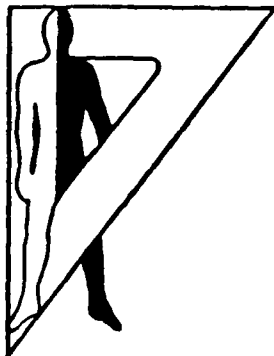


DTIC FILE COPY

4



AD

Technical Memorandum 8-88

AD-A202 370

COMPARATIVE GRAPHICS: HISTORY AND APPLICATIONS
OF PERCEPTUAL INTEGRALITY THEORY AND THE
PROXIMITY COMPATIBILITY HYPOTHESIS

C. Melody Carswell
Christopher D. Wickens
Aviation Research Laboratory
University of Illinois



November 1988
AMCMS Code 612716.H700011

Approved for public release;
distribution is unlimited.

U.S. ARMY HUMAN ENGINEERING LABORATORY
Aberdeen Proving Ground, Maryland

89 1 05 068

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGEForm Approved
OMB No 0704-0188
Exp. Date: Jun 30, 1986

1a. REPORT SECURITY CLASSIFICATION Unclassified		1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.	
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE		5. MONITORING ORGANIZATION REPORT NUMBER(S) Technical Memorandum 8-88	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)		7a. NAME OF MONITORING ORGANIZATION Human Engineering Laboratory	
6a. NAME OF PERFORMING ORGANIZATION Aviation Research Laboratory University of Illinois	6b. OFFICE SYMBOL (if applicable)	7b. ADDRESS (City, State, and ZIP Code) Aberdeen Proving Ground, MD 21005-5001	
6c. ADDRESS (City, State, and ZIP Code) 1 Airport Road Savoy, Illinois 61874		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (if applicable)	10. SOURCE OF FUNDING NUMBERS	
8c. ADDRESS (City, State, and ZIP Code)		PROGRAM ELEMENT NO. 6.27.16.A	PROJECT NO. 1L162716AH7D
		TASK NO.	WORK UNIT ACCESSION NO.
11. TITLE (Include Security Classification) Comparative Graphics: History and Applications of Perceptual Integrality Theory and the Proximity Compatibility Hypothesis			
12. PERSONAL AUTHOR(S) Carswell, C. Melody Wickens, Christopher D.			
13a. TYPE OF REPORT Final	13b. TIME COVERED FROM TO	14. DATE OF REPORT (Year, Month, Day) 1988, November	15. PAGE COUNT 85
16. SUPPLEMENTARY NOTATION			
17. COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB-GROUP	
01	03		
05	08		
		displays, S-C-R compatibility, graphs, compatibility of proximity, graphics, information integration.	
19. ABSTRACT (Continue on reverse if necessary and identify by block number)			
Interest has been recently renewed in the development and use of graphic displays for situations requiring the timely assimilation of large amounts of quantitative information. The present report traces the development of many of the graphic formats in common use today and reviews the experimental literature that compares alternative techniques. The proximity compatibility hypothesis is used to integrate the experimental work and is recommended as a framework to guide future experimentation and design decisions. Research issues regarding the appropriate functional classification of graphical formats--the designation of "graphical proximity"--are also discussed. <i>Keywords:</i>			
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS		21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Doris S. Eanes		22b. TELEPHONE (Include Area Code) 301-278-4478	22c. OFFICE SYMBOL SLCHE-SS-IR

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted.
All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

COMPARATIVE GRAPHICS: HISTORY AND APPLICATIONS OF PERCEPTUAL
INTEGRALITY THEORY AND THE PROXIMITY COMPATIBILITY HYPOTHESIS

C. Melody Carswell
Christopher D. Wickens
Aviation Research Laboratory
University of Illinois

November 1988

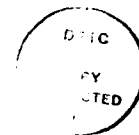
Approved

John D. Weisz
JOHN D. WEISZ
Director

Human Engineering Laboratory

Approved for public release;
distribution is unlimited.

U.S. ARMY HUMAN ENGINEERING LABORATORY
Aberdeen Proving Ground, Maryland 21005-5001



