fluktuation ist definiert als die Summe des Betrags der Abweichungen der Größenwerte von ihren jeweiligen Vorgänger-**

größenwerten. ($Fluktuation = \sum_{i=2}^n |x_i - x_{i-1}|$). Man hat mit Recht den Eindruck, daß die rechte Graphik

mehr Variabilität anzeigt als die linke Graphik, obwohl beide Graphiken die gleiche Standardabweichung aufweisen. Man würde die linke Graphik vielleicht eher als 2 stabile Zustände interpretieren, die in der Mitte im Niveau umgeschlagen sind. Simulationsversuche auf der Basis der Datengenerierungsregeln für die Konstruktion unserer experimentellen Graphen ergaben, daß die Fluktuation mit der Standardabweichung .50 ($n=1800$) korreliert.

Ziel des nachfolgenden Experimentes war es z.B. herauszubekommen, ob man obige Graphikpräsentation nicht besser in einer anderen Graphikvariante z.B. als Superlinie hätte zeigen sollen, um die Fluktuationsunterschiede beider Datenreihen zu verdeutlichen. Die Fluktuation ist vornehmlich im Zusammenhang mit Verläufen sinnvoll interpretierbar. Natürlich hätte man als grobes Streuungsmaß vielleicht auch die Spannweite wählen können und damit ein mehr verlaufsunabhängiges Variabilitätsmaß gefunden.

Die Fluktuationsunterschiede zweier Datenreihen sind sicher leichter wahrnehmbar, wenn regelmäßige Abweichungen (etwa Amplituden von Sinusschwankungen zweier Datenreihen oder Abweichungen von einer Referenzlinie) miteinander verglichen werden sollen. Dies wird aber hier nicht geprüft. Es stehen "empirisch realistische" Verläufe zur Testung an und dies erfordert sicher von den VPn einige Anstrengung, weil sich derartige Veränderungen leider häufig nicht zwingend spontan als einheitliche Strukturen wahrnehmen lassen.

In einem Liniendiagramm korreliert die Fluktuation vollkommen mit der Länge der Linie. So wurde den VPn die Fluktuation in der Instruktionsphase auch verdeutlicht. "Eine Graphik zeigt eine höhere Variabilität als eine andere, wenn sie insgesamt eine längere Linie aufweist." Beim Säulendiagramm wurden in der Instruktionsphase die Säulen mit Linien verbunden, um anhand eines Beispiels ganz anschaulich aufzuzeigen, welches von 2 Diagrammen die höhere Variabilität aufzeigt.

Es ist offensichtlich, daß sehr stark unterschiedlich fluktuierende Datenreihen trotz ihrer möglicherweise verschiedenen Strukturen dennoch schnell in der Wahrnehmung differenziert werden können. (siehe Abbildung 3). Um Unterschiede zwischen Graphikvarianten herauszubekommen, darf dieser Vergleich aber nicht zu trivial ausfallen. Durch eigenes Herumprobieren schien es dem Verfasser als anspruchsvolle Herausforderung für diese Fragestellung, den kritischen Unterschied zwischen der Datenreihe mit der höchsten und zweithöchsten Fluktuation auf Werte zwischen 20 und 30 % Unterschiedlichkeit festzulegen. Die Konstruktion eines ganz bestimmten Unterschieds (wie beim Summenvergleich) wäre programmtechnisch aufwendiger gewesen, weshalb darauf verzichtet wurde, obwohl die Bereichsvariation die interne Validität des Versuchsplans etwas schwächt.

8.2.2 Konstruktion der experimentellen Bedingungen

Ähnlich wie bei den anderen Versuchen wurden auch hier verschiedene Grundfunktionen mit Zufallsabweichungen herangezogen. So enthielt Aufgabe 1 nur ansteigende, Aufgabe 2 nur abfallende und Aufgabe 3 sehr komplexe Grundfunktionen. Welche der ansteigenden, abfallenden oder komplexen Grundfunktionen bei jeder VP zugrunde lagen, wurde durch Zufallsprozesse entschieden und war für alle experimentellen Bedingungen innerhalb einer VP parallelisiert. Erstaunlicherweise wurden zwischen den Aufgaben ähnlich wie bei Summenvergleich keine klaren konsistenten Leistungsunterschiede gefunden. Der Vergleich wurde für 2 und 4 Datenreihen durchgeführt. Jede dieser Testserien umfaßte 18 Präsentationen, jeweils 3 für insgesamt 6 experimentelle Bedingungen. Neben Supersäule und Superlinie wurden die Graphiktypen unter Juxtaposition in den Varianten horizontale und vertikale Positionierung getestet, wobei die Darbietungsreihenfolge der Präsentationen in jeder Testserie nach Zufall angeordnet wurde.

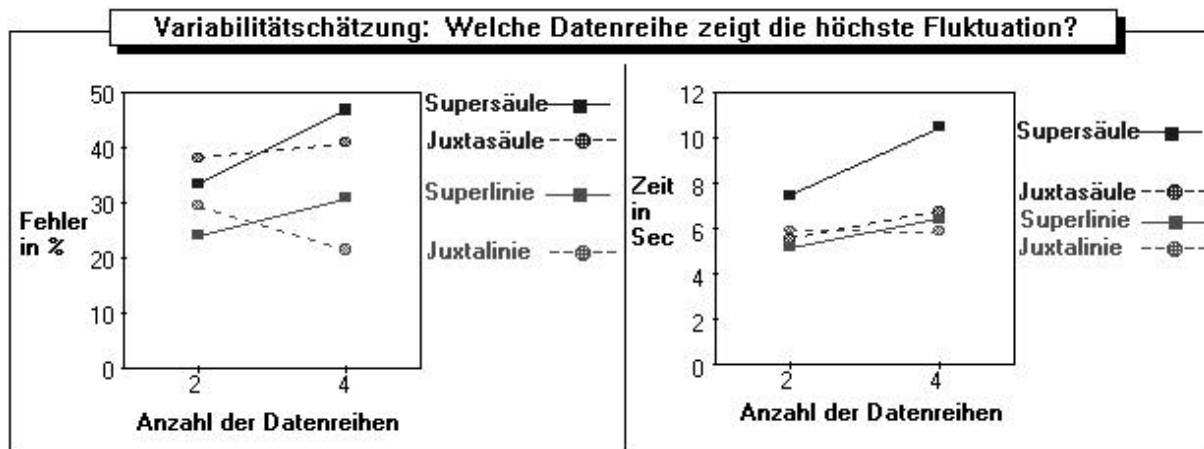
Aufgabe der VP war es stets entweder die Farbe oder die Diagrammnummer derjenigen Datenreihe zu finden, die insgesamt die höchste Fluktuation aufwies.

8.2.3 Ergebnisse

Die dreifaktorielle Varianzanalyse mit den Faktoren Anzahl der Datenreihen, Graphanordnung und Graphtyp mit der Genauigkeit als abhängiger Variablen (arcus-sinus transformierte Werte) ergab lediglich einen signifikanten Effekt, nämlich den Haupteffekt Graphtyp ($F(1,24) = 12,71$, $p = ,002$, $p < .01$), der, wie aus der Abbildung 4 hervorgeht, eindeutig zugunsten des Liniendiagramms geht.

