

[Bernhard Jacobs](mailto:b.jacobs@mx.uni-saarland.de), b.jacobs@mx.uni-saarland.de
Medienzentrum der philosophischen Fakultäten der Universität des Saarlandes

Zusammenstellung der Experimente zum Thema

Identifikation des größten Wertes in Tabelle und Säulendiagramm

- Version vom 26.8.2004 -

Im vorliegenden Dokument sind meine Experimente zusammen gestellt worden, die experimentell die Frage überprüften, ob der größte Wert in einer Präsentation besser in einem Säulendiagramm oder in einer Tabelle erkannt wird. Zudem werden alle Ergebnisse zu den Auswirkungen der Zahlenanordnung (horizontal, vertikal) sowie der Ziffernanzahl (2,3,4) in einer Tabelle mitgeteilt.

Inhaltsverzeichnis	Seite
Hypothesen, Versuchsaufbau und Einzelfallexperiment (1998) aus: http://www.phil.uni-sb.de/~jakobs/wwwartikel/tabgraph/maxwert.htm	2 - 15
Gruppenversuch mit 23 Versuchspersonen (1999) aus: http://www.phil.uni-sb.de/~jakobs/wwwartikel/tabgraph/maxwertalle.htm	16 - 24
Replikation des Gruppenversuchs mit 28 Versuchspersonen (2004) aus: http://www.phil.uni-sb.de/~jakobs/wwwartikel/tabgraph/maxwertalle2004.htm	25 - 29
Das zugrunde liegende Experiment als Toolbookdatei Sie benötigen zusätzlich das Runtime-Modul von Toolbook 3.0 . Nach Durchführung des Experimentes wird automatisch die Datei tabgraph.txt erzeugt, welche die Daten beinhaltet.	
<hr/>	
Kurzfassung: Identifying of Relations between values (1990) aus: Jacobs , B. (1990). Arbeitsbericht Nr. 3 des Medienzentrums	30

Autor: [Bernhard Jacobs](#), Medienzentrum der Philosophischen Fakultät der Universität Saarbrücken
created: 31.8.1998; last update 4.1.1999; [Download als Zip-Datei](#) (Version 4.1.1999) zum Offline-Lesen

URL des Originals: <http://www.phil.uni-sb.de/~jakobs/wwwartikel/tabgraph/maxwert.htm>

wichtige Zusatzinformationen

[Das entsprechende Experiment zusenden!](#)

Nach Durchführung des Experimentes wird automatisch die Datei tabgraph.txt erzeugt, welche die Daten beinhaltet.

Im Dezember 1998 wurde das Experiment als [Gruppenversuch](#) mit 23 Vpn durchgeführt und konnte überzeugend bestätigt werden.

Identifikation des größten Wertes in Tabelle und Säulendiagramm

[Zusammenfassung](#)

[Problemstellung](#)

[Versuchsaufbau und Versuchsablauf](#)

[Hypothesen](#)

[Ergebnisse](#)

[Diskussion](#)

[Literatur](#)

Zusammenfassung

Im Mittelpunkt der Arbeit steht die Hypothese, dass sich der Vorteil einer graphischen Präsentation von Daten gegenüber einer Tabelle mit zunehmender Komplexität - hier operationalisiert als Anzahl der Daten - erhöht. Die Aufgabenstellung für die Präsentationen Tabelle und Säulendiagramm bestand darin, jeweils den größten/höchsten Wert zu identifizieren. Im Rahmen eines Selbstversuchs in Form eines Einzelfallexperimentes konnte die erwartete ordinale Interaktion zwischen Präsentationsform und Anzahl der zu vergleichenden Daten eindeutig zu Gunsten des Säulendiagramms bestätigt werden. Weitere Befunde deuten an, dass in einer Tabelle die vertikale Anordnung von Zahlen etwas günstiger erscheint als die horizontale Anordnung. Obgleich Säulen schneller miteinander verglichen werden können als Zahlen, wird sich der Vorteil der Graphik in der Praxis nur dort zeigen, wo es gelingt, die Zuordnung der Variablen zu den Daten in der Graphik schnell zu vermitteln.

Problemstellung

Die graphische Präsentation von Daten erweist sich nur dann als vorteilhaft, wenn relevante Beziehungen zwischen den Daten durch die räumliche Anordnung einfacher als auf der Basis symbolisch vermittelter Zahlen zu erkennen sind. Die Graphik muss folglich im Hinblick auf die Beantwortung der anstehenden Fragestellung einer Tabelle überlegen sein, was sich insbesondere in einer schnelleren Entscheidungszeit zeigen müsste, da das Erkennen mit weniger Anstrengung und geringerem geistigem Aufwand verbunden ist.

Eine dieser Fragestellungen bezieht sich auf die Relation größer/kleiner zwischen den Daten. Die Frage nach dem größten Wert basiert auf einer derartigen Relation und erfordert zumindest bei Zahlen in der Regel mehrere Vergleichsoperationen. In einem Säulendiagramm springt der größte Wert hingegen unmittelbar ins Auge. Aufwendige und anstren-

gende Vergleiche sind in der Regel nicht erforderlich. Man kann die höchste Säule entdecken ohne im einzelnen jede tiefere Säule einzeln inspizieren zu müssen. Zumindest kann man ganze Bereiche von tieferen Säulen in einem Zug ausschließen. Das Säulendiagramm ist zur Identifikation des größten Wertes sicher eine der bestmöglichen unter den bekannten graphischen Präsentationen, wenngleich alle Graphikformate, welche die Daten "along a common scale" im Sinne von Cleveland (1985) anordnen, hier ebenfalls geeignet erscheinen. Allerdings konnte Jacobs (1995a) einen signifikanten, aber geringfügigen Zeitvorteil des Säulendiagramms gegenüber dem Liniendiagramm beim Entdecken des größten Wertes empirisch nachweisen. (siehe: [Locate the Maximum or Minimum Value within a Data Set!](#))

Insgesamt ist zu erwarten, dass die Identifizierung des größten Datenwertes in einem Säulendiagramm schneller vollzogen werden kann als in einer Tabelle. Diese Hypothese wurde in einer früheren Untersuchung (Jacobs 1990, Kurzfassung: [Identifying of Relations between values](#)) eindeutig empirisch bestätigt. In einer 12 Daten umfassenden Datenreihe konnte der größte bzw. niedrigste Datenwert in einem Säulendiagramm mit einer Effektstärke >1 schneller entdeckt werden als in einer entsprechenden Tabelle. Weitere Befunde zeigten, dass der Vorteil des Säulendiagramms mit zunehmender Komplexität anwächst. Ein Aspekt der Komplexität bezieht sich auf die Anzahl der zu vergleichenden Daten. Mit wenigen Ausnahmen gilt als allgemeine Empfehlung, Graphiken bevorzugt dann einzusetzen, wenn viele Daten dargestellt werden sollen (z.B. Gillan, Wickens, Hollands & Carswell (1998)), weil die räumliche Anordnung die Gruppierung der Daten sowie die Zusammenfassung zu graphischen Chunks wesentlich vereinfacht und insgesamt einen besseren Gesamtüberblick verschafft. Hohe Komplexität wäre somit ein Kriterium, die Daten graphisch zu präsentieren.

Insgesamt ist zu erwarten, dass der Unterschied zwischen Säulendiagramm und Tabelle im Hinblick auf die Fragestellung: "Identifiziere den größten Datenwert" mit zunehmender Anzahl der Daten anwächst. Dies entspricht der Erwartung einer ordinalen Interaktion zwischen Präsentationsform (Säulendiagramm, Tabelle) und Anzahl der Daten in der Präsentation.

Ein fairer Vergleich von Präsentationsformen setzt voraus, dass jede Präsentationsform in der für sie optimalen Realisation zugrunde gelegt wird. Hier kann man die Untersuchung von Jacobs (1990) vielleicht angreifen und argumentieren, Tabellen wären dort suboptimal dargestellt worden. Die Zahlen in der Tabelle waren horizontal angeordnet und als Zahlen fungierten meist Daten mit insgesamt 4 Zeichen (z.B.: 48.8, 69.5, 9.4 usw.). Die Zahlen selbst waren noch durch vertikale Linien unterbrochen, was möglicherweise auch störend wirken kann. Nach Ballstaedt (1997, S. 140, basierend auf Chaprakani & Ehrenberg (1976)) sind Zahlen innerhalb von Spalten besser zu vergleichen als Zahlen innerhalb von Zeilen. "Put figures to be compared in columns instead of rows.... It is easier to read figures down than across a page, because we are used to doing arithmetic that way." (Empfehlung von Ehrenberg 1981 nach Capoza.) Außerdem scheinen die Annahmen plausibel, Zahlen seien umso einfacher wahrzunehmen, je weniger Ziffern sie umfassen und der Dezimalpunkt erhöhe zusätzlich die Komplexität einer Zahl. Nicht ohne Grund wird für die Zahlendarstellungen in Tabellen empfohlen, die Zifferanzahl einer Zahl -vornehmlich die nach dem Dezimalpunkt- so weit zu reduzieren wie die Zuverlässigkeit der Genauigkeit eines Wertes dies rechtfertigt, um Vergleiche zwischen Zahlen zu erleichtern. "Ehrenberg (1977) and Wainer (1997) argue rather persuasively that numbers in tables should be rounded to two digits" (zitiert nach Natriello (1997)).

Versuchsaufbau und Versuchsablauf

Um die obigen Überlegungen umfassend zu prüfen, wurde der in Tabelle 1 verdeutlichte Versuchsaufbau gewählt.

Tabelle 1: Experimentelle Bedingungen und Anzahl der Präsentationen pro Bedingung				
		Präsentationsformen		
		Säulendiagramm	Tabelle (Zahlen)	
			horizontal	vertikal
Anzahl der zu vergleichenden Daten	4 Daten	8	6	6
	8 Daten	8	6	6
	16 Daten	8	6	6
	32 Daten	8	-	6

Wie aus der Tabelle 1 hervorgeht, bestand der Versuch aus 74 Präsentationen (Items). Eine horizontal angeordnete Tabelle mit 32 Daten war auf dem Bildschirm nicht mehr darstellbar, womit die Bedingung entfallen musste. Der Versuch enthält seine eigene Replikation, d.h. es wurde dafür gesorgt, dass die Hälfte der Bedingungsrealisationen im ersten (A) und die andere Hälfte im zweiten Durchgang (B) getestet wurde, um mögliche Trainingseffekte analysieren zu können. Die Reihenfolge aller Präsentationsbedingungen innerhalb von A und innerhalb von B wurde jeweils nach Zufall bestimmt. Alle Zahlen in der Tabelle enthielten entweder 2, 3 oder 4 Ziffern. Für den hier relevanten Vergleich der Präsentationsformen liegt ein faktorieller Versuchsplan zugrunde mit den Faktoren

Präsentationsformen (Säulendiagramm, Tabelle)
Anzahl der Daten (4,8,16,32)

Die Präsentationsform Tabelle wird hierbei über alle Variationen gemittelt und beinhaltet bei 32 Daten eine vertikale Anordnung der Zahlen. Konzentriert man sich nur auf die Präsentationsform Tabelle, so wird der Versuch hier als 3 faktorieller Versuchsplan gedeutet mit den Faktoren:

Präsentationsform (horizontale und vertikale Anordnung der Zahlen)
Anzahl der Daten (4,8,16)
Zifferngröße der Daten (2,3,4)

Konstruktion der Datenwerte in einer Präsentation

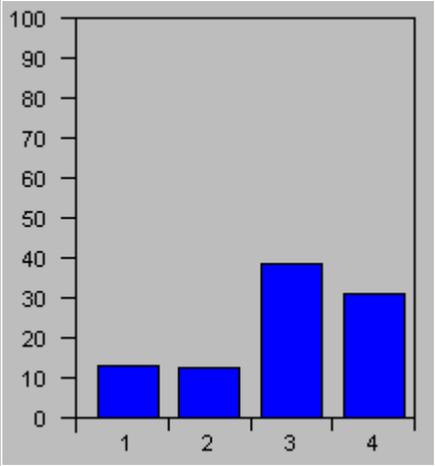
Um in einer Graphik den größten Wert entdecken zu können, muss der größte Wert sich vom zweitgrößten Wert so stark unterscheiden, damit er durch Wahrnehmung zuverlässig als größter Wert identifiziert werden kann. Natürlich lassen sich leicht Daten konstruieren, bei denen der größte Wert überhaupt nur in einer Tabelle sicher festgestellt werden kann, aber derartige Unterschiede interessieren hier nicht. Es wurde daher festgelegt, dass in jeder Präsentation der größte Werte mindestens 5% größer sein sollte als der zweitgrößte Wert. Für einen absolut validen Vergleich der Bedingungen benötigte man eigentlich eine Zufallsstichprobe von Daten aus dem Universum aller Daten. Diese Forderung ist gänzlich unrealistisch.