1. Title	
2. Author	
3. Subject	
4. Date	
5. Distribution	
6. Remarks	
7. Indexing	
8. Classification	
9. Approval	
10. Comments	
11. Signature	
12. Initials	
13. Date	
14. Remarks	
15. Indexing	
16. Classification	
17. Approval	
18. Comments	
19. Signature	
20. Initials	
21. Date	
22. Remarks	
23. Indexing	
24. Classification	
25. Approval	
26. Comments	
27. Signature	
28. Initials	
29. Date	
30. Remarks	
31. Indexing	
32. Classification	
33. Approval	
34. Comments	
35. Signature	
36. Initials	
37. Date	
38. Remarks	
39. Indexing	
40. Classification	
41. Approval	
42. Comments	
43. Signature	
44. Initials	
45. Date	
46. Remarks	
47. Indexing	
48. Classification	
49. Approval	
50. Comments	
51. Signature	
52. Initials	
53. Date	
54. Remarks	
55. Indexing	
56. Classification	
57. Approval	
58. Comments	
59. Signature	
60. Initials	
61. Date	
62. Remarks	
63. Indexing	
64. Classification	
65. Approval	
66. Comments	
67. Signature	
68. Initials	
69. Date	
70. Remarks	
71. Indexing	
72. Classification	
73. Approval	
74. Comments	
75. Signature	
76. Initials	
77. Date	
78. Remarks	
79. Indexing	
80. Classification	
81. Approval	
82. Comments	
83. Signature	
84. Initials	
85. Date	
86. Remarks	
87. Indexing	
88. Classification	
89. Approval	
90. Comments	
91. Signature	
92. Initials	
93. Date	
94. Remarks	
95. Indexing	
96. Classification	
97. Approval	
98. Comments	
99. Signature	
100. Initials	
101. Date	
102. Remarks	
103. Indexing	
104. Classification	
105. Approval	
106. Comments	
107. Signature	
108. Initials	
109. Date	
110. Remarks	
111. Indexing	
112. Classification	
113. Approval	
114. Comments	
115. Signature	
116. Initials	
117. Date	
118. Remarks	
119. Indexing	
120. Classification	
121. Approval	
122. Comments	
123. Signature	
124. Initials	
125. Date	
126. Remarks	
127. Indexing	
128. Classification	
129. Approval	
130. Comments	
131. Signature	
132. Initials	
133. Date	
134. Remarks	
135. Indexing	
136. Classification	
137. Approval	
138. Comments	
139. Signature	
140. Initials	
141. Date	
142. Remarks	
143. Indexing	
144. Classification	
145. Approval	
146. Comments	
147. Signature	
148. Initials	
149. Date	
150. Remarks	
151. Indexing	
152. Classification	
153. Approval	
154. Comments	
155. Signature	
156. Initials	
157. Date	
158. Remarks	
159. Indexing	
160. Classification	
161. Approval	
162. Comments	
163. Signature	
164. Initials	
165. Date	
166. Remarks	
167. Indexing	
168. Classification	
169. Approval	
170. Comments	
171. Signature	
172. Initials	
173. Date	
174. Remarks	
175. Indexing	
176. Classification	
177. Approval	
178. Comments	
179. Signature	
180. Initials	
181. Date	
182. Remarks	
183. Indexing	
184. Classification	
185. Approval	
186. Comments	
187. Signature	
188. Initials	
189. Date	
190. Remarks	
191. Indexing	
192. Classification	
193. Approval	
194. Comments	
195. Signature	
196. Initials	
197. Date	
198. Remarks	
199. Indexing	
200. Classification	
201. Approval	
202. Comments	
203. Signature	
204. Initials	
205. Date	
206. Remarks	
207. Indexing	
208. Classification	
209. Approval	
210. Comments	
211. Signature	
212. Initials	
213. Date	
214. Remarks	
215. Indexing	
216. Classification	
217. Approval	
218. Comments	
219. Signature	
220. Initials	
221. Date	
222. Remarks	
223. Indexing	
224. Classification	
225. Approval	
226. Comments	
227. Signature	
228. Initials	
229. Date	
230. Remarks	
231. Indexing	
232. Classification	
233. Approval	
234. Comments	
235. Signature	
236. Initials	
237. Date	
238. Remarks	
239. Indexing	
240. Classification	
241. Approval	
242. Comments	
243. Signature	
244. Initials	
245. Date	
246. Remarks	
247. Indexing	
248. Classification	
249. Approval	
250. Comments	
251. Signature	
252. Initials	
253. Date	
254. Remarks	
255. Indexing	
256. Classification	
257. Approval	
258. Comments	
259. Signature	
260. Initials	
261. Date	
262. Remarks	
263. Indexing	
264. Classification	
265. Approval	
266. Comments	
267. Signature	
268. Initials	
269. Date	
270. Remarks	
271. Indexing	
272. Classification	
273. Approval	
274. Comments	
275. Signature	
276. Initials	
277. Date	
278. Remarks	
279. Indexing	
280. Classification	
281. Approval	
282. Comments	
283. Signature	
284. Initials	
285. Date	
286. Remarks	
287. Indexing	
288. Classification	
289. Approval	
290. Comments	
291. Signature	
292. Initials	
293. Date	
294. Remarks	
295. Indexing	
296. Classification	
297. Approval	
298. Comments	
299. Signature	
300. Initials	
301. Date	
302. Remarks	
303. Indexing	
304. Classification	
305. Approval	
306. Comments	
307. Signature	
308. Initials	
309. Date	
310. Remarks	
311. Indexing	
312. Classification	
313. Approval	
314. Comments	
315. Signature	
316. Initials	
317. Date	
318. Remarks	
319. Indexing	
320. Classification	
321. Approval	
322. Comments	
323. Signature	
324. Initials	
325. Date	
326. Remarks	
327. Indexing	
328. Classification	
329. Approval	
330. Comments	