Wie man aus der Tabelle 5 entnehmen kann, entsprechen die Genauigkeitsunterschiede zwischen den Graphiktypen einer Streuung von ca. 1/3 bis ca. einer Standardabweichung. Bei 2 und bei 4 Datenreihen ist Juxtaline stets genauer als Juxtasäule und Superlinie stets genauer als Supersäule. Auch wenn sich diese numerischen Einzelunterschiede nicht alle streng statistisch absichern lassen, so ist doch der Haupteffekt "Graphtyp" eindeutig. Die Aufgaben zu dieser Fragestellung waren recht schwierig und haben manche VPn möglicherweise überfordert, wenn man bedenkt, daß für 2 Datenreihen die Zufallswahrscheinlichkeit einer richtigen Beantwortung bei 50% liegt.

Abbildung 4: Fehlerprozent und Zeit (Mittelwerte für alle Bedingungen)



Die Ergebnisse zur Entscheidungszeit fallen etwas anders aus. Hier sieht es so aus, daß alle Varianten mit Ausnahme von Supersäule vergleichbar schnelle Zeiten aufweisen. In der Varianzanalyse (Tabelle 4) führt dies dann zu schwer interpretierbaren Haupteffekten und zu mehreren Wechselwirkungen, die umständlich nachzuvollziehen sind.

Tabelle 4: Varianzanalyse zur Fragestellung: Finde die Datenreihe mit höchster Variabilität!

UV: Anzahl der Datenreihen (2,4,)
 Graphanordnung (Superposition, Juxtaposition)
 Graphtyp (Liniendiagramm, Säulendiagramm)

AV: Zeit (logarithmierte Werte; Testwerte = $\ln(\text{Zeit})$) für

Quelle der Variation	MS	df	F	p für F	Sign-N.
A Anzahl der Datenreihen	4,38	1,23	38,54	,000	p<.01 *
B Graphanordnung	,88	1,23	10,24	,004	p<.01 *
C Graphtyp	2,44	1,23	25,94	,000	p<.01 *
A x B	,37	1,23	10,27	,004	p<.01 *
A x C	,38	1,23	3,67	,068	p>.01 ns
B x C	1,52	1,23	11,26	,003	p<.01 *
A x B x C	,00	1,23	,00	,974	p>.01 *

* *signifikant auf dem 1 % Niveau; ns = nicht signifikant*

Die Graphik in Abbildung 4 zeigt auf einen Blick das wesentliche Ergebnis. Supersäule benötigt stets mehr Zeit als die übrigen Diagrammvarianten. Juxtapos, Juxtaline und Superlinie unterscheiden sich bzgl. der Zeit nachweislich nicht signifikant voneinander, weder bei 2 noch bei 4 Datenreihen, wie zwei einfaktorielle VA mit diesen Graphvarianten als Stufen klar ergeben hat (p jeweils >.10).

Alle diese Graphikvarianten unterscheiden sich aber signifikant von Supersäule bei jeder Stufe des Faktors "Anzahl der Datenreihen" (p<.05, t-tests ohne Alphaadjustierung, p-Werte schwanken zwischen .014 und .000x).

Tabelle 5: Mittelwerte und (Standardabweichungen) aller experimentellen Bedingungen zur Fragestellung: Finde die Datenreihe mit der höchsten Fluktuation ! (N=25)

Anzahl der Daten- reihen	Prozentsatz der korrekten Lösungen				Zeit in Sekunden			
	Juxta- linie	Super- linie	Juxta- säule	Super- säule	Juxta- linie	Super- linie	Juxta- säule	Super- säule
2	71 (20)	76 (27)	62 (25)	67 (32)	5.9 (3.6)	5.1 (3.0)	5.5 (3.1)	7.4 (5.0)
4	79 (20)	69 (26)	59 (24)	53 (25)	5.9 (2.3)	6.4 (3.1)	6.7 (2.8)	10.4 (4.7)

Unter Berücksichtigung beider Meßvariablen geht die Schlußfolgerung aus diesem Experiment ziemlich klar in die Richtung, daß für Vergleiche von Variabilitäten in Datenreihen das Liniendiagramm besser geeignet ist als das Säulendiagramm und daß es keine Rolle spielt, ob die Datenreihen im Liniendiagramm in Superposition oder Juxtaposition angeordnet sind.

8.2.4 Unterschiede der Positionierung der Graphiken unter Juxtaposition.

Die Daten für 2 und 4 Datenreihen wurden zusammengefaßt, wodurch die Aufgabenanzahl für jede Juxtavariante auf 6 Präsentationen anstieg. Innerhalb eines Graphiktyps wurden für horizontale und vertikale Anordnung stets sehr hoch vergleichbare Ergebnisse sowohl für die Genauigkeit als auch für die Zeit gefunden. Sicherheitshalber durchgeführte T-Tests zwischen den einzelnen Positionierungsvarianten ergaben alle klar insignifikante Resultate (größter t-Wert überhaupt: $t = .37$ $p = .714$). Für den Variabilitätsvergleich von Datenreihen spielt es demnach keine Rolle, ob die Diagramme unter Juxtaposition übereinander oder untereinander angeordnet sind.

8.3 Größenvergleiche bei einer Rubrik.

Gelegentlich interessiert im Falle mehrerer Datenreihen, wie die Größenrelationen der Daten zu einem bestimmten Zeitpunkt aussehen. So will man etwa aus jährlichen Verlaufsanalysen verschiedener Autofirmen ablesen, welche Firma im Jahre 1993 die meisten Autos verkauft hat. Carswell (1992) subsumiert solche Fragen unter local comparison. Wenn sich der lokale Vergleich auf Elemente aus verschiedenen Datenreihen bezieht, ist die Graphanordnung Superposition aus mehreren Gründen Juxtaposition überlegen:

1. Werden die Diagramme unter Juxtaposition nebeneinander angeordnet, so muß gegenüber Superposition mehrmals die zu vergleichende Rubrik identifiziert werden und die zu vergleichenden Werte stehen weiter auseinander (Effekt von Distance nach Cleveland (1985)).

2. Werden die Diagramme unter Juxtaposition untereinander angeordnet, so können die zu vergleichenden Werte nicht anhand derselben gemeinsamen Skala verglichen werden. (siehe Clevelands Unterscheidung: Position along a common scale vs. position along identical nonaligned scales 1985, S. 232 ff). Zudem hat Superposition auch hier einen Proximity-Vorteil, weil die zu vergleichenden Werte in engerer räumlichen Beziehung stehen.

Die Überlegenheit von Superposition gegenüber Juxtaposition wurde in der Untersuchung von Schutz (1961b) für den Graphtyp Liniendiagramm klar demonstriert. Der Effekt konnte bei 2, 3 und 4 Datenreihen statistisch überzeugend nachgewiesen werden. Insofern war die Erwartung eines Vorteils von Superposition gegenüber Juxtaposition ganz eindeutig.