Die Datenwerte jeder Präsentation basieren auf Zufallszahlen, wobei gewisse Restriktionen festgelegt wurden, z.B. um zu verhindern, dass die größten Zahlen meistens mit der Ziffer 9 beginnen. Die Reaktionsgeschwindigkeit zum Erkennen der größten Zahl ist nicht für alle möglichen Vergleichszahlen äquivalent und hängt von verschiedenen Bedingungen ab ([Zwisler 1993](#)). Es sind auch Daten denkbar, welche systematisch die Tabelle benachteiligen

würden, z.B. wenn bei einer Zahlenkolonne von 4 ziffrigen Zahlen 66% der Zahlen als erste Ziffern etwa 93 aufweisen würden und die größte Zahl auch mit 93 beginnen würde. Dann wären viele Vergleichsoperationen notwendig wären, um die größte Zahl zu identifizieren. Derartige Konstellationen sind beim Versuch allerdings sehr unwahrscheinlich. Es bleibt zu hoffen, dass die konstruierten Zahlen einen fairen Vergleich der Präsentationsbedingungen ermöglichen. Das verwandte Zufallsprinzip produziert jedenfalls eine große Fülle unterschiedlicher Zahlen und verhindert so die Häufung systematischer Zahlenkonstellationen, die eine Bedingung unverhältnismäßig bevor- oder benachteiligen würden. Die Werte für das Säulendiagramm basieren auf dem gleichen Konstruktionsprinzip wie die Generierung 3 ziffriger Zahlen, wobei die Säulen aus den Zufallszahlen als Prozentsätze der Größenachse (Ordinate, y-Achse) definiert wurden. Der Vergleich der Bedingungen basiert nicht auf identischen Daten, sondern auf Zufallsrealisationen aus derselben Grundmenge. Auf diese Weise entfällt das Problem der Erinnerung und erhöht sich die Generalisierbarkeit.

Besonderheiten der graphischen Darstellungen und Beispiele

Die experimentellen Datenpräsentationen (siehe Beispiel in Abbildung 1) unterscheiden sich von den üblichen Darstellungen in der Praxis dadurch, dass sie weder eine Überschrift noch weitere Angaben dazu beinhalten, was die Zahlenwerte im einzelnen bedeuten. Die Präsentationen beschränken sich nur auf die unbedingt notwendigen Angaben, ein bestimmtes Datum einer Rubrik (Kategorie) zuzuordnen zu können. Damit reduziert sich die Fragestellung hauptsächlich auf ein Problem der Wahrnehmung, ob man eher die Größe von Zahlen oder die Höhe von Säulen miteinander vergleichen kann. Derartige Operationen müssen auch in ökologisch validen Präsentationen durchgeführt werden, aber man kann nicht einfach davon ausgehen, dass Tabellen und Graphiken bei den sonstigen Anforderungen identisch sind - was noch diskutiert wird - und die Ergebnisse bedenkenlos generalisieren. Andererseits begründen ja gerade die unterschiedlichen Wahrnehmungsbedingungen (räumlich vs. symbolisch) die Verwendung von Graphiken.

Abbildung 1: Beispiele für die experimentellen Präsentationen zur Fragestellung Bei welcher Rubrik (Kategorie) liegt der größte Datenwert ?																																																		
Säulendiagramm, 4 Säulen	Tabelle, 8 Daten, 2 Ziffern; horizontale Anordnung der Zahlen	Tabelle, 16 Daten, 4 Ziffern; vertikale Anordnung der Zahlen																																																
	<table border="1" data-bbox="687 1619 1114 1733"> <tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td></tr> <tr><td>24</td><td>11</td><td>33</td><td>59</td><td>16</td><td>34</td><td>25</td><td>11</td></tr> </table>	1	2	3	4	5	6	7	8	24	11	33	59	16	34	25	11	<table border="1" data-bbox="1182 1402 1385 1928"> <tr><td>1</td><td>2750</td></tr> <tr><td>2</td><td>1622</td></tr> <tr><td>3</td><td>6047</td></tr> <tr><td>4</td><td>1069</td></tr> <tr><td>5</td><td>3446</td></tr> <tr><td>6</td><td>1025</td></tr> <tr><td>7</td><td>3855</td></tr> <tr><td>8</td><td>2735</td></tr> <tr><td>9</td><td>4090</td></tr> <tr><td>10</td><td>1524</td></tr> <tr><td>11</td><td>1637</td></tr> <tr><td>12</td><td>3895</td></tr> <tr><td>13</td><td>5359</td></tr> <tr><td>14</td><td>8620</td></tr> <tr><td>15</td><td>2711</td></tr> <tr><td>16</td><td>6818</td></tr> </table>	1	2750	2	1622	3	6047	4	1069	5	3446	6	1025	7	3855	8	2735	9	4090	10	1524	11	1637	12	3895	13	5359	14	8620	15	2711	16	6818
1	2	3	4	5	6	7	8																																											
24	11	33	59	16	34	25	11																																											
1	2750																																																	
2	1622																																																	
3	6047																																																	
4	1069																																																	
5	3446																																																	
6	1025																																																	
7	3855																																																	
8	2735																																																	
9	4090																																																	
10	1524																																																	
11	1637																																																	
12	3895																																																	
13	5359																																																	
14	8620																																																	
15	2711																																																	
16	6818																																																	

Versuchsablauf

Nach einer entsprechenden Instruktion und der Möglichkeit mehrere Beispiele auszuprobieren, begann der eigentliche Versuch. Die Vp erhielt stets die Anweisung, aus der Präsentation den größten Wert so schnell wie möglich, aber dennoch korrekt zu erfassen. Die Abfolge einer Präsentation gestaltete sich wie folgt

1. Fragestellung: Welche Rubrik umfasst den größten Datenwert ?
2. Klick auf die Leertaste = Präsentation erscheint, Zeitmessung beginnt.
3. Klick auf die Leertaste = Präsentation verschwindet, Zeitmessung stoppt.
4. Eingabe der Rubrik für das Datum mit dem größten Wert.

Versuchsperson

Als alleinige Versuchsperson fungierte der Autor dieses Beitrags. Mithin handelt es sich um ein Einzelfallexperiment. Um die Zuverlässigkeit des Gesamtversuchs zu erhöhen, wurde der Versuch an 2 Tagen insgesamt 10-mal durchgeführt. Gleichzeitig sollten so Trainingseffekte abgeschätzt werden können.

Zentrale Hypothesen

1. Wegen der sehr hohen Erfahrung der Vp im Umgang mit Tabellen und Graphiken ist im Verlauf der Versuchsdurchgänge keine Leistungssteigerung mehr festzustellen.
2. Der Vorteil des Säulendiagramms gegenüber der Tabelle wächst mit zunehmender Anzahl der zu vergleichenden Datenwerte (=ordinale Interaktion zwischen Präsentationsform und Anzahl der Daten). Das ist die wichtigste Fragestellung dieser Untersuchung.
3. Zahlen sind in vertikaler Anordnung besser zu vergleichen als in horizontaler Anordnung
4. Je mehr Ziffern eine Zahl umfasst, desto schwieriger ist die Identifikation des größten Wertes.

Ergebnisse

Ergebnisse Säulendiagramm vs. Tabelle

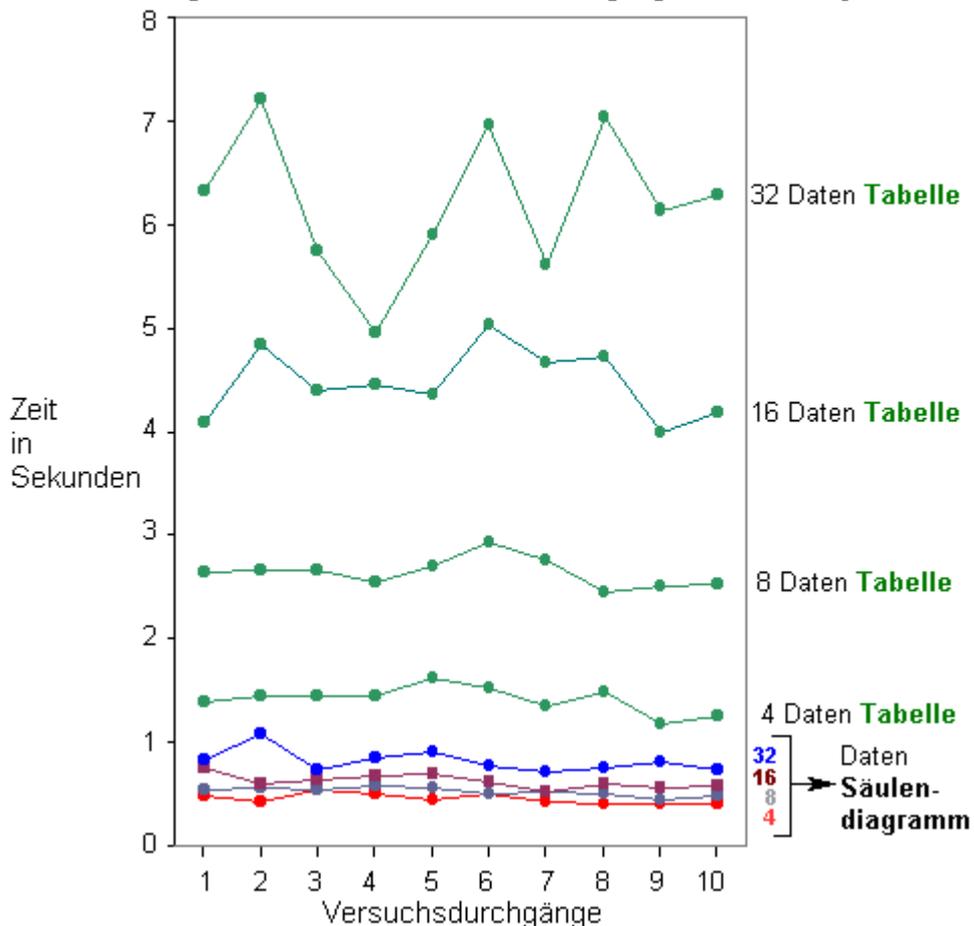
Die Darstellung der Ergebnisse bezieht sich vornehmlich auf die benötigte Zeit. Differenzierte Genauigkeitsanalysen für einzelne Bedingungen sind wegen zu geringer Fehleranzahl schwer durchführbar. Eine Analyse der Genauigkeiten aller erhobenen Daten ergab grob betrachtet folgendes Bild:

Säulendiagramme: 99 % richtig
 Tabellen bis 8 Zahlen: 97 % richtig
 Tabellen 16-32 Zahlen: 87 % richtig

Die Abbildung 2 zeigt die Ergebnisse der Vp Jacobs für die Zeiten über 10 Versuchsserien hinweg. Die graphische Darstellung erweist sich in der Tat als sehr nützlich, was sicher auch an den Daten selbst bzw. an der Solidität des Versuchsaufbaus liegt. Es lassen sich mehrere wichtige Relationen auf einen Blick erkennen.

Abbildung 2

Zeit zur Identifikation des größten Wertes für Tabelle und Säulendiagramm über 10 Versuchsdurchgänge bei einer Vp



Anmerkung: Jeder Datenpunkt repräsentiert den Mittelwert aus mehreren Präsentationen (genauerer siehe [Tabelle 1](#))

- Die Zeiten für die Tabelle sind über alle Versuche hinweg höher als die für die Säulendiagramme
- Je mehr Daten, desto schwieriger ist der größte Wert zu identifizieren.
- Die Zeitunterschiede bzgl. der Datenanzahl sind bei der Tabelle konsistent höher als beim Säulendiagramm. (Dies unterstreicht die Interaktion zwischen Präsentationsform und der Anzahl der zu vergleichenden Daten)
- Die Zeitrelationen aller 8 Bedingungen sind über alle Versuche weitgehend vergleichbar. Ein Versuchsdurchgang scheint - zumindest für Vp Jacobs- ein ziemlich reliables Ergebnis zu liefern.
- Es gibt bei keiner der 8 Bedingungen eine deutlich erkennbare Zeitreduktion im Verlauf der Versuchsdurchgänge. Das Leistungslimit ist offensichtlich erreicht und nähert sich der Asymptote.
- Die Versuchsperson Jacobs kann offenbar in einem Säulendiagramm schneller den größten Wert aus 32 Daten identifizieren als in einer Tabelle aus 4 Zahlen die größte Zahl entdecken. Dies hat sie jedenfalls in 10 Versuchen nachgewiesen.

Hinweis:

Die größere Variabilität bei der Bedingung 32 Daten Tabelle geht zum Teil auf eine geringere Zuverlässigkeit zurück, da hier jedem Datenpunkt nur 6 Präsentationen (nur die vertikalen Tabellen) zugrunde liegen, allen anderen Tabellendaten jedoch 12.

Die zentrale Hypothese des Versuchs lautete, dass die Unterschiede zwischen den Präsentationsformen mit wachsender Komplexität deutlicher zu Gunsten des Säulendiagramms in Erscheinung treten müssten. Dies geht bereits aus Abbildung 2 hervor, allerdings in einer für Interaktionsdarstellungen ungewohnten Präsentation. Aus diesem Grunde sollen die Ergebnisse nun noch in der dafür besser geeigneten Abbildung 3 dargestellt werden, wobei die Daten zusätzlich noch so aggregiert werden sollen, damit die zentrale Hypothese prägnant und mit höchstmöglicher Zuverlässigkeit mit der Empirie konfrontiert werden kann. Deshalb werden die Ergebnisse für über alle Versuchsdurchgänge gemittelt.

Abbildung 3: Tabelle und graphische Darstellung der Ergebnisse für die Vp Jacobs

Zeit in Sekunden zur Identifizierung des größten Wertes für unterschiedliche viele Daten in Tabelle und Säulendiagramm.