331. Signature	
332. Initials	
333. Date	
334. Remarks	
335. Indexing	
336. Classification	
337. Approval	
338. Comments	
339. Signature	
340. Initials	
341. Date	
342. Remarks	
343. Indexing	
344. Classification	
345. Approval	
346. Comments	
347. Signature	
348. Initials	
349. Date	
350. Remarks	
351. Indexing	
352. Classification	
353. Approval	
354. Comments	
355. Signature	
356. Initials	
357. Date	
358. Remarks	
359. Indexing	
360. Classification	
361. Approval	
362. Comments	
363. Signature	
364. Initials	
365. Date	
366. Remarks	
367. Indexing	
368. Classification	
369. Approval	
370. Comments	
371. Signature	
372. Initials	
373. Date	
374. Remarks	
375. Indexing	
376. Classification	
377. Approval	
378. Comments	
379. Signature	
380. Initials	
381. Date	
382. Remarks	
383. Indexing	
384. Classification	
385. Approval	
386. Comments	
387. Signature	
388. Initials	
389. Date	
390. Remarks	
391. Indexing	
392. Classification	
393. Approval	
394. Comments	
395. Signature	
396. Initials	
397. Date	
398. Remarks	
399. Indexing	
400. Classification	
401. Approval	
402. Comments	
403. Signature	
404. Initials	
405. Date	
406. Remarks	
407. Indexing	
408. Classification	
409. Approval	
410. Comments	
411. Signature	
412. Initials	
413. Date	
414. Remarks	
415. Indexing	
416. Classification	
417. Approval	
418. Comments	
419. Signature	
420. Initials	
421. Date	
422. Remarks	
423. Indexing	
424. Classification	
425. Approval	
426. Comments	
427. Signature	
428. Initials	
429. Date	
430. Remarks	
431. Indexing	
432. Classification	
433. Approval	
434. Comments	
435. Signature	
436. Initials	
437. Date	
438. Remarks	
439. Indexing	
440. Classification	
441. Approval	
442. Comments	
443. Signature	
444. Initials	
445. Date	
446. Remarks	
447. Indexing	
448. Classification	
449. Approval	
450. Comments	
451. Signature	
452. Initials	
453. Date	
454. Remarks	
455. Indexing	
456. Classification	
457. Approval	
458. Comments	
459. Signature	
460. Initials	
461. Date	
462. Remarks	
463. Indexing	
464. Classification	
465. Approval	
466. Comments	
467. Signature	
468. Initials	
469. Date	
470. Remarks	
471. Indexing	
472. Classification	
473. Approval	
474. Comments	
475. Signature	
476. Initials	
477. Date	
478. Remarks	
479. Indexing	
480. Classification	
481. Approval	
482. Comments	
483. Signature	
484. Initials	
485. Date	
486. Remarks	
487. Indexing	
488. Classification	
489. Approval	
490. Comments	
491. Signature	
492. Initials	
493. Date	
494. Remarks	
495. Indexing	
496. Classification	
497. Approval	
498. Comments	
499. Signature	
500. Initials	
501. Date	
502. Remarks	
503. Indexing	
504. Classification	
505. Approval	
506. Comments	
507. Signature	
508. Initials	
509. Date	
510. Remarks	
511. Indexing	
512. Classification	
513. Approval	
514. Comments	
515. Signature	
516. Initials	
517. Date	
518. Remarks	
519. Indexing	
520. Classification	
521. Approval	
522. Comments	
523. Signature	
524. Initials	
525. Date	
526. Remarks	
527. Indexing	
528. Classification	
529. Approval	
530. Comments	
531. Signature	
532. Initials	
533. Date	
534. Remarks	
535. Indexing	
536. Classification	
537. Approval	
538. Comments	
539. Signature	
540. Initials	
541. Date	
542. Remarks	
543. Indexing	
544. Classification	
545. Approval	
546. Comments	
547. Signature	
548. Initials	
549. Date	
550. Remarks	
551. Indexing	
552. Classification	
553. Approval	
554. Comments	
555. Signature	
556. Initials	
557. Date	
558. Remarks	
559. Indexing	
560. Classification	
561. Approval	
562. Comments	
563. Signature	
564. Initials	
565. Date	
566. Remarks	
567. Indexing	
568. Classification	
569. Approval	
570. Comments	
571. Signature	
572. Initials	
573. Date	
574. Remarks	
575. Indexing	
576. Classification	
577. Approval	
578. Comments	
579. Signature	
580. Initials	
581. Date	
582. Remarks	
583. Indexing	
584. Classification	
585. Approval	
586. Comments	
587. Signature	
588. Initials	
589. Date	
590. Remarks	
591. Indexing	
592. Classification	
593. Approval	
594. Comments	
595. Signature	
596. Initials	
597. Date	
598. Remarks	
599. Indexing	
600. Classification	
601. Approval	
602. Comments	
603. Signature	
604. Initials	
605. Date	
606. Remarks	
607. Indexing	
608. Classification	
609. Approval	
610. Comments	
611. Signature	
612. Initials	
613. Date	
614. Remarks	
615. Indexing	
616. Classification	
617. Approval	
618. Comments	
619. Signature	
620. Initials	
621. Date	
622. Remarks	
623. Indexing	
624. Classification	
625. Approval	
626. Comments	
627. Signature	
628. Initials	
629. Date	
630. Remarks	
631. Indexing	
632. Classification	
633. Approval	
634. Comments	
635. Signature	
636. Initials	
637. Date	
638. Remarks	
639. Indexing	
640. Classification	
641. Approval	
642. Comments	
643. Signature	
644. Initials	
645. Date	
646. Remarks	
647. Indexing	
648. Classification	
649. Approval	
650. Comments	
651. Signature	
652. Initials	
653. Date	
654. Remarks	