Aufgrund der Befunde von Jacobs (1990) wurde angenommen, daß der Graphtyp Säulendiagramm für diese Fragestellung dem Liniendiagramm überlegen ist. Er fand heraus, daß man in einem Säulendiagramm schneller als im Liniendiagramm entscheiden konnte, welcher Wert von 2 Datenreihen bei einer bestimmten Rubrik der größere ist (Graphanordnung Superposition). Der gleiche Graphtypunterschied wurde auch für Juxtaposition angenommen. Es erscheint plausibel davon auszugehen, daß Säulen insgesamt prägnanter und als Längen unmittelbarer wahrnehmbar sind als die Kreise im Liniendiagramm, für die quasi virtuelle Längen (Abstand Kreis zur Grundlinie) generiert werden müssen. Diese Argumentation gilt insbesondere für eine juxtapositionierte Anordnung, bei der die Graphen untereinander liegen. Die Ergebnisse von Jacobs (1990) widersprechen allerdings den Befunden von Gaylin & Casali (1988). Die Autoren untersuchten superpositionierte Graphtypen mit mehreren Datenreihen und fanden keinerlei Unterschiede bei Fragestellungen des point comparison, "which required evaluation of two oder more points of data to determine greater-than or lesser-than relationships".

8.3.1 Konstruktion der experimentellen Bedingungen.

Die Graphanordnung Juxtaposition wurde in den Versionen "Diagramme nebeneinander" und "Diagramme untereinander" getestet. Mithin waren in diesem Experiment 6 verschiedene Graphikvarianten beteiligt. Alle Graphikvarianten beinhalteten 4 Datenreihen und 8 Rubriken. Die 4 Datenreihen wiesen bei Aufgabe 1 einen überwiegend ansteigenden Verlauf, bei Aufgabe 2 einen überwiegend abfallenden Verlauf und bei Aufgabe 3 sehr komplexe Verläufe auf. Insgesamt wurden somit 18 Präsentationen zur Beurteilung vorgelegt. Nach Zufall wurde festgelegt, bei welcher Rubrik die Größenwerte der 4 Datenreihen verglichen werden sollten (= Zielrubrik). Alle experimentellen Bedingungen für eine VP waren hinsichtlich Grundfunktionen und Zielrubriken parallelisiert. Da die Zufallsprozesse bei der nächsten VP ebenfalls wirkten, führt dies dazu, daß in the long run eine repräsentative Stichprobe aus allen möglichen Rubriken getestet wurden. Es sollte stets die Datenreihe mit dem höchsten Größenwert auf der Zielrubrik identifiziert werden. Der höchste Größenwert wurde stets so festgelegt, daß er 10% höher als der zweithöchste Größenwert ausfiel. Wegen der Zufallsprozesse bei der Konstruktion der einzelnen Datenreihen, war auch die Farbe bzw. die Diagrammnummer mit der Zieldatenreihe hinreichenden Variationen ausgesetzt.

8.3.2 Ergebnisse:

Als Testwerte für die Graphiktypen unter Juxtaposition wurden die Ergebnisse beider Juxtapositionsversionen zunächst zusammengefaßt um einen Vergleich mit Superposition zu ermöglichen. Tabelle 6 zeigt die Ergebnisse.

Tabelle 6: Mittelwerte und (Standardabweichungen) zur Fragestellung: Welche von 4 Datenreihen erzielt zu einem bestimmten, vorgegebenen Zeitpunkt den höchsten Wert?; N=25)

	Säule	Linie
Genauigkeit in %		
Superposition	92 (15)	85 (24)
Juxtaposition	81 (23)	84 (18)
Zeit in Sekunden		
Superposition	3,0 (0,8)	3,0 (1,2)
Juxtaposition	6,1 (2,0)	6,9 (2,2)

Der Prozentsatz der korrekten Lösungen wurde arcsin-transformiert (Neuer Testwert = $\arcsin(\text{alter Testwert})$), die Zeitwerte wurden logarithmiert (neuer Testwert = $\ln(\text{alter Testwert})$). Mit den transformierten Werten wurden zweifaktorielle VAn mit Meßwiederholung auf den Faktoren Graphanordnung (Superposition, Juxtaposition) und Graphtyp (Säulendiagramm, Liniendiagramm) durchgeführt.

Beide Analysen ergaben einen signifikanten Haupteffekt für den Faktor 'Graphanordnung' (Genauigkeit $F(1,24)=16,77$; $p=,000$; $p<.01$; Zeit: $F(1,24)=245,56$; $p=,000$; $p<.01$). Alle übrigen Effekte waren insignifikant.

Gesonderte t-Tests ergaben keinerlei Unterschiede zwischen den Graphiktypen innerhalb einer Anordnung: Supersäule unterscheidet sich nicht signifikant von Superlinie (Genauigkeit: $t(24)=,96$; $p=,348$; $p>.01$; Zeit: $t(24)=,28$; $p=,782$; $p>.01$), Juxtaposition nicht signifikant von Juxtaline (Genauigkeit; $t(24)= -,21$; $p=,832$; $p>.01$; Zeit: $t(24)= 2,09$; $p=,048$; $p>.01$).

Wie die Tabelle 6 aufzeigt, sticht vor allem der klare Zeitvorteil für die Graphanordnung Superposition ins Auge. Die praktische Bedeutsamkeit dieses Effektes steht außer Frage. Man braucht für Juxtaposition ca. doppelt so lange wie für Superposition. Die Unterschiede bzgl. der Genauigkeit sind natürlich abhängig von den Unterschieden der Datenwerte bei der Zielrubrik. In diesem Experiment war der maximale Größenwert der Zielrubrik 10% größer als der zweitgrößte Größenwert, insbesondere deshalb, um bei Juxtaposition noch genaue Schätzungen zu ermöglichen. Bereits bei 10% Unterschied lassen sich Genauigkeitseffekte nachweisen. Man kann die Genauigkeitsunterschiede zwischen den Graphanordnungen durch geringere Prozentsätze sicher vergrößern und möglicherweise eine Differenz ermitteln, bei der unter Superposition noch zuverlässige Angaben möglich sind, unter Juxtaposition hingegen die Differenzierungsfähigkeit klar überfordert ist. Insgesamt bestätigen die Ergebnisse die Befunde von Schutz (1961b), was das Liniendiagramm angeht. Hier wurde lediglich noch darüber hinaus festgestellt, daß das Säulendiagramm analoge Ergebnisse liefert.

Im Gegensatz zu Schutz (1961b) wurden hier klare Genauigkeitsunterschiede zwischen Superposition und Juxtaposition statistisch nachgewiesen. Aus dem Versuchsaufbau von Schutz (1961b) geht hervor, daß der Bereich möglicher Größenwerte viel geringer als hier ausfiel und sich die möglichen Größenwerte viel deutlicher als hier unterschieden, so daß praktisch überhaupt keine Fehler auftreten konnten. Zudem war bei Schutz im Gegensatz zu hier ein Gitternetz unterlegt, was insbesondere die Aufgabe unter Juxtaposition erheblich erleichtert und in jedem Fall zu empfehlen ist, wenn man mit Juxtaposition auch Fragestellungen des point comparisons beantwortet wissen will.