Mittelwert (M) und Streuung (s)

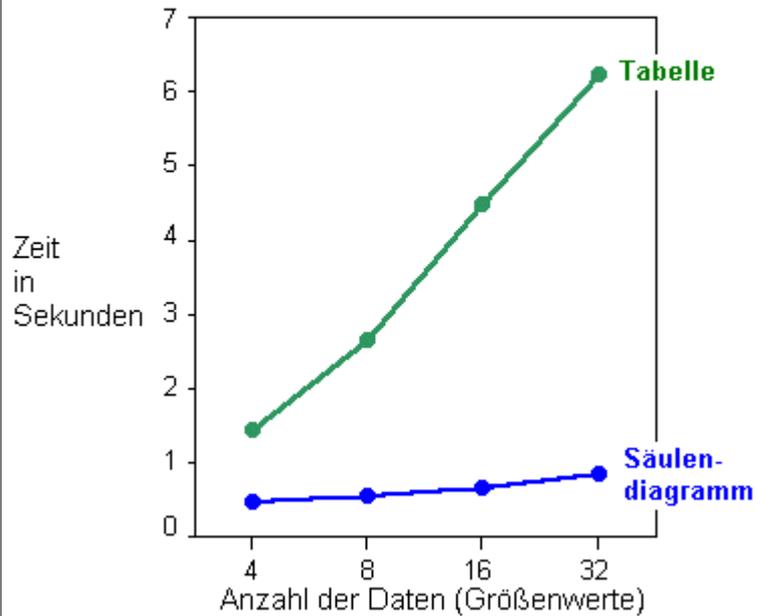
		Anzahl der Daten			
		4	8	16	32
Tabelle	M	1.42	2.63	4.47	6.22
	s	0.13	0.14	0.34	0.70
Säulendiagramm	M	0.46	0.52	0.63	0.82
	s	0.04	0.04	0.07	0.11

Der Mittelwert **M** ergibt sich aus den Mittelwerten aus 10 Versuchsserien.

Die Standardabweichung **s** bezieht sich auf die Streuung über die Mittelwerte aus den 10 Versuchsserien.

Insgesamt liegen den Daten die Einschätzungen aus 740 Präsentationen zugrunde.

Interaktion zwischen Präsentationsform und Anzahl der zu vergleichenden Daten für die Fragestellung: "Identifiziere den größten Wert in der Präsentation!"



Wie aus der Abbildung 3 hervorgeht, konnte die Interaktion zwischen Präsentationsform und Komplexität - hier operationalisiert als die Anzahl der Daten - für die Vp Jacobs eindeutig bestätigt werden. Der Unterschied zwischen den Präsentationsformen wächst mit zunehmender Anzahl der zu vergleichenden Daten zugunsten des Säulendiagramms. Anders ausgedrückt: Die benötigte Zeit steigt mit wachsender Zahl der zu vergleichenden Daten bei der Tabelle deutlich steiler an als im Säulendiagramm (= ordinale Interaktion). Dies spricht für eine überwiegend unmittelbare Wahrnehmung des größten Wertes im Säulendiagramm und eine eher sequentielle Suche nach dem größten Wert in der Tabelle. Die Schnelligkeit im Säulendiagramm geht mit Sicherheit nicht zu Lasten der Genauigkeit, da in der tabellarischen Darstellung mehr Fehler gemacht wurden. Auch ohne Signifikanzprüfung ist die Zuverlässigkeit dieser Interaktion, wie aus den Standardabweichungen der einzelnen Versuchsmittelwerte abzuleiten ist, erstaunlich hoch und auf Einzelfallniveau außergewöhnlich hinreichend überzeugend. Wie aus Abbildung 2 hervorgeht, ist diese Interaktion auch an der Leistungsgrenze zu erwarten, so dass man nicht argumentieren kann, mangelnde Erfahrungen im Umgang mit einer der Präsentationsformen könnten hier verzerrend einwirken.

Ergebnisse zur Tabellenanordnung

Bei allen Tabellenvergleichen wurden nur Tabellen bis 16 Zahlenwerte zur Auswertung verwendet.

Horizontale versus vertikale Zahlenanordnung in der Tabelle

In Hypothese 3 wurde vermutet, dass Zahlen innerhalb einer Spalte (vertikale Tabellenanordnung) besser miteinander verglichen werden können als Zahlen innerhalb einer Zeile (horizontale Tabellenanordnung).

Tabelle 2: Ergebnisse zur Zahlenanordnung in der Tabelle: Mittelwerte aus allen Versuchsdurchgängen

	Zahlen- anordnung	Anzahl der Daten		
		4	8	16
Fehler in Prozent	horizontal	3	7	18
	vertikal	0	2	5
Zeit in Sekunden	horizontal	1.5	2.7	4.9
	vertikal	1.2	2.6	4.0

Hinweis: Jede Bedingung basiert auf 60 Präsentationen

Tabelle 2 scheint diese Hypothese tendenziell zu stützen, da alle Vergleiche eher für die vertikale Tabellenanordnung sprechen. Die Unterschiede sind aber nicht sehr groß. Lediglich bei 16 zu vergleichenden Daten ist ein relativ zuverlässiger Vorteil zugunsten der Zahlenvergleiche innerhalb der Spalte auszumachen. (Hier betragen die Standardabweichungen der Zeiten bezogen auf die Mittelwerte über die Versuchsserien 0.55 bzw. 0.36). Dieses Ergebnis steht im Einklang mit der Einschätzung von Ehrenberg (1977, S.282) "Figures are easier to follow reading down a column than across a row, especially for a larger number of items." In der Praxis wird der Vorteil der vertikalen Positionierung der Zahlen vermutlich deutlich höher sein, weil die Zahlen innerhalb einer Reihe meist weiter auseinander liegen, allein schon deshalb, um die Beschriftung der Spalten bequemer vornehmen zu können. Im Experiment wurde aber darauf geachtet, die räumliche Nähe der Zahlen in einer Reihe so eng wie irgendetmöglich zu gestalten, was der Regel 5 von Ehrenberg (1977, S.285) entgegenkommt: "The rule is that figures which are meant to be compared should be placed close together".

Einfluss der Zifferanzahl

Die letzte Hypothese postulierte schwierigere Vergleiche mit wachsender Zifferanzahl der Zahlen. Wie aus Tabelle 3 ersichtlich, zeigen sich die erwarteten Unterschiede, wenn auch nicht in einer dramatischen Größenordnung. Aber immerhin beträgt der Unterschied zwischen 2 und 3 Ziffern im Durchschnitt eine halbe Sekunde.

Tabelle 3: Fehler und Zeiten für Größenvergleiche mit 2,3 und 4 ziffrigen Zahlen in einer Tabelle.

	Anzahl der Ziffern pro Zahl		
	2	3	4
Fehler in Prozent	3	7	7
Zeit in Sekunden	2.4	2.9	3.2
(s für Zeit)	(0.15	0.32	0.25)

 Hinweis: Jede Bedingung basiert auf 120 Präsentationen
 s für Zeit = Standardabweichung über die 10 Versuchsserien.

Das Problem der Zifferanzahl stellt sich nicht beim Größenvergleich von Säulen in einem Säulendiagramm, da alle Zahlen als Anteile an der Ordinate festgelegt werden und da spielt es keine Rolle, ob man Einer oder Millionen miteinander vergleicht. Der tabellarische Genauigkeitsvorteil bei der Mitteilung exakter Daten konfliktiert somit gelegentlich mit der Leichtigkeit, die Daten miteinander vergleichen zu können. Zur Verbesserung der tabellarischen Präsentation sollte man aber, wo immer möglich, Regel 1 von Ehrenberg (1977,S.281) beherzigen: "Rounding to Two Significant or Effective Digits."

Diskussion

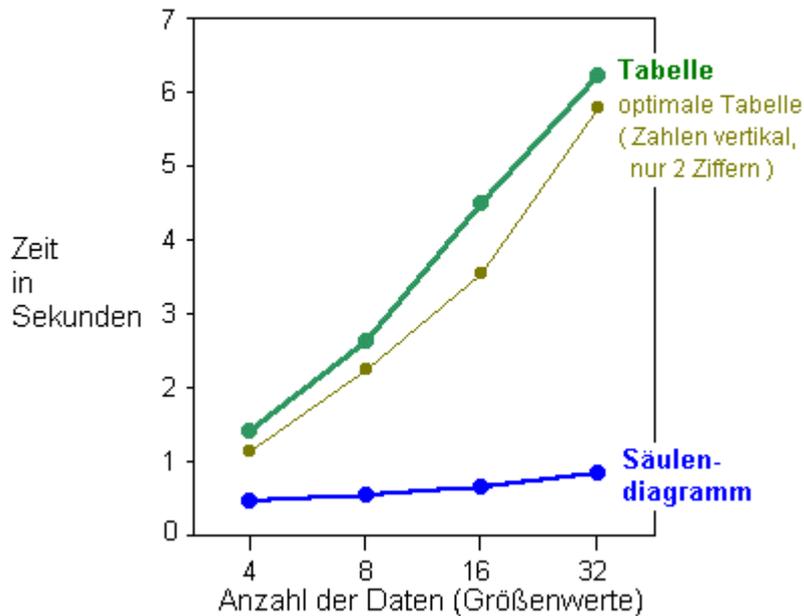
Aussagekraft der Untersuchung

Obgleich die Daten alle von einer Vp stammen, schätze ich die Aussagekraft des Experimentes ziemlich hoch ein. Die Randomisierung der Darbietungsreihenfolge aller Bedingungen und die Zufallsprozeduren zur Erstellung des Reizmaterials sorgen für eine recht hohe interne Validität. Die große Zahl der Bedingungsreplikationen stärkt die Zuverlässigkeit der Befunde. Es mangelt der Untersuchung, da Einzelfall orientiert, eindeutig an externer Validität im Sinne der Populationsrepräsentativität, was mir im gegebenen Fall allerdings nicht so gravierend erscheint, da vornehmlich Wahrnehmungsprozesse untersucht werden, die bei den meisten Menschen mit gewissen Variationen ähnlich verlaufen müssten. dass der Forscher selbst die Vp darstellt, ist meiner Meinung nach die größte Schwäche dieser Studie. Die Hypothesen waren dem Forscher vor dem Experiment bekannt und auch wenn hier das Bemühen um ein faires Vorgehen behauptet wird, so kann mehr oder weniger unbewußtes, "hypothesenkonformes" Verhalten sicher nicht ausgeschlossen werden.

Diese Einwände lassen sich ganz einfach durch ein Gruppenexperiment mit hinreichend vielen Vpn ausräumen, wofür sich vielleicht einmal die Gelegenheit bietet und werden im folgenden als technisch lösbar eingeschätzt. [Wie die entsprechende Untersuchung, durchgeführt im Dez. 1998 zeigte, ist die Übereinstimmung des Einzelfalles mit dem [Gruppenversuch](#) als sehr gut einzuschätzen (Jacobs 1999)]

wichtigstes Ergebnis

Interaktion zwischen Präsentationsform und Anzahl der zu vergleichenden Daten für die Fragestellung: "Identifiziere den größten Wert in der Präsentation !"



Die zentrale Interaktionshypothese des mit wachsender Datenanzahl steigenden Vorteils von Säulenvergleichen im Säulendiagramm gegenüber Zahlenvergleichen in der Tabelle ist ganz eindeutig bestätigt worden und es wäre ganz unwahrscheinlich, dieses Ergebnis nicht in einem soliden Gruppenexperiment replizieren zu können. **Man kann den größten Wert unterschiedlich hoher Säulen schneller und mit weniger Anstrengung identifizieren als den größten Wert unter verschiedenen Zahlen und der Vorteil für die Säulen steigt mit wachsender Zahl der Vergleich-**

objekte. Auch wenn man als Tabellenpräsentation nur Vergleiche innerhalb einer Spalte verwendet und der Empfehlung von Ehrenberg und Wainer weiter folgend nur zweiziffrige Zahlen in die Analyse aufnimmt, sieht die Interaktion zwischen Präsentationsform (optimale Tabelle, Säulendiagramm) und Anzahl der Daten nicht wesentlich anders aus. Das Säulendiagramm ist auch der optimal gestalteten Tabelle klar überlegen.

Generalisierung der Ergebnisse

Die wichtigere Frage erscheint mir, ob man dieses Ergebnis einfach auf Tabellen und Säulendiagramme verallgemeinern kann, wie man sie üblicherweise vorfindet, denn die Interpretation einer Präsentation erschöpft sich nicht im Vergleich von Säulen oder Zahlen. Man muss die Bedeutung der Daten verstehen. "...the real problem of new data is not that of presenting it well, but of having first to understand it". (Ehrenberg 1977, S.295) Insbesondere die Zuordnung von Datum zur Bezeichnung, für was das Datum steht, also etwa die Fragen "Was wird hier dargestellt? Was ist die abhängige oder unabhängige Variable" Was misst X- bzw. Y-Achse, Was bedeutet diese Kategorie unter der x-Achse? usw. sind durch die Untersuchung gar nicht erfasst. Hier gibt es aber unterschiedliche Orientierungsvorteile für Tabelle und Graphiken. Bei einer Datenreihe treten diese Unterschiede nicht besonders klar in Erscheinung und in gut konzipierten Präsentationen beider Präsentationsformen müsste man relativ eindeutig und schnell die Variablenbedeutung erkennen können. Insofern glaube ich, dass die Zeitvorteile für das Säulendiagramm gegenüber der Tabelle sich vornehmlich dann auf ökologisch valide Ergebnisdarstellungen generalisieren lassen, wenn

- relativ klar ist, was die Säulen repräsentieren (z.B. wiederkehrende Präsentationen am Wahlabend)
- nur eine Datenreihe vorgegeben wird (konventionelles Säulendiagramm)
- diese Datenreihe viele Datenelemente enthält.

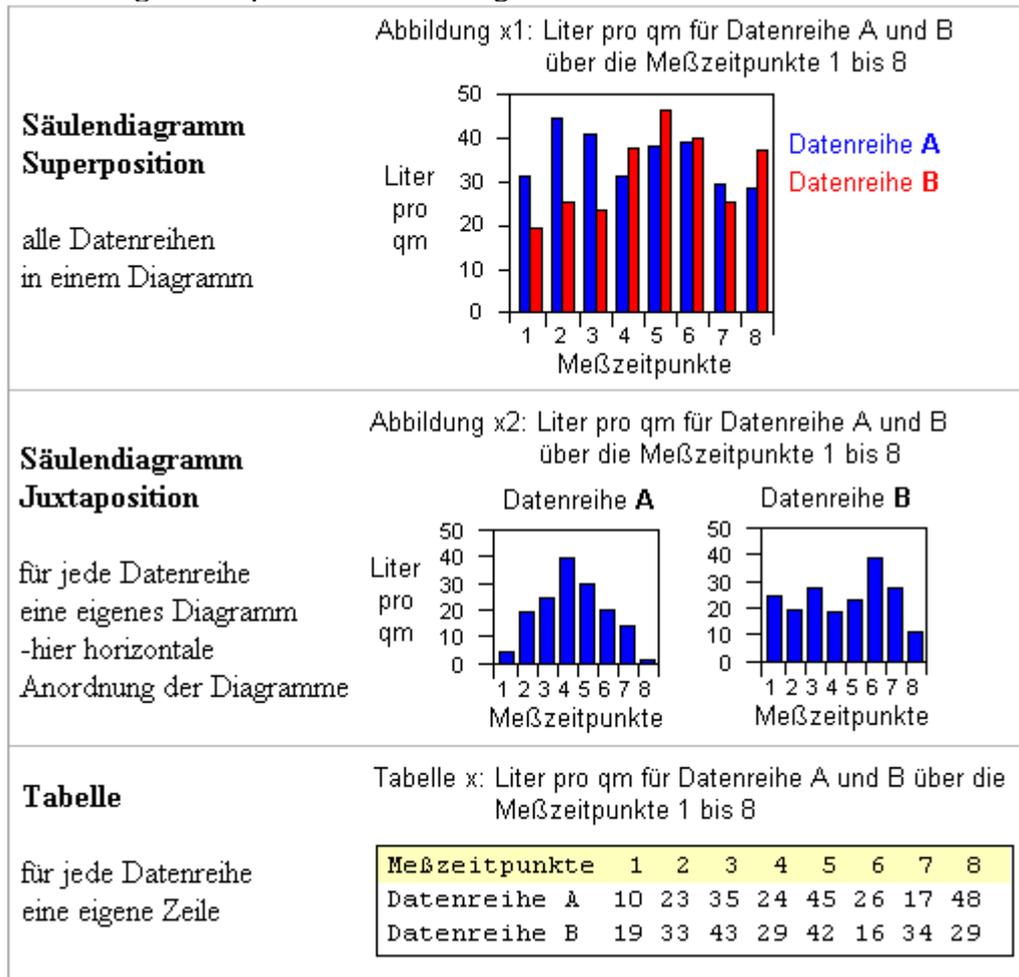
Die hier getestete Fragestellung bezieht sich auf größer/kleiner- Relationen zwischen den Daten und speziell auf die Identifikation des größten Wertes. Man kann aus den Ergebnissen auch ableiten, dass der Vorteil des

Säulendiagramms ansteigt, je mehr derartige Vergleiche angestellt werden müssen. Dies hat sich etwa in einer früheren Untersuchung gezeigt, als alle Datenwerte der Größe nach in eine Rangordnung gebracht werden mussten und wäre auch klar zu erwarten, wenn man beispielsweise nach dem zweitgrößten Wert in der Präsentation fragen würde. Bei der Frage nach dem kleinsten Wert, der in einem Säulendiagramm genau so schnell wie der höchste Wert identifiziert werden kann, sind ebenfalls ähnliche Unterschiede zur Tabelle zu erwarten.

Mehrere Datenreihen in einer Präsentation

Schwieriger wird es schon, wenn mehrere Datenreihen in der Präsentation enthalten sind. Abbildung 4 verdeutlicht den einfachsten Fall mit 2 Datenreihen für die Präsentationsformen Säulendiagramm und Tabelle.

Abbildung 4: Beispiele für Säulendiagramme und Tabelle mit 2 Datenreihen.



Meyer, Shinar. & Leiser (1997) konnten in 2 Experimenten konsistent nachweisen, dass man aus einer Präsentation mit 3 Datenreihen von jeweils 5 Datenwerten den größten Wert innerhalb einer bestimmten Datenreihe in einer Tabelle schneller identifizieren konnte als in einem superpositionierten Säulendiagramm (Fragebeispiel: "Wo liegt der höchste Wert für Datenreihe B?"). Die verlangte Anforderung setzt voraus, dass man zunächst die Datenreihe identifizieren muss, aus der heraus man die Größenvergleiche vornimmt. Diese Orientierungsleistung ist in der Tabelle einfacher zu leisten, insbesondere dann, wenn man in einem superpositionierten Säulendiagramm mit Legenden arbeitet und die Bezeichnung der Datenreihen nicht direkt in der Graphik anbringt, wie das Milroy & Poulton (1978) empfehlen und/oder die Unterscheidung der Säulen für jede Datenreihe suboptimal gestaltet. Diese Orientierung in der Graphik kann man aber wesentlich verbessern, wie aus den hier verwendeten Graphiken am Beispiel von Liniendiagrammen aus Abbildung 2 oder 3 hervorgeht. Auch in Abbildung 4 scheint mir etwa die Identifikation "Welche Daten repräsentieren die

der Datenreihe A" im superpositionierten Säulendiagramm besser zu gelingen als in einer konventionellen Legende. Aber so gut wie in einer Tabelle kann man die Zuordnung "Welche Variable repräsentiert welche Datenreihe, bzw. welche Daten gehören zu welcher Datenreihe" in einem superpositionierten Säulendiagramm meiner Meinung nach überhaupt nicht gestalten.

Lohse (1993, S. 220) hat mit Hilfe der Erfassung von Augenbewegungen z.B. festgestellt, dass man in einem superpositionierten Linien - bzw. Säulendiagramm über 1 Sekunde mehr Zeit als in einer Tabelle benötigt, um die Zuordnung der für die Fragestellung relevanten Variablenbezeichnungen zu den entsprechenden Datenreihen zu lesen. Hierbei lag eine sehr umfangreiche superpositionierte Datenpräsentation von 6 Datenreihen mit je 12 Datenwerten zugrunde. Unter dieser Voraussetzung ist es auch verständlich, dass ein einfacher Vergleich von 2 Daten im Hinblick auf die Relation größer in der Tabelle schneller vollzogen werden konnte als im Säulen- und Liniendiagramm (S.219). Demgegenüber fand Jacobs (1990), dass man schneller den größten Wert von 2 Säulen als von 2 Zahlen identifizieren kann, wenn die Zuordnung in beiden Präsentationsformen vergleichbar gut gelingt. Wie Lohse 1993 auf S. 220 mitteilt, ist die benötigte Zeit für den reinen Datenvergleich - also abzüglich des Lesens von Ordinate, Abszisse, Legende usw. - für das bessere der untersuchten Säulendiagramme mit 2.17 Sekunden auch kürzer als die entsprechende Zeit für die beste Tabelle mit 2.72 Sekunden.

In einem superpositionierten Säulendiagramm ist es schwieriger, den höchsten Wert einer bestimmten Datenreihe zu finden, da neben den eigentlich zu vergleichenden Säulen noch weitere für den Vergleich irrelevante Säulen stören. Verwendet man für die Datenreihen - statt monocromer Schraffuren- unterschiedlich farbige Säulen wie hier in [Abbildung 4](#), rechne ich dennoch mit einem Vorteil des Säulendiagramms. Bzgl. dieser Anforderung wären juxtaponierte Säulendiagramme höchstwahrscheinlich effizienter und hier dürfte der höchste Wert mindestens so schnell wie, bei optimaler Gestaltung der graphischen Präsentation und vielen Daten deutlich schneller als in einer Tabelle identifiziert werden können.

Dafür aber sind Größer/kleiner-Vergleiche zwischen den Datenreihen für die einzelnen Kategorien auf der x-Achse ("Fragebeispiel für [Abbildung 4](#): "Welche Datenreihe zeigt den höchsten Wert zum Messzeitpunkt 3") im juxtaponierten Säulendiagramm deutlich schlechter wahrnehmbar als unter Superposition. Das ist nach Cleveland (1985) auch zu erwarten, weil die Daten nun auf der Basis der Position an der gleichen, aber nicht gemeinsamen Skala (position along identical, nonaligned scale) abgeschätzt werden müssen. Die Unterlegenheit von Juxtaposition konnte von Jacobs (1995b) klar empirisch bestätigt werden (siehe Kurzfassung: [Welche Datenreihe zeigt bei einer bestimmten Rubrik den höchsten Wert?](#)). Der Größenvergleich ist unter Justaposition sicherlich auch schlechter erkennbar als in einer Tabelle. Meyer, Shinar. & Leiser (1997) fanden bzgl. dieser Fragestellung sogar schnellere Reaktionszeiten für die Tabelle als für das (meiner Meinung nach nicht optimal gestaltete,) superpositionierte Säulendiagramm. Mit wachsender Anzahl der Datenreihen wird die Anwendung des superpositionierten Säulendiagramms immer problematischer und etwa bei 32 Datenreihen lässt sich der größte Wert in einer Tabelle immerhin noch finden, aber ein Säulendiagramm dieser Art überhaupt nicht mehr darstellen.

Von Extremfällen abgesehen, kann man im Normalfall aber davon ausgehen. Je höher die Komplexität im Sinne der wachsenden Anzahl von Größer/kleiner-Vergleichen, desto eher wird sich das Säulendiagramm gegenüber der Tabelle trotz aller Zuordnungsprobleme auch in der Praxis durchsetzen und handelt es sich um wiederkehrende Präsentationen wird das Säulendiagramm in jedem Falle besser abschneiden. Coll (1992) gab recht komplexe Fragestellungen des Typs "relative information question" vor, deren Beantwortung im wesentlichen die Nutzung der Relation "kleiner, größer" erfordern und konnte für diese Fragestellungen statistisch signifikante Vorteile des Säulendiagramms gegenüber der Tabelle nachweisen, obwohl superpositionierte Säulendiagramme mit monocromen Säulen zur Anwendung kamen.

Gute Vergleichsmöglichkeiten visueller Repräsentationen reichen nicht aus

Optimale Wahrnehmungseigenschaften visueller Datenrepräsentationen für bestimmte Vergleiche (der eigentliche Grund für Wahl graphischer Präsentationen) konfliktieren gelegentlich mit sonstigen praktischen Darstellungsproblemen und erschweren dann die kognitive Orientierung. So können etwa die Rubriken (Kategorien) auf der X-Achse eines Säulendiagramms nicht viel Text aufnehmen, wenn viele Säulen dargestellt werden sollen oder wenn möglichst viele Säulen eng beieinander liegen sollen, damit die Vergleiche auch einfach vorzunehmen sind. Dies ist meiner Meinung nach der größte Nachteil des Säulendiagramms. Ist die Identifikation des Namens jeder Kategorie zum Verständnis der Daten essentiell, kann dieser Name aber infolge Platzmangels nicht deutlich lesbar gestaltet werden (und wird dann durch kryptische Abkürzungen ersetzt oder sind weitere Legenden erforderlich), dann stellt sich die Frage, ob sich die Anwendung des Säulendiagramms noch lohnt. Eine diesbzgl. bessere Orientierung liefert hier das Balkendiagramm (die Balken liegen hier horizontal untereinander) und für Bezeichnungstext der Rubriken ist genügend Platz vorhanden. Man kann aber annehmen, dass die Größenvergleiche von Balken in Balkendiagrammen ähnlich gut gelingen wie die Säulenvergleiche von Säulen in Säulendiagrammen. Eine andere Alternative zum Säulendiagramm wäre in diesem Fall das Dot-Chart von Cleveland (1985)

Dasselbe gilt für die Anordnung von Daten in einer Tabelle. Hier bietet die vertikale Anordnung der Daten in den Zeilen einer Spalte ähnliche Vorteile wie das Balkendiagramm zur Beschriftung der Kategorien. Zudem deutete sich an, dass die größte Zahl in vertikaler Darstellung der Zahlen etwas besser identifiziert werden konnte. Die günstigere Vergleichsmöglichkeit in der vertikalen Tabellendarstellung dürfte sich für andere Relationen (etwa Summen -oder Differenzbildung, Vergleiche zwischen den ersten Ziffern hinter dem Dezimalpunkt) noch wesentlich deutlicher herauskristallisieren. Allerdings könnte es erhebliche kognitive Verwirrung hervorrufen, einen Verlauf beispielsweise in vertikaler Tabellendarstellung anzuordnen, weil wir es gewohnt sind, Verläufe von links nach rechts zu betrachten. [Ehrenberg stellt auch Verläufe vertikal dar und zwar von oben beginnend (1) nach unten zunehmend (...10).]