Wieder Erwarten konnten keine Graphypunterschiede nachgewiesen werden. Hier widersprechen die Ergebnisse den Befunden von Jacobs (1990). Es war angenommen worden, daß insbesondere die Identifizierung der Zielrubrik, eine Voraussetzung für den Größenvergleich, beim Säulendiagramm schneller gelingt, weil die Säulen direkt über den Rubriken stehen, beim Liniendiagramm aber erst eine Zuordnung zu den entsprechenden Kreisen vorgenommen werden muß.

Bei Jacobs (1990) lagen 12 Rubriken zugrunde und ein Großteil der Daten basierte auf reinen Zufallswerten für jede Rubrik. Dies führte zu Präsentationen mit vielen Überlappungen (häufige Zickzackkurven) und hoher Konfusion. Möglicherweise zeigt sich ein Graphypeffekt erst dann, wenn die Zuordnung von Rubrik zu den dazugehörigen Kreisen schwieriger wird. Dies ist etwa dann der Fall, wenn die Abstände zwischen den Rubriken enger werden. Die Abstände zwischen den Rubriken waren bei Juxtaposition wesentlich enger als bei Superposition, es wurden aber dennoch keine Graphypunterschiede gefunden. Offenbar waren die Abstände zwischen den Rubriken groß genug, um eine klare Zuordnung von Zielrubrik zu den Zielkreisen vornehmen zu können. Dies war aber auch beabsichtigt. Denn es hat keinen Sinn, Graphiken so darzustellen, daß ihre Schwächen besonders betont werden.

Möglicherweise lassen sich übereinander liegende Linien mit Kreisen aber auch etwas schneller als nebeneinander liegende Säulen vergleichen, was man ja empirisch testen kann. Die Ergebnisse scheinen jedenfalls die Auffassung zu stützen, daß Liniendiagramm sei für die untersuchte Fragestellung geeignet, wenn eine klare Zuordnung von Rubrik zu Größenwert garantiert werden kann. Diese Zuordnung kann z.B. auch durch eine 2. Abszisse oberhalb der Graphik oder noch besser und, vor allem bei vielen Rubriken empfehlenswert, durch ein Gitternetz wesentlich verbessert werden.

8.3.3 Positionierung der Graphiken unter Juxtaposition

Die Graphpositionierung unter Juxtaposition war hier in den Varianten "Diagramme nebeneinander" sowie "Diagramme untereinander" angeordnet. Dies spielt für die untersuchte Fragestellung offenbar überhaupt keine Rolle. Eine zweifaktorielle VA mit den Faktoren Graphpositionierung (untereinander, nebeneinander) und Graphvariante (Juxtaline, Juxtasäule) lieferte sowohl für die Genauigkeit, wie für die Zeit auf allen möglichen Faktoren insignifikante Ergebnisse ($p > .05$). Eventuell gleichen sich Vor- und Nachteile beider Varianten aus. Der Vorteil der vertikalen Anordnung zur Identifizierung der Zielrubrik wird durch den Nachteil beim Größenvergleich der Datenreihen ausgeglichen.

Wenn die Unterschiede der Größenwerte bei der Zielrubrik anwachsen (z.B. Unterschied zwischen höchsten und zweithöchstem Wert 25% oder mehr beträgt und der maximale Wert so sehr spontan wahrgenommen werden kann), dann müßte die vertikale Positionierung Vorteile gegenüber der horizontalen Anordnung zeigen. Möglicherweise ist die vertikale Anordnung für diese Fragestellung auch dann effizienter, wenn die Graphiken ein Gitternetz beinhalten, da so der Größenvergleich erleichtert wird.

8.4 Die Schätzung von Größenwerten (point reading)

Nichts kann Daten genauer darstellen als Zahlen. Daten sind quasi Zahlen. In der Regel dient eine Tabelle als Grundlage für die Erstellung von Graphiken, woraus bereits folgt, daß die graphische Darstellung immer nur höchstens so genau sein kann wie die Zahlen der zugrunde liegenden Tabelle. Hinzu kommt, daß die Genauigkeit der räumlichen Wahrnehmung bald auf ihre Grenzen stößt und Wahrnehmungstäuschungen vorkommen können (z.B. Poulton (1985)). Deshalb wäre es auch unangebracht, überhaupt eine Graphik zu erwägen, wenn das Hauptziel der Präsentation in der Mitteilung der exakten Daten bestünde. Wenn hier dennoch Schätzungen von Größenwerten getestet werden, so liegt der praktische Grund darin, daß man neben dem eigentlichen Zweck der Graphik

zusätzlich noch möglichst genau die Daten mitteilen will. Die durchgeführte Untersuchung kann hierzu dem Graphikdesigner empirisch fundierte Anhaltspunkte für das angestrebte Genauigkeitsniveau geben und klären helfen, ob man zur Graphik zusätzlich Zahlen mitzuteilen braucht oder nicht.

Jacobs (1989) hat recht gründlich die Schätzung von Größenwerten beim Säulendiagramm erfaßt. Die Abweichungen vom tatsächlichen Wert halten sich in Grenzen und betragen beim konventionellen Säulendiagramm im Durchschnitt 1,21% der Größenachse. Durch ein unterlegtes Gitternetz läßt sich dieser Fehleranteil in etwa noch halbieren. Gelegentlich ist die Schätzgenauigkeit für verschiedene Graphikformate getestet worden. Dabei wurden die Ergebnisse jedoch nicht immer mitgeteilt, sondern mit den Ergebnissen anderer Fragestellungen zusammengefaßt (z.B. Culbertson & Powers (1959)). Casali und Gaylin (1988) untersuchten mehrere Graphiktypen und konnten keine signifikanten Unterschiede zwischen Superlinie und konventioneller Supersäule nachweisen. Lediglich das dreidimensionale Säulendiagramm in Superposition erwies sich für die Schätzung von Größenwerten signifikant zeitaufwendiger als die sonstigen Diagramme. Schutz (1961b, S. 117) fand für point reading keine Unterschiede zwischen Superlinie und Juxtalnie. Allerdings konnten die möglichen Punktwerte nur bestimmte, leicht ablesbare Werte annehmen.