Zieht man weiterhin ins Kalkül, dass sich verschiedene Präsentationsformen unterschiedlich gut für verschiedene Fragestellungen eignen, dann steht man bei der Auswahl einer Präsentation häufig vor einem Ambivalenzkonflikt, der nur durch die Setzung klarer Prioritäten entschieden werden kann. Die Klarheit dessen, für was die Daten stehen und die Fairness der Ergebnisdarstellung sind dabei unverzichtbare Forderungen und wenn es hier Bedenken gibt, diese Forderungen überzeugend in eine Graphik zu überführen, dann sollte man im Zweifelsfalle lieber eine Tabelle darstellen.

Hinweis: Im Dezember 1998 wurde das Experiment als [Gruppenversuch](#) mit 23 Vpn durchgeführt (Jacobs (1999))

Literatur

Ballstaedt, S. P. (1997). Wissensvermittlung. Die Gestaltung von Lernmaterial. Beltz. Psychologie Verlags Union. Weinheim.

Capoza, D. (?). Numeracy Explained: How to Make Sense of Data. (Online on Internet: <http://www.cstudies.ubc.ca/facdev/services/newsletter/88/oct88-5.html> [1.9.1998])

- Chaparakani, T.K. & Ehrenberg, A.S. (1976). Numerical information processing. London: Business School. (zitiert nach einer Email von Ballstaedt vom 26.8.1998)
- Cleveland, W. S. (1985). The Elements of Graphing Data. Monterey, California:Wadsworth advanced Books and Software.
- Coll, J. H. (1992). An experimental study of the efficacy of tables versus bar graphs with respect to type of task. *Information Management* 23, pp. 45-51
- Ehrenberg, A.S.C. (1977). Rudiments of numeracy. *Journal of the Royal Statistical Society, Series A*, 140, pp. 277-297.
- Ehrenberg, A.S.C. (1981) "The Problem of Numeracy," *The American Statistician*. 35(2) 67-71.
- Gillan, D. J., Wickens, C. D., Hollands, J. G., & Carswell, C. M. (1998). Guidelines for presenting quantitative data in HFES publications. *Human Factors*, 40, 28-41.
- 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, B. (1994). Graphische vs. tabellarische Präsentation von statistischen Daten. *Zeitschrift für Pädagogische Psychologie*, 8, 73-84. (Kurzfassung von Jacobs 1990)
- Jacobs, B. (1995a). Die Wahrnehmung besonderer Relationen in Säulendiagramm und Liniendiagramm unter Superposition und Juxtaposition. Saarbrücken. Arbeitsberichte des Medienzentrums der Universität des Saarlandes, Nr. 17. ([pdf-Datei](#) (625 KB))
- Jacobs, B. (1995b). Globale Vergleiche, lokale Vergleiche und Größenschätzungen in Liniendiagramm und Säulendiagramm unter Superposition und Juxtaposition. Saarbrücken. Arbeitsberichte des Medienzentrums der Universität des Saarlandes, Nr. 16. ([pdf-Datei](#) (260 KB))
- Jacobs, B. (1999). [Identifikation des größten Wertes in Tabelle und Säulendiagramm \(Gruppenversuch\)](#) . URL: <http://www.phil.uni-sb.de/~jakobs/wwwartikel/tabgraph/maxwertalle.htm> [4.1.1999]
- Lohse, G. L. (1993). Eye Movement-Based Analyses of Graphs and Tables: The next Generation. *Proceedings of the Fourteenth International Conference on Information Systems*. Orlando, FL. 213-224.
- Meyer, J., Shinar, D. & Leiser, D. (1997). Multiple factors that determine performance with tables und graphs. *Human Factors*, 1997,39(2), 268-286
- Milroy, R. & Poulton E.C. (1978). Labelling Graphs for Improved Reading Speed. *Ergonomics*, 21, 55-61.
- Natriello, G. (1997). Table Manners: Preparing Accessible Tables. *Teachers College Record*, Vol 98, Nr. 4 (online in Internet: <http://tcrecord.tc.columbia.edu/forrec/vol98/summer97.htm> [1.9.1998])
- Wainer, H. (1997). Improving tabular displays, With NAEP tables as examples and inspirations. *Journal of Educational and Behavioral Statistics*, 22, pp. 1-30.(zitiert nach Natriello, G. (1997)).
- Zwisler, R. (1993) Mentale Arithmetik ,(online in Internet [24.9.1998] unter URL: http://www.psychologie.uni-regensburg.de/~zwr02102/scripts/Mentale_Arithmetik.html

Identifikation des größten Wertes in Tabelle und Säulendiagramm (Gruppenversuch)

[Zusammenfassung](#)

[Ziel der Untersuchung](#)

[Versuchspersonen und Versuchsdurchführung](#)

[Validität und Reliabilität des Versuchs](#)

[Zusatzhypothese zum Trainingseffekt](#)

[Vergleich Tabelle vs. Säulendiagramm](#)

[Der Einfluss des Trainings](#)

[Der Einfluss der Zahlenanordnung in der Tabelle](#)

[Der Einfluss der Zifferanzahl](#)

[Zusammenfassung der wichtigsten Ergebnisse](#)

[weitere Hinweise](#)

Zusammenfassung

In einem früheren Selbstversuch wurde nachgewiesen, dass das Säulendiagramm der Tabelle beim Identifizieren des maximalen Wertes überlegen ist. Zudem ergaben sich gewisse Hinweise dafür, dass Zahlen in einer Tabelle besser miteinander verglichen werden können, wenn die Zahlen vertikal statt horizontal angeordnet sind. Die entsprechenden Hypothesen wurden hier in einem Gruppenexperiment erneut überprüft und konnten eindeutig bestätigt werden.

Ziel der Untersuchung

In einem früheren Artikel mit dem Titel [Identifikation des größten Wertes in Tabelle und Säulendiagramm](#) habe ich in das Thema eingeführt, die Versuchsdurchführung beschrieben und die Hypothesen näher expliziert. Die empirische Überprüfung wurde in Form eines mehrfach wiederholten Einzelfallexperimentes an mir selbst vorgenommen. Obgleich auf eine statistische Testung ganz verzichtet wurde, ergaben sich augenscheinlich massive, eindeutige Befunde in Richtung der zentralen Interaktionshypothese. Sie lautet, dass der Vorteil des Säulendiagramms gegenüber der Tabelle beim Identifizieren des maximalen Wertes mit wachsender Anzahl der zu vergleichenden Größenwerte ansteigt. Daneben wurde einige, auf Ehrenberg basierende Hypothesen zur Verbesserung von Tabellen getestet, die zwar in der erwarteten Richtung lagen, deren Ausmaß und Zuverlässigkeit allerdings nicht klar eingeschätzt werden konnte. Sinn dieser Ausführungen ist es, die damals aufgestellten Hypothesen hier auf konventionelle Weise in einem Gruppenversuch zu testen. Zum einen können so einzelne, weniger deutliche Ergebnisse auf ihre Zuverlässigkeit geprüft und die Generalisierbarkeit der Befunde im Hinblick auf die Vpn-Repräsentativität abgeschätzt werden. Die Ausführungen beschränken sich dabei vornehmlich auf die notwendigen Angaben zu Versuchspersonen und Versuchsdurchführung, sowie auf die Mitteilung der Ergebnisse, sofern ansonsten keine neuen Aspekte eingeführt werden.

Versuchspersonen und Versuchsdurchführung

Als Versuchspersonen dienten überwiegend Teilnehmer(innen) des vom Verfasser geleiteten Seminars Versuchsplanung. Hinzu kamen einige Student(inn)en bzw. wissenschaftliche Mitarbeiter(innen), die sich zur Teilnahme überreden ließen. Mit einem Teil der Vpn wurde

der Versuch im Computerübungsraum des Cip-Pools durchgeführt, ein Teil absolvierte das Experiment in Einzelversuchen. 4 Vpn besorgten sich das Experiment via Internet, führten es zu Hause am eigenen Computer durch und schickten die Ergebnisse via Attachment zu. Die wenigsten Versuchspersonen bearbeiteten das Experiment unter idealen Bedingungen. Es herrschten überwiegend mehr oder weniger normale Alltags- bzw. Arbeitsbedingungen. 4 Vpn wurden durch erheblichen Schlagbohrhammerlärm ("Durchbohren der Betondecke mit schwerem Bohrgerät") beeinträchtigt. Es gibt eben kein Grundrecht auf Ruhe, auch nicht an einer Universität. Letztlich ist es nur der besonderen Konzeption des Experiments zu verdanken, dass diese Widrigkeiten zwar stören, aber die interne Validität des Versuchs wenigstens nicht ernsthaft belasten. Der Versuch dauerte ca. 10 bis 15 Minuten. Am Versuch nahmen 25 Vpn teil. Nach Inspektion der Daten mit Boxplots lieferten 2 Versuchspersonen sehr signifikante Ausreißer (extrem überlange Zeiten) auf mehreren Variablen. Ihre Ergebnisse wurden deshalb vollständig aus der Datenbank entfernt, so dass zur Auswertung die Daten von 23 Vpn zugrunde liegen. Alle hier durchgeführten Analysen basieren folglich auf einem N von 23.

Validität und Reliabilität des Versuchs

Aus der Beschreibung des [Versuchsaufbaus](#) geht hervor, dass die Versuchsplanvalidität durch ein aufwendiges Messwiederholungsexperiment gesichert wurde. Wie schon früher ausgeführt, ist auch die Kontentvalidität der Daten für den Gesamtversuch als sehr gut einzuschätzen, weil die Größenwerte innerhalb jeder Präsentation stets Zufallsstichproben aus der nach gewissen Kriterien eingeschränkten Grundmenge darstellen und somit insgesamt einen sehr großen Bereich möglicher Datendarstellungen repräsentieren.

Jede Bedingung wurde mehrfach getestet. Betrachtet man die Präsentationen als Items und den Durchschnitt der Items unter einer Bedingung als Testwert, so lassen sich für die Zeiten Reliabilitäten berechnen, die hier der Einfachheit halber als Korrelationen zwischen erstem und zweiten Durchgang im Form von Splithalf-Reliabilitäten berechnet wurden.

Tabelle 1: Splithalf-Reliabilitäten der Zeiten für die wichtigsten erhobenen Bedingungen des Gesamtversuchs

Anzahl der Größenwerte	4	8	16	32
Tabelle	.82	.90	.91	.84
Säulendiagramm	.81	.94	.84	.67

Erklärungsbeispiel: Die Bedingung "aus 8 möglichen Daten den größten Wert in der Tabelle zu finden" wurde im Gesamtversuch 12-mal getestet. Die Reliabilität für den Gesamtwert dieser Bedingung beträgt .90.

Wie aus der Tabelle 1 hervorgeht, lassen sich Personen hinsichtlich ihrer Schnelligkeit durch die erhobenen Bedingungen recht gut differenzieren. Zum anderen könnte man die Korrelationen auch so deuten, dass hier die einzelnen Stufen der UV's recht zuverlässig implementiert wurden.

Zusatzhypothese zum Trainingseffekt

Im Computereperiment wird der gleiche Versuch zweimal hintereinander durchgeführt, was die Vpn gar nicht bemerken. Im ersten Versuch werden 37 Präsentationen bearbeitet. Mithin kann überprüft werden, ob dadurch im nachfolgenden Versuch Zeitgewinne erzielt werden. Meyer, Shinar und Leiser (1997) konnten entsprechende Trainingseffekte für Tabelle, Säulendiagramm und Liniendiagramm nachweisen. Besonders kritisch wäre die Entde-

ckung einer Interaktion zwischen den Präsentationsformen und dem Training zu bewerten, weil man dann behaupten könnte, es würden 2 Präsentationsformen miteinander verglichen, bei denen das Leistungspotential der einen Präsentation noch längst nicht erreicht wäre. Ein fairer Vergleich von Präsentationsformen setzt aber vergleichbare Erfahrung mit bzw. hinreichende Beherrschung der beiden Präsentationsformen voraus. Es hat ja z.B. auch wenig Sinn, die Auswirkungen von gesprochener Sprache bzw. gedrucktem Text auf das Lernen bei Kindern mit mangelnder Lesefähigkeit zu testen. Das Problem ist aus der Medienforschung bekannt und wird ständig ignoriert, in dem man etwa Präsentationsvergleiche mit Computernovizen zwischen Hypertext und konventionellem Lernmedium durchführt. Es wird ein Trainingseffekt erwartet, dessen Ausmaß aber nicht eingeschätzt werden kann. Zudem wird erhofft, dass keine Interaktion zwischen Präsentationsform und Training vorliegt.

Vergleich Säulendiagramm vs. Tabelle

Neben dem klar erwarteten Einfluss der Präsentationsform sowie der Anzahl der zu vergleichenden Daten, steht im Mittelpunkt der Überprüfung die zentrale Interaktionshypothese: **Der Vorteil des Säulendiagramms gegenüber der Tabelle beim Identifizieren des maximalen Wertes wächst mit zunehmender Anzahl der zu vergleichenden Datenwerte** (=ordinale Interaktion zwischen Präsentationsform und Anzahl der Daten). Das ist die wichtigste Fragestellung dieser Untersuchung.

Anzahl der Fehler

Da insgesamt recht wenige Fehler gemacht wurden, die Verteilung der Fehlerwerte dadurch notgedrungen ganz massiv von der Normalverteilung abweicht und erhebliche Varianzheterogenität vorliegt, wurde auf eine varianzanalytische Berechnung der Fehlerwerte verzichtet. Tabelle 2 zeigt die Ergebnisse der einzelnen Bedingungskonstellationen.

Tabelle 2: Fehler in Prozent für Tabelle und Säulendiagramm beim Identifizieren des maximalen Größenwertes aus 4, 8, 16 und 32 Datenelementen

Anzahl der Daten	4	8	16	32	Durchschnitt
Tabelle:	3	6	11	14	8
Säulendiagramm:	1	1	1	1	1
Wilcoxon z= :	1.4	2.5	2.7	3.0	
5% signifikant:	ns	*	*	*	

Augenscheinlich liegt die erwartete Interaktion für die gemachten Fehler in der erwarteten Richtung. Die Fehlerprozentage für die Tabelle steigen mit wachsender Anzahl der Daten an, während die Fehler im Säulendiagramm konstant bei 1% verharren. Ab 8 Daten aufwärts werden in der Tabelle signifikant mehr Fehler gemacht als im Säulendiagramm. Weil die Testung hier mit einem nichtparametrischen Verfahren durchgeführt wurde, liegt nur ein Signifikanzniveau von 5% zugrunde.

benötigte Zeit

Wegen der relativ geringen Fehlerzahl werden die Zeiten grundsätzlich aus allen Präsentationsvorgaben verwendet. Es spielt demnach keine Rolle, ob die Frage korrekt oder falsch beantwortet wurde. Spearman-Rangkorrelationen zwischen Fehler und Zeiten von $r = -.35$ für die Tabelle und von $r = .27$ für das Säulendiagramm sind beide insignifikant. Die Entscheidungszeit hat demnach keinen Einfluss auf das Ausmaß der Fehler. Mithin kann der Fehler bei der Auswertung der Zeiten vernachlässigt werden.

Zur gemeinsamen Überprüfung mehrerer Hypothesen wird der Versuchsplan als 3- faktorielles Design mit Messwiederholung auf allen Faktoren gedeutet.

1. **Präsentationsform** mit den Stufen Tabelle und Säulendiagramm
2. **Anzahl der zu vergleichenden Daten** mit den Stufen 4, 8, 16 und 32 Größenwerte
3. **Replikation** mit den Stufen 1. Durchgang und 2. Durchgang zur Abschätzung von Trainingseffekten.

Wegen Verletzung der Anwendungsvoraussetzungen wird, wo immer nötig, nach Geenhouse-Geiser korrigiert.

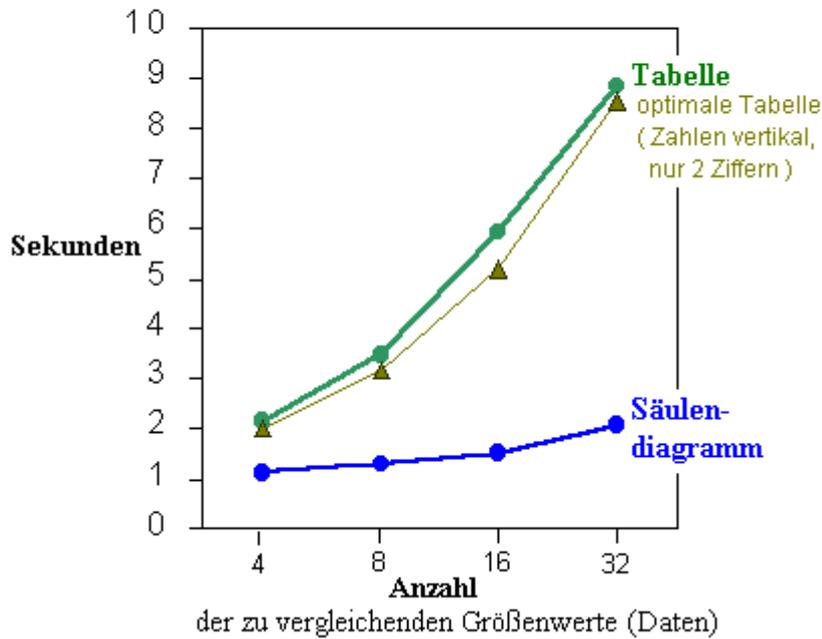
Tabelle	3:	Ergebnis	der			Varianzanalyse
Zeiten für die Fragestellung: Identifiziere den maximalen Größenwert!						
		df	F	p	Signifikanz	
Präsentation (Tabelle, Graphik)	(1 , 22)	230.0	.000	*		
Anzahl der Daten (4,8,16,32)	(1.2, 27.4)	125.0	.000	*		
Replikation (1.,2. Durchgang)	(1 , 22)	9.4	.006	*		
Präsentation x Anzahl der Daten	(1.3, 28)	78.7	.000	*		
Präsentation x Replikation	(1 , 22)	0.3	.61	ns		
Replikation x Anzahl der Daten	(2 , 44)	1.7	.2	ns		
Präsentation x Anzahl x Replik.	(1.6, 36)	0.5	.59	ns		

Anmerkung: df (Zähler,Nenner) korrigiert nach Geenhouse-Geiser, * = signifikant auf dem 1% Niveau

Alle Befunde liegen in der erwarteten Richtung. Zunächst soll die überzeugend nachgewiesene Interaktion zwischen Präsentationsform und Anzahl der Daten in Form von Tabellen und Graphiken näher spezifiziert werden:

Abbildung 1

Benötigte Zeit zur Beantwortung der Fragestellung:
Identifiziere den maximalen Größtenwert!
 Mittelwerte von 23 Vpn; repeated Measurement



Während die Zeiten für die Tabelle mit wachsender Anzahl der zu vergleichenden Größenwerte steil ansteigen, ist der Anstieg im Säulendiagramm sehr flach.

Die grün dargestellte Tabelle beinhaltet alle getesteten tabellarischen Präsentationen und umfasst Zahlen mit 2,3 und 4 Ziffern.

Der Verlauf für die nach Ehrenberg optimale Tabelle - einer Teilmenge der grün dargestellten Tabelle - weicht nur geringfügig von der normalen Tabelle ab.

Die Übereinstimmung mit dem früher durchgeführten [Einzelfallexperiment](#) sowie einem späteren [Replikationsexperiment](#) im Jahre 2004 ist hervorragend

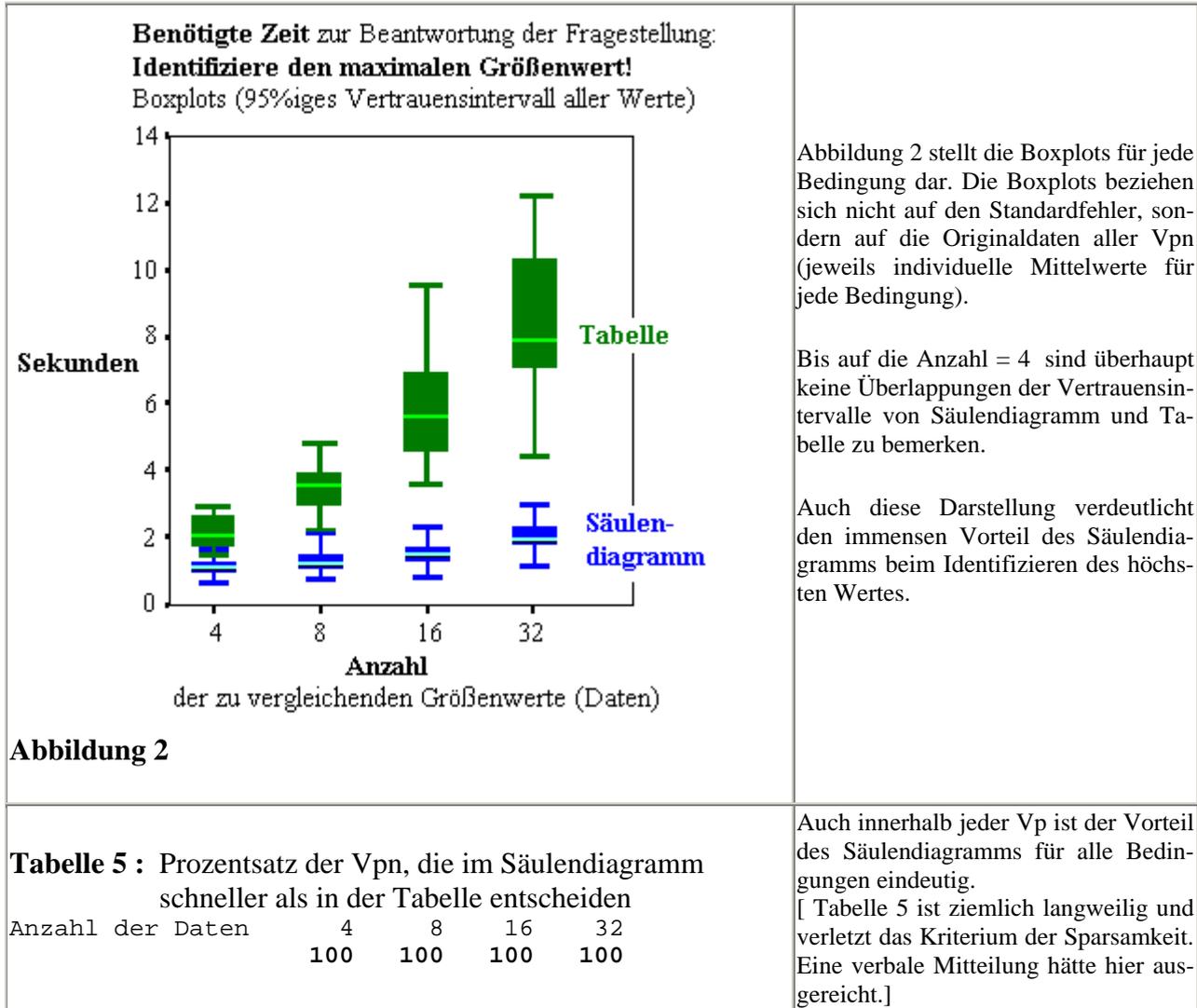
Tabelle 4: Mittelwerte M und Standardabweichung s für die Zeiten in **Sekunden**

Anzahl der Daten	4	8	16	32	Durchschnitt
Tabelle M	2.1	3.5	5.9	8.9	5.1
s	0.5	0.7	1.6	3.1	1.3
Säulendiagramm M	1.1	1.3	1.5	2.0	1.5
s	0.3	0.4	0.4	0.5	0.4
Effektstärke d	2.6	4.0	3.8	3.1	3.8

Anmerkung: Die der Effektstärke zugrunde liegende Streuung ist die gemeinsame Streuung für Tabelle und Säulendiagramm

Tabelle 4 lässt abschätzen, dass die in der Graphik erkennbaren Unterschiede in der Tat sehr bedeutsam sind.

Auf multiple Mittelwertsvergleiche wird verzichtet und stattdessen gleich die Effektstärken mitgeteilt, aus denen erkennbar ist, dass das Säulendiagramm bei jeder Bedingung der Tabelle eindeutig überlegen ist.



Die Unterschiede zwischen den Präsentationsformen sind derart massiv, dass man nur schwerlich den Eindruck verdrängen kann, was hier untersucht worden war, sei mehr oder weniger ziemlich trivial. Dass sich diese Trivialität jedoch so wunderbar manifestierte, hängt mit dem methodischen Vorgehen zusammen. Keine einzige Person hat bereits ab 4 zu vergleichenden Datenwerten den maximalen Wert in der Tabelle eher gefunden als im Säulendiagramm. Diese Aussage bezieht sich die Erwartungswerte (- wenn man so will auf die wahren Werte -), die hier als Mittelwerte aus mehreren Präsentationen geschätzt bzw. berechnet wurden und deren Reliabilität in dieser Untersuchung recht ordentlich einzuschätzen ist. Man stelle sich vor, den Versuchspersonen nur ca. 3 Items mit relativ wenigen zu vergleichenden Daten vorgegeben und den Versuch nicht als Messwiederholungsdesign, sondern mit unabhängigen Stichproben gemacht zu haben, wie das in der Graphforschung nicht selten geschieht.

Es kann kein Zweifel darüber bestehen, dass man mindestens ab 4 Datenwerten den maximalen Größtenwert eher anhand der Höhe von Rechtecken als auf der Basis der Größe von Zahlen identifizieren kann. Wenn andere Forscher, z.B. Meyer, Shinar & Leiser (1997) Ergebnisse mitteilen, wonach die Tabelle beim Identifizieren des maximalen Wertes aus 5 Datenwerten günstigere Ergebnisse liefert als das Säulendiagramm, dann muss dies an Faktoren liegen, die nichts mit dem eigentlichen Wahrnehmungs- bzw. Vergleichsprozess zu tun haben, sondern die diesem vorausgehen. (siehe dazu: [frühere Diskussion zur Generalisierung der Ergebnisse.](#))

Der Einfluss des Trainings

Wie erwartet, erzielten die Probanden im zweiten Durchgang etwas schnellere Ergebnisse (siehe Haupteffekt Replikation in [Tabelle 3](#)). Der signifikante Zeitgewinn ist in Tabelle und Säulendiagramm absolut gesehen vergleichbar, was auch durch die insignifikante Wechselwirkung zwischen Präsentationsform und Replikation zum Ausdruck kommt. Das Ausmaß des Zeitgewinns von jeweils ca. 0,2 Sekunden muss allerdings unabhängig von Signifikanz-erwägungen als recht schwach eingestuft werden.

Der Einfluss der Zahlenanordnung in der Tabelle

Da die 32 Daten umfassende tabellarische Bedingung nur in der vertikalen Anordnung getestet wurde, beschränkt sich der Vergleich der Zahlenanordnungen in der Tabelle nur auf Präsentationen bis zu 16 Daten. Tabelle 6 zeigt die Fehler für die einzelnen Bedingungen.