Pinker (1990) ist der Ansicht, beim Liniendiagramm sei es schwieriger, absolute Meßwerte zu bestimmen als beim Säulendiagramm, weil graphische Orientierungen zur Bestimmung des genauen Ablesewertes fehlten. Er hat dabei jedoch ein reines Liniendiagramm ohne Rubrikenspezifikationen im Auge. Ohne derartige klare Hinweise kann man den abzuschätzenden Punkt natürlich schwer finden. Gerade deshalb wurden hier ja auch kleine Kreise in die Linien gelegt, damit der Größenwert der Rubrik klar lokalisierbar ist. Jacobs (1994 S.23) äußerte ebenfalls die Vermutung, das Säulendiagramm sei beim Ablesen von Größenwerten dem Liniendiagramm überlegen und meint in diesem Zusammenhang das Punkteliniendiagramm. Er begründet diese Vermutung jedoch nicht näher. Vor der Darstellung weitergehenden Spekulationen sollen daher die Daten zunächst die Frage entscheiden. Die wenigen oben berichteten empirischen Studien lassen eher keine Unterschiede zwischen allen hier untersuchten Graphiktypen erwarten.

8.4.1 Konstruktion der experimentellen Bedingungen:

Die erste Testserie galt dem Vergleich Säulendiagramm vs. Liniendiagramm.

Jedes Diagramm umfaßte 12 Rubriken. Für jede Graphikpräsentation waren 5 Schätzungen vorgesehen. Zunächst wurden nach Zufall die den Daten zugrunde liegenden Kurvenverläufe für die jeweils 5 Präsentationen ausgewählt. Jede VP erhielt eine Funktion mit überwiegend ansteigenden Werten, eine mit überwiegend abfallenden Werten, eine u-förmige oder umgekehrt u-förmige Grundfunktion sowie 2 Funktionen mit mindestens tritonem Grundverlauf.

In der nachfolgenden Testserie wurden die 4 Graphikvarianten, jeweils mit 4 Datenreihen gegeneinander getestet. Jede Graphikvariante umfaßte 10 Rubriken und 4 Aufgaben.

Als Datenkonstellationen für die 4 Datenreihen wurden aus einem zugrunde liegenden Universum ähnlich wie oben nach Zufall bestimmte Grundfunktionen gezogen. Für Aufgabe 1 nur ansteigende Funktionen, für Aufgabe 2 nur fallende Funktionen, für Aufgabe 3 2 ansteigende und 2 fallende Funktionen. Aufgabe 4 lieferte ein sehr komplexes Bild und beinhaltete 4 Datenreihen, wovon jede einen tritonem oder quadritonem Grundverlauf aufwies.

Für beide Testserien gilt:

Die Größenwerte konnten nur ganze Zahlen annehmen. Die Ordinate wies eine Spannweite von 0 bis 100 auf, die Zahlen an der Ordinate waren in 10-er Schritten angebracht und es war festgelegt worden, daß die Größenwerte nur ganze Zahlen zwischen 1 und 100 annehmen können. Nach Zufall wurde

entschieden, an welchen Rubriken die VP die Größenwerte schätzen sollten. Die Präsentationen für alle zu vergleichenden Graphikvarianten waren hinsichtlich der Grundfunktionen und der zu schätzenden Rubriken parallelisiert und wurden in zufälliger Anordnung dargeboten.

8.4.2 Untersuchungsvorgehen:

Neben den üblichen Erklärungen und Beispielen vor dem eigentlichen Experiment erhielten die VPn auch den Hinweis, den exakten Datenwert einer Rubrik im Liniendiagramm in der Punktmitte des kleinen Kreises zu lokalisieren. Die VP erhielt vor jeder Darbietung die Aufforderung, einen bestimmten Größenwert der nachfolgenden Graphik so schnell wie möglich, aber dennoch korrekt abzuschätzen: (z.B.: "Wie hoch ist der Niederschlag zum Meßzeitpunkt 4?; Wie hoch ist der Niederschlagswert zum Meßzeitpunkt 8 bei der roten Datenreihe (bzw. beim Diagramm 2)?"). Durch einen die Zeit initialisierenden, von der VP zu bestimmenden Tastendruck erschien die Graphik. Der nächste Tastendruck beendete die Zeitmessung und brachte die Graphik zum Verschwinden. Nun mußte die VP den Zahlenwert eintippen und bestätigen. Als Fehler wird hier der Betrag der Abweichung der Schätzung vom zutreffenden Wert erfaßt.

$$\text{Fehler} = |\text{Schätzung} - \text{korrekter Wert}|$$

Wegen der Ordinatenkonstruktion entspricht dieser Fehler zugleich dem Prozentsatz gemessen an der Ordinatenhöhe:

8.4.3 Datenauswertung

Die Untersuchungsmethode provoziert dann sehr hohe Fehlerwerte, wenn die VP etwa infolge von Vergessen oder Verwechslung die falsche Rubrik einschätzt. Deshalb wurden aus der Datenanalyse zunächst solche Werte entfernt, bei denen der Verdacht einer derartigen Verwechslung sehr wahrscheinlich wahr. Erkennbar wurde die Abweichung etwa daran, daß solche Werte im Boxplot von Tukey deutlich als Extremwerte in Erscheinung traten. Als Entscheidungskriterium für die Elimination aus den Daten wurde ein Fehlerwert >6% der Größenskala festgelegt. Eine VP wurde bei einer Fragestellung gänzlich aus der Analyse genommen, weil sie gehäuft Extremwerte aufwies. Insgesamt wurden für beide Fragestellungen jeweils etwa 5% aller Daten eliminiert.

8.4.4 Ergebnisse

8.4.4.1 Ergebnisse zur Fragestellung: Wie groß ist Größenwert einer vorgegebenen Rubrik

Sowohl für die Genauigkeit wie für die Zeit erzielten Liniendiagramm und Säulendiagramm vergleichbare Werte, wie die insignifikanten t-Tests der Tabelle 7 belegen.

Tabelle 7: Mittelwerte, Standardabweichungen und t-Tests zur Fragestellung: "Wie groß ist der Größenwert einer vorgegebenen Rubrik ? (z.B. "Wieviel Niederschlag ist zum Meßzeitpunkt 3 gefallen?"); N=25

	Fehler Schätzung-korrekt		Zeit in Sekunden	
	M	s	M	s
Säulendiagramm	1.16	0.79	5.21	1.36
Liniendiagramm	1.25	0.82	5.54	1.13
t(24) =	-1.20 (ns)		-0.51 (ns)	

8.4.4.2 Ergebnisse zur Fragestellung: Wie groß ist bei der Datenreihe x der Größenwert einer vorgegebenen Rubrik ? (z.B. "Wieviel Niederschlag ist zum Meßzeitpunkt 3 bei der roten Datenreihe gefallen?")

Berechnet man eine zweifaktorielle VA mit den Faktoren Graphyp und Graphanordnung, so erhält man einen signifikanten Genauigkeitseffekt für die Graphanordnung zugunsten von Superposition ($F(1,23)=8,94$; $p=,007$; $p<,01$). Die entsprechende VA für die Entscheidungszeiten erbrachte aber überhaupt keine signifikanten Effekte. Der Genauigkeitsvorteil für Superposition könnte in der gegenüber Juxtaposition fast doppelt so hohen Größenachse (Ordinate) begründet sein. Dadurch erhöht sich die Auflösung der Skala für Superposition, was nach Jacobs (1989) die Genauigkeit verbessert. Dieser Vorteil geht aber weniger auf die Anordnung der Daten, die eher Genauigkeitsverluste für Superposition erwarten ließe, sondern auf die hier mit der Anordnung konfundierte Präsentationsart zurück. Man kann auf einer Bildschirmseite die Skala von Superposition eben größer gestalten als die Skalen von Juxtaposition.