Tabelle 6: Fehler in Prozent

Identifizieren des maximalen Größenwertes aus 4, 8 und 16 Datenelementen für vertikale und horizontale Anordnung der Zahlen in der Tabelle

Anzahl der Daten	4	8	16	Durchschnitt
vertikal:	1	4	10	5
horizontal:	5	7	12	8

Die statistische Testung beschränkt sich auf den Durchschnitt. Der z-Wert des Wilcoxon-Tests von 2.3 unterbietet das erforderliche Signifikanzniveau von 5% zugunsten der vertikalen Zahlenanordnung in der Tabelle. Um den erwarteten Zeitvorteil der vertikalen Zahlenanordnung gegenüber der horizontalen Zahlenanordnung in der Tabelle umfassend zu überprüfen, wurde der Versuch als 2 faktorieller Versuchsplan mit Messwiederholung auf allen Faktoren gedeutet:

1. Anordnung der Zahlen in der Tabelle (horizontal, vertikal)
2. Anzahl der zu vergleichenden Daten (4,8,16)

Tabelle 7: Zweifaktorielle Varianzanalyse mit den Faktoren Tabellenanordnung und Anzahl der Daten (Zeit in Sekunden)

	df	F	p	Signifikanz
Anordnung (horizontal,vertikal)	(1, 22)	19.7	.000	*
Anzahl der Daten (4,8,16)	(1, 24)	169.0	.000	*
Anordnung x Anzahl der Daten	(1.5, 32)	7.1	.006	*

Entscheidend ist hier nur der signifikante Nachweis der Tabellenanordnung sowie die knapp signifikante Bestätigung der Interaktion von Tabellenanordnung und Anzahl der Daten. Die genaue Bedeutung der Unterschiede geht aus Tabelle 8 hervor.

Tabelle 8: Der Einfluss der Zahlenanordnung in der Tabelle
Mittelwerte M und Standardabweichung s für die Zeiten zum Erkennen des maximalen Wertes aus 4, 8 und 16 Daten bei vertikaler und horizontaler Anordnung der Zahlen in der Tabelle.

Anzahl der Daten		4	8	16	Durchschnitt
vertikal	M	2.0	3.3	5.5	3.6
	<i>s</i>	0.6	0.8	1.9	1.0
horizontal	M	2.2	3.6	6.4	4.1
	<i>s</i>	0.4	0.7	1.5	0.8
t(22)		2.6	2.6	3.9	
1% Signifikanz		*	*	*	
Effektstärke d		.35	.41	.51	

Wie aus Tabelle 8 hervorgeht, ist es erwartungsgemäß etwas leichter, die Daten in vertikaler Anordnung in einer Tabelle zu vergleichen. Um die Unterschiede zwischen den Tabellenanordnungen gesondert für jede Anzahl der zu vergleichenden Daten auf Signifikanz zu testen, wäre eigentlich ein multipler Mittelwertsvergleich notwendig gewesen, den SPSS 8.0 für das zugrunde gelegte Design aber leider nicht rechnet. Aus diesem Grunde wurden vereinfachend und - für manchen Methodiker unannehmbar - 3 T-Tests für abhängige Stichproben durchgeführt, um die Unterschiede dann auch noch ohne Alpha-Adjustierung einseitig auf dem 1% Niveau gegeneinander zu testen.

Konzentriert man sich auf die Effektstärken, so weist die vertikale Zahlenanordnung in der Tabelle einen schwachen bis mittleren Vorteil gegenüber der horizontalen Anordnung aus. Es sei nochmals darauf hingewiesen, dass die Bedingung für die horizontale Anordnung ausgesprochen günstig konzipiert war, weil der Abstand zwischen den Zahlen möglichst gering gewählt wurde. In realen Tabellen kann diese Spaltenenge häufig nicht zur Anwendung kommen, weil dann zu wenig Platz zur Kennzeichnung der Variablen bleibt. Insofern wird die praktische Bedeutsamkeit des Vorteils der vertikalen Anordnung hier unterschätzt.

Der Einfluss der Zifferanzahl

Eine einfaktorielle Varianzanalyse, bei denen die Anzahl der Ziffern die Stufen der UV und die Zeit die AV darstellten, ergab einen hochsignifikanten F-Wert von 38.2, der mit Geenhouse-Geiser korrigierten Freiheitsgraden [$F(1.5,33.2) = 38.2$] das erforderliche Signifikanzniveau von 1% klar unterschreitet. Alle möglichen statistischen Vergleiche zwischen den Zeiten für die Zahlen mit verschiedener Zifferanzahl unterscheiden sich signifikant auf dem 1 % Niveau (nach Bonferroni Mehrfachvergleichsanpassung, die SPSS 8 für dieses Design gnädig bereitstellt).

Tabelle 9: Der Einfluss der Zifferanzahl

Mittelwerte **M** und Standardabweichung **s** für die Zeiten zum Erkennen des maximalen Wertes bei 2, 3 und 4 ziffrigen Zahlen

Anzahl der Ziffern	Zeit in Sekunden	
	M	s
2	3.4	0.8
3	3.9	0.9
4	4.3	1.0

Man erkennt aus der Tabelle 9, dass jede weitere Ziffer die Zeit um ca. 0.5 Effektstärken anwachsen lässt. Bei 2 gegenüber 4 Ziffern ist das schon eine Standardabweichung und bedenkt man, dass man aus einer Tabelle vielleicht noch weitere Beziehungen entdecken will, so sind schnell einige Sekunden beisammen. Die Empfehlung von Ehrenberg, die Ziffergröße der Zahlen, wo immer möglich, auf 2 signifikante Ziffern zu reduzieren, kann daher nur nachdrücklich unterstützt werden und wurde in diesem Artikel ziemlich konsequent umgesetzt.

Zusammenfassung der wichtigsten Ergebnisse

Unter der Voraussetzung, es existiere in einer Präsentation ein eindeutig wahrnehmbarer maximaler Wert und dem Leser sei unmittelbar mit der Vorgabe einer Datenpräsentation völlig klar

- welche Daten er vergleichen soll,
- im Hinblick auf welche Fragestellung der Vergleich vorzunehmen ist,
- und dass der Vergleich möglichst schnell aber dennoch genau entschieden werden soll,

dann ist folgendes zu erwarten:

1. Die Identifizierung des maximalen Wertes kann in einem Säulendiagramm schneller und genauer als in einer Tabelle vorgenommen werden. Der Schnelligkeitsvorteil des Säulendiagramms ist bereits ab 4 zu vergleichenden Datenwerten eindeutig und wächst sehr deutlich mit steigender Anzahl der zu vergleichenden Größenwerte. Die Zeitunterschiede zwischen den Präsentationsformen sind, gemessen an der Effektstärke, außerordentlich hoch.
2. Es ist einfacher, Zahlen tabellarisch in vertikaler als in horizontaler Anordnung zu vergleichen. Zumindest gilt dies für die Identifikation des maximalen Wertes. Das Ausmaß des Unterschieds liegt im niedrigen bis mittleren Bereich. Die Zahlenvergleiche werden mit zunehmender Anzahl der Zahlensziffern schwieriger. Alle Ergebnisse stützen die Vorschläge von Ehrenberg, Zahlen für Präsentationszwecke vornehmlich vertikal anzuordnen und auf möglichst wenige signifikante Ziffern zu runden.

Hinweise

- Die [Literaturangaben](#) sind aus dem früheren Artikel zu entnehmen.
- Das Computerexperiment findet der interessierte Leser unter <http://www.phil.uni-sb.de/~jakobs/wwwartikel/tabgraph/maxwert.htm>.
- Ein [weiterer Gruppenversuch \[2004\]](#) konnte die hier gefundenen Ergebnisse replizieren.

Autor: [Bernhard Jacobs](#), Medienzentrum der Philosophischen Fakultäten der Universität des Saarlandes created: 21.1.2004

URL des Originals: <http://www.phil.uni-sb.de/~jakobs/wwwartikel/tabgraph/maxwertalle2004.htm>

[Theorie, Hypothesen und Einzelfallexperiment](#) | [früherer Gruppenversuch Jacobs \(1999\)](#)

Identifikation des größten Wertes in Tabelle und Säulendiagramm [Replikationsexperiment 2004]

Zusammenfassung

In einem Selbstversuch sowie einem darauf folgenden Gruppenversuch wurde nachgewiesen, dass das Säulendiagramm der Tabelle beim Identifizieren des maximalen Wertes deutlich überlegen ist. Zudem ergaben sich gewisse Hinweise dafür, dass Zahlen in einer Tabelle besser miteinander verglichen werden können, wenn die Zahlen vertikal statt horizontal angeordnet sind. Die entsprechenden Hypothesen wurden hier in einem weiteren Gruppenexperiment erneut überprüft und konnten eindeutig bestätigt werden.

Ziel der Untersuchung

In einem früheren Artikel mit dem Titel [Identifikation des größten Wertes in Tabelle und Säulendiagramm](#) habe ich in das Thema eingeführt, die Versuchsdurchführung beschrieben und die Hypothesen näher expliziert. Im Mittelpunkt der Überlegungen stand die Hypothese, der Vorteil des Säulendiagramms gegenüber der Tabelle beim Identifizieren des maximalen Wertes steige mit wachsender Anzahl der zu vergleichenden Größenwerte an. In einem sich anschließenden [Gruppenversuch](#) wurde diese Interaktionshypothese überzeugend empirisch untermauert. Daneben konnten einige, auf Ehrenberg basierende Hypothesen zur Verbesserung von Tabellen belegt werden.

Obleich eine erneute Replikation wegen des rigorosen Versuchsaufbau und der Klarheit der Befunde eigentlich nicht zwingend notwendig erscheint, habe ich die sich bietende Gelegenheit genutzt, den Versuch erneut durchzuführen. Er könnte insbesondere die Zuverlässigkeitsbewertung bei solchen Fragestellungen verbessern, die im vorherigen Gruppenversuch nur kleine Unterschiede hervorbrachten, etwa die, ob vertikal angeordnete Zahlen besser zu vergleichen sind als horizontal angeordnete Zahlen.

Versuchspersonen und Versuchsdurchführung

Als Versuchspersonen dienten 28 Teilnehmer(innen) des vom Verfasser geleiteten Seminars Versuchsplanung (WS 2003/2004). Im Rahmen eines Arbeitsauftrags wurden die TeilnehmerInnen verpflichtet, den Versuch am Computer durchzuführen. Zu diesem Zweck habe ich [die notwendigen Dateien](#) via Email-Attachment an die StudentInnen verschickt mit der Aufforderung, das Experiment bei sich zu Hause oder sonst wo selbstständig durchzuführen und mir die Daten anonymisiert bis zu einem bestimmten Abgabetermin zuzusenden. Die Probanden erhielten von mir keinerlei spezielle Informationen oder Anleitungen, wie sie beim Versuch im einzelnen vorgehen sollten. Alle notwendigen Instruktionen waren dem Computereperiment zu entnehmen.

Diese Art der Versuchsdurchführung unterscheidet sich natürlich von einem kontrollierten Vorgehen im Labor. Dennoch machen die Daten einen sehr vertrauenswürdigen Eindruck. Sie wurden per Augenschein auf gravierende Ausreißer abgecheckt. Anschließend wurde entschieden, alle Daten zu verwenden. Alle hier durchgeführten Analysen basieren folglich auf einem N von 28.

Die Ergebnisdarstellung wird recht knapp gehalten, wobei aus Vergleichsgründen die Tabellennummerierung vom [vorherigen Gruppenversuch](#) übernommen und zusätzlich mit [2004] gekennzeichnet wurden.

Vergleich Säulendiagramm vs. Tabelle

Neben dem klar erwarteten Einfluss der Präsentationsform ("Säulendiagramm günstiger als Tabelle") sowie der Anzahl der zu vergleichenden Daten ("mit zunehmender Datenanzahl mehr Fehler und Zeitbedarf"), steht im Mittelpunkt der Überprüfung die zentrale Interaktionshypothese: **Der Vorteil des Säulendiagramms gegenüber der Tabelle beim Identifizieren des maximalen Wertes wächst mit zunehmender Anzahl der zu vergleichenden Datenwerte** (=ordinale Interaktion zwischen Präsentationsform und Anzahl der Daten).

Anzahl der Fehler

Tabelle 2 zeigt die Fehler für die einzelnen Bedingungskonstellationen.

Tabelle 2 [2004]: Fehler in Prozent für Tabelle und Säulendiagramm beim Identifizieren des maximalen Größenwertes aus 4, 8, 16 und 32 Datenelementen

Anzahl der Daten	4	8	16	32	Durchschnitt
Tabelle:	5.7	12.6	17.3	17.9	12.7
Säulendiagramm:	0.4	0.9	1.8	1.8	1.2
Wilcoxon z= :	3.5	4.3	3.9	3.3	
5% signifikant:	*	*	*	*	

Augenscheinlich liegt die Interaktion in der erwarteten Richtung, wenngleich man noch höhere Fehlerwerte bei der 32 Zahlen umfassenden Tabelle erwartet hätte. Die Fehlerprozent für die Tabelle steigen mit wachsender Anzahl der Daten deutlich an, während sich die Fehler im Säulendiagramm nur marginal erhöhen. Bereits ab 4 Daten aufwärts werden in der Tabelle signifikant mehr Fehler gemacht als im Säulendiagramm. Weil die Testung hier mit einem nichtparametrischen Verfahren durchgeführt wurde, liegt nur ein Signifikanzniveau von 5% zugrunde.