Beim Liniendiagramm ließe sich ein Vergleich der Graphikanordnungen mit vergleichbarer Größenachse durchführen (wie Schutz (1961b) dies auch getan hat), beim Säulendiagramm ist dies kaum möglich, weil dann die Säulenbreiten beider Anordnungen oder aber der Abstand zur Ordinate unterschiedlich sein müßten.

Tabelle 8: Mittelwerte, Standardabweichungen und F-Tests zur Fragestellung: Wie groß ist bei einer bestimmten Datenreihe der Größenwert einer vorgegebenen Rubrik ? (z.B. "Wieviel Niederschlag ist zum Meßzeitpunkt 3 bei der roten Datenreihe gefallen?"); N=25

	Fehler Schätzung-korrekt		Zeit in Sekunden	
	M	s	M	s
Supersäule	1.07	0.58	5.40	1.49
Superlinie	1.13	0.61	5.55	1.67
Juxtasäule	1.34	0.49	5.56	1.32
Juxtalinie	1.44	0.79	5.68	1.46
F(3,69)=:	2.38 ns		0.49 ns	

Grundsätzlich lassen sich keinerlei signifikante Unterschiede zwischen den einzelnen Graphikvarianten statistisch erhärten. Die einfaktorielle VA mit den 4 Graphikvarianten als Stufen des Faktors Graphvariante ergab sowohl für die Genauigkeit ($F(3,69)=2,38$, $p=,077$, $p>,01$), wie auch für die Zeit ($F(3,69)=0,49$, $p=0,670$, $p>,01$) insignifikante Effekte. Zwar zeigt sich in Tabelle 8 durchgängig eine leichte numerische Überlegenheit für den Graphyp Säulendiagramm, jedoch sind diese hauchdünnen Unterschiede, wenn überhaupt, dann höchstens von theoretischem Interesse. Es liegen gewiß unterschiedliche Wahrnehmungsbedingungen der einzelnen Graphiktypen zugrunde, deren spezielle Vor- bzw. Nachteile sich aber offenbar weitgehend ausgleichen.

Angesichts der relativ geringen praktischen Bedeutung von point reading für Auswahl einer Graphik, der durchweg passablen Genauigkeits- und der sehr vergleichbaren Zeitwerte für alle Graphikvarianten scheint die Schlußfolgerung angemessen, alle hier getesteten Varianten als im Prinzip für diese Fragestellung ziemlich gleichwertig einzustufen. Es macht wenig Sinn, point reading als entscheidendes Auswahlkriterium für eine der hier getesteten Graphvarianten zu verwenden.

Zur Einschätzung der absoluten Fehlerwerte, die hier doch recht optimistisch stimmen und beim Säulendiagramm eine hohe Übereinstimmung mit den Werten aus der Untersuchung von Jacobs (1989) aufzeigen, ist zu bedenken, daß sehr günstige Bedingungen hinsichtlich der Ordinatewahl und Einteilung vorliegen und die recht hohe Ordinate eine sehr gute Auflösung bietet. Die Ergebnisse können nicht einfach auf Bedingungen generalisiert werden mit kleiner, ungünstiger Ordinatehöhe und einer Zahleneinteilung, die z.B. Interpolieren erschwert, oder auf Graphiken mit deutlich mehr als 12 Rubriken.

Theoretisch ist dieser Befund für mich überraschend. Denn die Zurordnung von Größenwert zur dazugehörigen Rubrik und damit auch die Identifizierung der vertikalen Position, deren Größe eingeschätzt werden soll, erscheint beim Säulendiagramm einfacher. Zudem bietet das Säulenende, quasi als kleine Linie, die in der Wahrnehmung zur Ordinate verlängert werden muß, eine bessere Abschätzungsmöglichkeit als der Mittelpunkt eines Kreises. Auch scheint mir die Lokalisation des zu schätzenden Wertes beim Säulendiagramm transparenter. Man muß den VPn nicht explizit mitteilen, wo der exakte Wert lokalisiert ist, den es zu schätzen gilt, was bei manchen Symbolen im Liniendiagramm auch Kopfzerbrechen mit sich bringen kann). Bei näherer Inspektion einiger Graphiken fiel dann aber doch ein Vorteil für das Liniendiagramm auf. Die Wahrscheinlichkeit, störende Einflüsse zwischen Ordinate und Schätzpunkt aufzufinden, ist beim Liniendiagramm deutlich niedriger. Dieser Unterschied wird z.B. bei einem konstant abfallenden Trend ganz deutlich. Hier kann bei jeder Rubrik im Liniendiagramm störungsfrei eine virtuelle Linie vom jeweiligen Kreis zur Ordinate gezogen werden, beim Säulendiagramm hingegen nimmt die Anzahl der störenden Säulen mit zunehmendem Rubrikenposition zu. Wie Jacobs (1989) feststellen konnte, sinkt die Genauigkeit der Schätzung mit zunehmender Rubrikenanzahl. Dafür ist nicht nur und vielleicht nicht einmal in erster Linie der Abstand der Säule zur Ordinate entscheidend. Denn Jacobs (1989) konnte feststellen, daß keine Genauigkeitsunterschiede zwischen einer kleinen und einer großen Gesamtlänge der Rubrikenachse auftraten.

9.0 Zusammenfassende Schlußfolgerungen:

9.1 Wahrnehmung von Summen bzw. Mittelwerten.

Der günstige Graphiktyp zur Wahrnehmung von Summenunterschieden von Datenreihen hängt von der Graphanordnung ab:

Unter Juxtaposition eignet sich das Säulendiagramm eindeutig eher als das Liniendiagramm, da Juxtasäule stets genauere Werte lieferte als Juxtalinie. Aus diesem Ergebnis ließe sich eventuell auch die Hypothese ableiten, daß globale Vergleiche (Gruppendaten) innerhalb einer Datenreihe im Säulendiagramm besser gelingen als im Liniendiagramm. Dies ist darüber hinaus auch deshalb zu erwarten, weil die Zuordnung der Daten zu den Rubriken im Säulendiagramm besser vorzunehmen ist als im Liniendiagramm, wodurch die Identifikation der zu vergleichenden Abschnitte besser gelingen könnte.

Offenbar ist es einfacher für die Wahrnehmung, nebeneinander stehende Säulen graphisch zu addieren als Linienabschnitte. Theoretisch interessant wäre die Frage, ob man die Addition über Flächen oder Positionen vornimmt. Sicher veranschaulichen mehrere nebeneinander stehende Säulen eher den Eindruck einer Gesamtfläche, auch wenn dieses Gebilde keine einheitliche Form aufweist. Bei einer Linie muß man, um Linienabschnitte als Flächen aufzufassen virtuelle Linien der Endpunkte zur x-Achse hinzu denken. Zur Einschätzung vom Summe wird möglicherweise eher eine Position als mittlerer Punkt auf der Linie angepeilt. Die so geschätzten Punkte sind wegen der räumlichen Nähe der Datenreihen unter Superlinie besser miteinander zu vergleichen als unter Juxtalinie. Verständlich wären so auch die Unterschiede der Graphpositionierung innerhalb von Juxtalinie. Sie tendierten übereinstimmend (jedoch nicht klar signifikant) in die Richtung "günstiger wenn die Diagramme nebeneinander stehen". Dann liegen im Sinne von Cleveland (1985) die Vergleiche auf einer "common alignment scale", während sie in horizontaler Positionierung auf einer common nonaligned scale abgeschätzt werden müßten.

Die Interpretation der Summenunterschiede beider Graphiktypen unter Superposition muß differenziert gesehen werden. Erstaunlich war der Befund, daß Supersäule bei 2 Datenreihen numerisch die höchste Genauigkeit unter allen Graphikvarianten bei noch relativ akzeptabler Schnelligkeit erzielte, weil bei Supersäule die räumliche Nähe auseinandergerissen wird und die Zusammengehörigkeit zu einer Datenreihe stets gestört ist. Vielleicht addiert die Wahrnehmung dann die Menge der Farbe und vergleicht z.B. ob mehr Rot oder Blau in der Präsentation vorhanden ist. Zwar benötigt Supersäule bei vielen Datenreihen (≥ 4) mehr Zeit als Superlinie oder Juxtasäule, aber der Zeitunterschied zu diesen Graphvarianten ist deutlich geringer als z.B. bei der Einschätzung von Verläufen. Wählt man jedoch die Graphanordnung Superposition und will dort viele Datenreihen unterbringen, so ist Superlinie Supersäule auch bei Mittelwertsvergleichen überlegen.

9.2 Wahrnehmung von Variabilität

Die Wahrnehmung von Variabilität gelingt im Liniendiagramm zuverlässig besser als im Säulendiagramm. Die Messung von Variabilität war hier an eine mindestens ordinal-skalierte X-Achse gebunden und für die Einschätzung von Fluktuation innerhalb eines Verlaufs vermutlich ein recht brauchbares Maß. Die Fluktuation wurde als Summe des Betrags der Abweichung aller Größenwerte von den jeweiligen Vorgängergrößenwerten definiert. Im Liniendiagramm entsprach dieses Maß exakt der Länge der Linie und diese ist offenbar unmittelbarer wahrzunehmen als die Summe, der Durchschnitt oder ganz global die Unterschiedlichkeit aus mehreren Differenzen, die im Säulendiagramm abzuschätzen sind.

Im Versuch hatte die zu identifizierende Datenreihe mit höchster Fluktuation im Durchschnitt eine um 25 % höhere Fluktuation als die Datenreihe mit zweithöchster Fluktuation. Die gefundenen Zeiten lassen erkennen, daß ein derartiger Vergleich selten spontan gelingt. Es wurden recht viele Fehler gemacht, so daß offenbar ziemlich große Unterschiede notwendig sind, um zuverlässig zutreffende Entscheidungen zu erzielen. Eine derart mangelnde Präzision kann den Statistiker sicher nicht befriedigen, weswegen die numerische Mitteilung der entsprechenden Statistiken gewiß oft wünschenswert ist.

Variabilität ist ein sehr abstrakter, globaler Begriff, der Unterschiedliches bedeuten kann. Er drängt sich in der Wahrnehmung nicht zwingend als einfaches, einheitliches Gebilde auf. Variabilität mußte in dieser Untersuchung zunächst einmal eindeutig definiert werden und war insofern ein konstruiertes Maß und in dieser Form den VPn keineswegs geläufig. Möglicherweise sind Punktediagramme für die spontane Wahrnehmung der Variabilität bzw. der Variabilitätsunterschiede von Datenreihen besser geeignet (zu optimalen Symbolen von Punkten in graphischen Darstellungen siehe Tremmel 1992).

9.3 Schlußfolgerung zu Summen -und Variabilitätsvergleichen.

Es ist ganz klar, daß die hier analysierten Graphikpräsentationen dann nicht optimal sind, wenn der Vergleich von Mittelwerten oder Streuungen mehrerer Datenreihen von zentralem Interesse sind. Für solche Spezialfragestellungen ist z.B. das Boxplot von Tukey wesentlich effizienter und informativer (siehe dazu auch Cleveland's dot chart ;1985, S. 292). Nur handelt es dabei bereits um die Darstellung statistisch verarbeiteter Informationen und nicht mehr um die Ausgangsdaten. In dieser Untersuchung ging es aber um den Prozeß einer graphischen Mittelwertsbildung und Fluktuationseinschätzung. Dieser kann das Auge und vielleicht auch den Geist ganz schön anstrengen. Spontane Einschätzungen werden aber sicher mit zunehmender Unterschiedlichkeit um so wahrscheinlicher und können so einen ersten groben Eindruck vermitteln. Um die Wahrnehmung der Unterschiede zu fördern, kann (bzw. muß) man einem Vorschlag Kosslyn's (1994) folgend, notfalls die Ordinatenbegrenzungen so modifizieren, daß trotz höhere Auflösung eine faire Graphik entsteht.

9.4 Replikation des Experimentes von Schutz (1961b)

Das Experiment von Schutz (1961b) konnte bzgl. der hier getesteten Aspekte im wesentlichen erfolgreich repliziert werden. Was Schutz für das Liniendiagramm feststellte, gilt in gleichem Maße für das Säulendiagramm. In der Anordnung Superposition ist es wegen der räumlichen Nähe deutlich einfacher als unter der Graphanordnung Juxtaposition, zu entscheiden, welche Datenreihe bei welcher Rubrik den höchsten Wert aufweist. Die Unterschiede zwischen den Graphanordnungen lassen sich sowohl bei der Genauigkeit als auch bei der Zeit statistisch überzeugend nachweisen. Die Unterschiede sind ohne Frage von praktischer Bedeutsamkeit und sind derartige Vergleiche von zentraler Relevanz, sollten die Daten in Superposition dargestellt werden.

Bei Juxtaposition spielt es hier keine Rolle, ob die Diagramme nebeneinander oder untereinander angeordnet sind. Eventuell gleichen sich Vor- und Nachteile beider Varianten aus. Der Vorteil der vertikalen Anordnung zur Identifizierung der Zielrubrik wird durch Nachteil beim Größenvergleich der Datenreihen ausgeglichen.

Sollten aus sonstigen Gründen dennoch Juxtaposition statt Superposition zu Anwendung kommen und dabei auch Größenvergleiche zwischen Elementen mehrerer Datenreihen vorgenommen werden können, dann ist ein Gitternetz in jedem Falle angebracht.