Die Fehlerwerte liegen in dieser Untersuchung bei der Tabelle etwas höher als im ersten Gruppenexperiment. Dies könnte daran liegen, dass die Teilnehmer die Aufgaben mit etwas weniger Sorgfalt bearbeitet haben, da sie das Experiment nicht im Labor absolvierten und einige Teilnehmer offenbar in Terminnot gerieten.

benötigte Zeit

Zur gemeinsamen statistischen Überprüfung mehrerer Hypothesen wird der Versuchsplan als 3- faktorielles Design mit Messwiederholung auf allen Faktoren gedeutet.

1. **Präsentationsform** mit den Stufen Tabelle und Säulendiagramm
2. **Anzahl der zu vergleichenden Daten** mit den Stufen 4, 8, 16 und 32 Größenwerte
3. **Replikation** mit den Stufen 1. Durchgang und 2. Durchgang zur Abschätzung von Trainingseffekten.

Wegen Verletzung der Anwendungsvoraussetzungen wird, wo immer nötig, nach Geenhouse-Geiser korrigiert.

Tabelle 3: Ergebnis der Varianzanalyse [2004]

Zeiten für die Fragestellung: Identifiziere den maximalen Größenwert!

	df	F	p	Signifikanz
Präsentation (Tabelle, Graphik)	(1, 27)	252.0	.000	*
Anzahl der Daten (4,8,16,32)	(1.5, 41)	183.0	.000	*
Replikation (1.,2. Durchgang)	(1, 27)	6.7	.015	ns
Präsentation x Anzahl der Daten	(1.4, 37)	133.0	.000	*
Präsentation x Replikation	(1, 27)	0.0	.96	ns
Replikation x Anzahl der Daten	(1.5, 41)	0.2	.79	ns
Präsentation x Anzahl x Replik.	(1.8, 36)	0.2	.77	ns

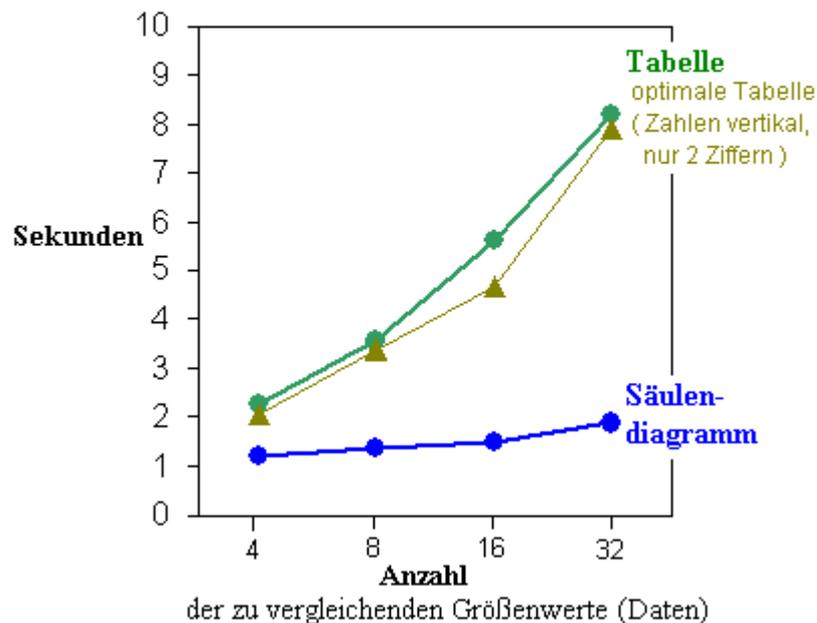
Anmerkung: df (Zähler,Nenner) korrigiert nach Geenhouse-Geiser, * = signifikant auf dem 1% Niveau

Abbildung 1 [2004]

Replikationsexperiment 2004

Benötigte Zeit zur Beantwortung der Fragestellung:**Identifiziere den maximalen Größenwert!**

Mittelwerte von 28 Vpn; repeated Measurement



Sowohl die statistischen Ergebnisse der dreifaktoriellen Varianzanalyse, wie die Verdeutlichung der wichtigsten Beziehungen in Abbildung 1 entsprechen in hervorragendem Maße den früheren Befunden.

Der Einfluss des Trainings

Wie erwartet, erzielten die Probanden im zweiten Durchgang etwas schnellere Ergebnisse. Der Haupteffekt Replikation in [Tabelle 3](#) unterbietet aber knapp nicht das erforderliche Signifikanzniveau von 1%. Der Zeitgewinn ist in Tabelle und Säulendiagramm absolut gesehen vergleichbar, was auch durch die insignifikante Wechselwirkung zwischen Präsentationsform und Replikation zum Ausdruck kommt. Das Ausmaß des Zeitgewinns von jeweils ca. 0,18 Sekunden (im vorhergehenden Gruppenversuch 0,20 Sekunden) muss allerdings unabhängig von Signifikanzenerwägungen als recht schwach eingestuft werden.

Der Einfluss der Zahlenanordnung in der Tabelle [2004]

Da die 32 Daten umfassende tabellarische Bedingung nur in der vertikalen Anordnung getestet wurde, beschränkt sich der Vergleich der Zahlenanordnungen in der Tabelle nur auf Präsentationen bis zu 16 Daten. Tabelle 6 zeigt die Fehler für die einzelnen Bedingungen.

Tabelle 6 [2004]: Fehler in Prozent

Identifizieren des maximalen Größenwertes aus 4, 8 und 16 Datenelementen für

vertikale und horizontale Anordnung der Zahlen in der Tabelle

Anzahl der Daten	4	8	16	Durchschnitt
vertikal:	2	8	16	9
horizontal:	10	17	17	14

Die statistische Testung bezieht sich auf den Durchschnitt. Der z-Wert des Wilcoxon-Tests von $z=2.6$ unterbietet das erforderliche Signifikanzniveau von 5% zugunsten der vertikalen Zahlenanordnung in der Tabelle. Auch hier findet man eine sehr gute Übereinstimmung zum früheren Gruppenversuch.

Um den erwarteten Zeitvorteil der vertikalen Zahlenanordnung gegenüber der horizontalen Zahlenanordnung in der Tabelle umfassend zu überprüfen, wurde der Versuch als 2 faktorieller Versuchsplan mit Messwiederholung auf allen Faktoren gedeutet:

1. Anordnung der Zahlen in der Tabelle (horizontal, vertikal)
2. Anzahl der zu vergleichenden Daten (4,8,16)

Tabelle 7 [2004]: Zweifaktorielle Varianzanalyse mit den Faktoren Tabellenanordnung und Anzahl der Daten (Zeit in Sekunden)

	df	F	p	Signifikanz
Anordnung (horizontal,vertikal)	(1, 27)	41.6	.000	*
Anzahl der Daten (4,8,16)	(1.4, 38)	207.0	.000	*
Anordnung x Anzahl der Daten	(1.7, 46)	5.9	.007	*

Entscheidend ist hier nur der signifikante Nachweis der Tabellenanordnung sowie die knapp signifikante Bestätigung der Interaktion von Tabellenanordnung und Anzahl der Daten. Die genaue Bedeutung der Unterschiede geht aus Tabelle 8 hervor.

Tabelle 8 [2004]: Der Einfluss der Zahlenanordnung in der Tabelle (2004)

Mittelwerte M und Standardabweichung s für die Zeiten zum Erkennen des maximalen Wertes aus 4, 8 und 16 Daten bei vertikaler und horizontaler Anordnung der Zahlen in der Tabelle.

Anzahl der Daten	4	8	16	Durchschnitt
vertikal M	2.2	3.3	5.1	3.5
s	0.6	0.9	1.8	1.1
horizontal M	2.4	3.8	6.1	4.1
s	0.8	1.1	1.5	0.0
$t(27)$	2.7	3.5	4.7	6.4
1% Signifikanz	*	*	*	*
Effektstärke d	.34	.41	.54	.52

Genauso wie im früheren Gruppenversuch zeigt sich auch hier, dass Zahlen in einer Tabelle schneller miteinander verglichen werden können, wenn sie vertikal statt horizontal angeordnet sind. Der Vorteil wächst mit steigender Anzahl der Daten.

Der Einfluss der Zifferanzahl [2004]

Eine einfaktorielle Varianzanalyse, bei denen die Anzahl der Ziffern die Stufen der UV und die Zeit die AV darstellten, ergab einen hochsignifikanten F-Wert, der mit Geenhouse-Geiser korrigierten Freiheitsgraden [$F(1.7,54.) = 19.5$] das erforderliche Signifikanzniveau von 1% klar unterschreitet. Alle möglichen statistischen Vergleiche zwischen den Zeiten für die Zahlen mit verschiedener Zifferanzahl unterscheiden sich signifikant auf dem 5 % Niveau (nach Bonferroni Mehrfachvergleichsanpassung)

Tabelle 9 [2004]: Der Einfluss der Zifferanzahl

Mittelwerte **M** und Standardabweichung **s** für die Zeiten zum Erkennen des maximalen Wertes bei 2, 3 und 4 ziffrigen Zahlen

Anzahl der Ziffern	Zeit in Sekunden	
	M	s
2	3.5	0.9
3	3.8	1.0
4	4.2	1.2

Man erkennt aus der Tabelle 9, dass jede weitere Ziffer die Zeit um ca. 0.33 Effektstärken anwachsen lässt. Varianzanalyse und Mittelwertsvergleiche führen zu den gleichen Ergebnissen wie im vorherigen Gruppenversuch. Dort benötigte jede weitere Ziffer jedoch ca. 0.5 Effektstärken mehr Zeit.

Schlussbetrachtung

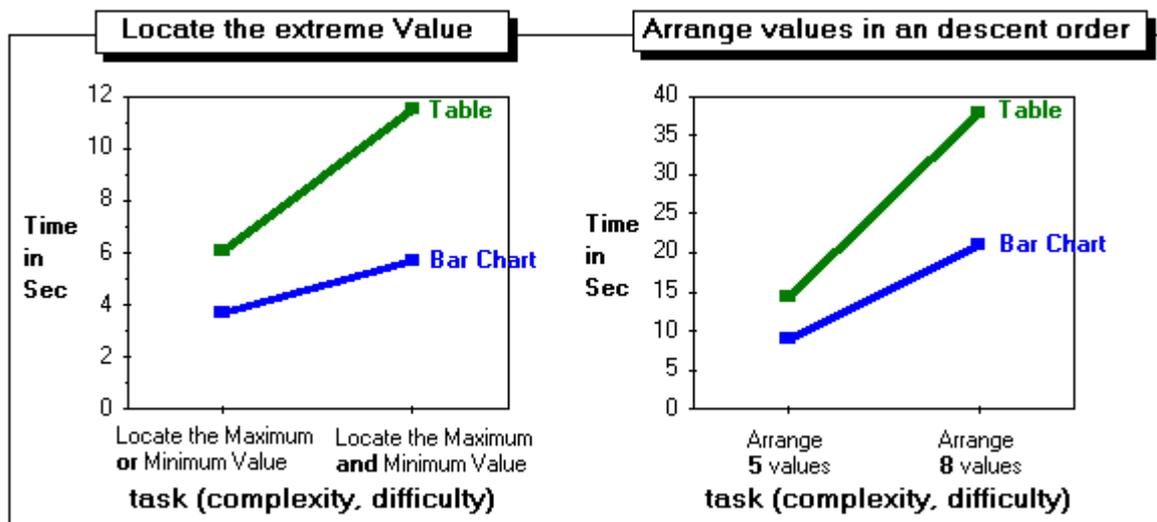
Das Computereperiment wurde eigentlich für Lehrzwecke genutzt, um Studenten die Vorteile eines Wiederholungsdesigns an einem konkreten Beispiel zu demonstrieren. Nebenbei ergab sich so die Möglichkeit einer erneuten Hypothesenüberprüfung unter natürlichen Studienbedingungen. Insgesamt bestätigen die Befunde die aufgestellten Hypothesen recht deutlich und zeigen eine sehr hohe Ähnlichkeit mit den früheren Ergebnissen. Insofern liegt hier eine ausgesprochen erfolgreiche, wenngleich keineswegs überraschende, Replikation vor.

Kurzfassung: Identifying of Relations between values (Jacobs 1990)

„Bigger" and „smaller" were the relations between values we were concentrated on in this investigation. 40 experimental subjects were asked to identify the maximum or minimum value in a bar graph or a table (each 12 values). The complexity was also varied. The easier question was to identify either the minimum or the maximum value. The more difficult question was to identify both, the minimum and the maximum value.

The second kind of question was to put 5 values (easy task) or 8 values (hard task) in a descent order. Each value was generated by chance for every experimental subject.

Results: Proved superiority of the bar graph to the table



Important Results

- The bar graph achieved always a significantly higher percentage of right answers than the table**
 Locate the extreme value: $[F(1,39)=22.11; p<.001]$
 Arrange values in an descent order: $[F(1,38) = 11.8; p<.001]$
- The table needed significantly more time than the bar chart.**
 Locate the extreme value: $(F(1,39)=112, p< .001)$.
 Arrange values in a descent order: $[F(1,38) = 83.8; p<.001]$
 The times divergences between the kinds of presentation always correspond an effect size of more than 1.
- The interactions between the kinds of presentation and the complexity according to time could be proved statistically on the promille-level.**
 Locate the extreme value: $[F(39,1)=29,42; p<.001]$
 Arrange values in an descent order: $[F(1,38) = 40.7; p<.001]$

Source: 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 Philosophischen Fakultät der Universität des Saarlandes. Nr. 3