Die Schätzung von Größenwerten gelingt erstaunlich präzise, so daß man in vielen Fällen auf die Angabe von Zahlen verzichten kann. Die Schätzung ist unter beiden Graphikanordnungen weitgehend gleich gut bzw. die gefundenen Unterschiede sind praktisch ohne nennenswerten Belang. Entgegen der Erwartung eines Vorteils für das Säulendiagramm konnten keine signifikanten Unterschiede zwischen den Graphtypen gefunden werden. Man kann in Liniendiagramm und Säulendiagramm Größenwerte vergleichbar gut und schnell einschätzen. Ebenso erkennt man in einer Präsentation mit mehreren Datenreihen im Liniendiagramm und Säulendiagramm mit gleicher Präzision und in gleicher Schnelligkeit die Datenreihe mit dem höchsten Wert bei einer bestimmten Rubrik.

Literatur

- Bertin, J. (1974). Graphische Semiologie. Diagramme, Netze, Karten. (übersetzt von G.Jensch, D. Schade, W. Scharfe). Berlin: Walter de Gruyter.
- Carswell, C. M. (1992). Choosing specifiers: An Evaluation of the basic tasks model of graphical perception. *Human Factors*, 34, 535-554.
- Casali, J. G. & Gaylin, K. B. (1988). Selected graph design variables in four interpretation tasks: a microcomputer-based pilot study. *Behaviour and Information Technology*, 7, 31-49.
- Cleveland, W. S. (1985). *The Elements of Graphing Data*. Monterey, California: Wadsworth advanced Books and Software.
- Croxtan, R. E. & Stein, H. (1932). Graphik comparison by bars, squares, circles and cubes. *Journal of American Statistical Association*, 27, 54-60.
- Jacobs, B. (1989). Schnelligkeit und Genauigkeit beim Abschätzen von Größenwerten aus einem Säulendiagramm. Saarbrücken. Arbeitsberichte des Medienzentrums der Universität des Saarlandes, Nr. 2.

- Jacobs, B. (1990). Ein Vergleich der Auswirkungen graphischer und tabellarischer Präsentationsformen auf die Schnelligkeit und Genauigkeit beim Erkennen und Interpretieren statistischer Daten. Saarbrücken: Arbeitsberichte des Medienzentrums der Universität des Saarlandes, Nr. 3.
- Jacobs. (1994). Der Einfluß von Graphtyp und Graphanordnung auf das Graphverstehen bei der Analyse von Verläufen. Saarbrücken: Arbeitsberichte des Medienzentrums der Universität des Saarlandes, Nr. 13.
- Jacobs, B. (1994b). Graphische vs. tabellarische Präsentation von statistischen Daten. Zeitschrift für Pädagogische Psychologie, 8, 73-84.
- Jacobs. (1995). Experimentelle Analysen zur Wahrnehmung von Kurvenverläufen und Kurvenvergleichen im Säulendiagramm und Liniendiagramm unter Superposition und Juxtaposition.. Saarbrücken: Arbeitsberichte des Medienzentrums der Universität des Saarlandes, Nr. 15.
- Kosslyn, St. M. (1994). Elements of Graph Design. New York: W.H. Freeman and Company.
- Poulton, E. C. (1985). Geometric illusions in reading graphs. Perception & Psychophysics, 37, 543-548.
- Pinker, St. (1990). A theory of graph comprehension. In R. Freedle (Hrsg.), Artificial intelligence and the future of testing (S. 73-126). Hillsdale, NJ: Erlbaum.
- Rinck, M. (1989). Die Strukturierung von Wissen durch statistische Graphen. Dissertation. Universität Marburg.
- Schutz (1961 b). An Evaluation of Methods for Presentation of Graphic Multiple Trends - Experiment III. Human Factors, 3, 108- 119.
- Schnotz, W. (1994). Wissenserwerb mit logischen Bildern. In B. Weidenmann (Hrsg.), Wissenserwerb mit Bildern (S. 95-147). Bern: Hans Huber.
- Tremmel, L. (1992). Untersuchungen zu optimalen Symbolen in graphischen Darstellungen. Hamburg: Verlag Dr. Kovac.
- Wainer, H. (1992). Understanding graphs and tables. Educational Researcher, 21, 14-23.

Glossar

Datenreihe:	Hier Datengrundlage für einen Kurvenverlauf. Ein Liniendiagramm umfaßt 3 Datenreihen, wenn es 3 Kurvenverläufe (Linien) beinhaltet
Superposition:	Graphische Präsentation, bei der mehrere Datenreihen (multiple panels) in einer Graphik dargestellt werden (multiple line diagram nach Schutz (1961b), z.B.: 3 Linien in einem Liniendiagramm).
Juxtaposition:	Graphische Präsentation, bei der jede Datenreihe in einem eigenen Diagramm dargestellt wird (multiple graph diagram nach Schutz (1961b), z.B.: 3 kleine Säulendiagramme nebeneinander).
Graph(ik)typ:	Spezielles Format einer graphischen Präsentation, welches hauptsächlich die räumlichen Spezifikationen der Daten bestimmt (z.B. Liniendiagramm, Säulendiagramm, Punktediagramm)
Graph(ik)anordnung:	Spezielle Anordnung in einem Graphtyp (Superposition, Juxtaposition).
Supersäule:	Säulendiagramm in Superposition.
Juxtasäule:	Säulendiagramm in Juxtaposition.
Superlinie:	Liniendiagramm in Superposition.
Juxtalinie:	Liniendiagramm in Juxtaposition.
Graphvariante:	Oberbegriff für jede mögliche graphische Präsentationen. Supersäule, Juxtasäule usw. sind Graphvarianten.
Diagramm:	Präsentation von Daten. Hier wird darunter meist eine einzelne Graphik unter Juxtaposition verstanden. Eine Graphanordnung mit 4 Datenreihen unter Juxtaposition beinhaltet 4 Diagramme.
Graphpositionierung	auch Diagrammpositionierung=Anordnung der Einzeldiagramme unter Juxtaposition. (Einzelgraphiken untereinander oder nebeneinander)
Aufgabe: Item:	Eine mit Hilfe einer Graphikpräsentation konkret zu beantwortende Frage = Aufgabe.
Testwert:	Summe bzw. Mittelwert aus mehreren Aufgaben für eine Fragestellung.
Testserie	Experimenteller Ablauf, der für die Vp als Einheit aufgefaßt werden kann. Eine Testserie umfaßt Instruktionen zum Ablauf, Beispielanwendungen zu bestimmten Bedingungen und die Testphase. In einer Testserie werden mehrere Bedingungen geprüft. So besteht etwa der Summenvergleich zweier Datenreihen für alle Bedingungen aus einer Testserie, der Größenvergleich von 4 Datenreihen aus einer weiteren Testserie.
virtuelle Linie	Eine nicht in der Graphik vorhandene, aber vom Leser in der Vorstellung konstruierte Linie, welche für die Aufgabenlösung notwendig oder hilfreich ist. z.B. eine senkrechte Linie vom kleinen Kreis in einem Liniendiagramm zur Abszisse, um den Kreis der Rubrikenposition zuordnen zu können.