CONTENTS

CHAPTER ONE	GRAPHICAL TECHNIQUES: HISTORY AND PROSPECTS . . .	3
	The Rise and Fall of Statistical Graphics	5
	Graphics and Gadgets	9
	A Graphical Renaissance	11
CHAPTER TWO	COMPARATIVE GRAPHICS	17
	Graphs for Locating Exact Information	19
	Graphs for Simple Comparisons	25
	Graphs for Information Synthesis	28
	Graphs for Complex Comparisons	30
	Mediating Variables in Comparative Graphics	32
CHAPTER THREE	THE PROXIMITY COMPATIBILITY HYPOTHESIS	37
	The Concept of Compatibility	37
	Experiments on Proximity Compatibility	41
	Evidence for the Proximity Compatibility Hypothesis	42
CHAPTER FOUR	DEFINITIONS OF GRAPHICAL PROXIMITY	45
	Task Proximity	47
	Graphical Proximity: An Overview	49
	Dimensional Dependence	50
	Dimensional Homogeneity	61
	Object Proximity	65
CHAPTER FIVE	SUMMARY OF APPLIED AND EXPERIMENTAL ISSUES IN GRAPHIC DESIGN	73
REFERENCES		75

FIGURES

1.1	Chronology of Major Graphical Formats Through the Nineteenth Century	7
1.2	Multivariate Statistical Graphs of the Twentieth Century	13
3.1	Some Examples of Mapping Proximity	39
3.2	Advantages of Low- and High-Proximity Graphs Used in Four Types of Tasks	43
4.1	Hypothesized Effect of Display Proximity on Performance for an Integration Task and an Independent Processing Task	60

TABLES

2.1	Summary of Studies Comparing Graphical Formats	20
4.1	Compatible and Incompatible Matches of Task and Display Proximity	46
4.2	Dimensional Dependencies Used as Graphical Proximity Estimates in the Proximity Compatibility Hypothesis	58
4.3	Dimensional Homogeneity Used as a Measure of Display Proximity in the Proximity Compatibility Hypothesis	64
4.4	Object Proximity Used as a Measure of Display Proximity in the Proximity Compatibility Hypothesis . .	69

CHAPTER 1

GRAPHICAL TECHNIQUES: HISTORY AND PROSPECTS

For some years now, those persons curious enough to be wooed into empirical exploits and those enthusiastic enough to publicize their carefully collected knowledge have acclaimed the benefits of graphical presentation of quantitative data. Graphics, they argue, allow the assimilation of information in ways tables of numbers fail to match. One of the earliest advocates of graphical displays, William Playfair, was quite adamant on this point. In his ground-breaking book of 1786, the Commercial and Political Atlas, Playfair disparages the use of tables:

Information, that is imperfectly acquired, is generally as imperfectly retained; and a man who has carefully investigated a printed table, finds, when done, that he has only a very faint and partial idea of what he has read; and that like a figure imprinted on sand, is soon totally erased and defaced (p. 3).

With regard to his own experimental "charts," on the other hand, Playfair writes:

On inspecting any one of these charts attentively, a sufficiently distinct impression will be made, to remain unimpaired for a considerable time, and the idea which does remain will be simple and complete, at once indicating the duration and amount (p. 4).

Or even more to the point is the commentary of two nineteenth century economists (as quoted by Wainer & Thissen, 1981): "Getting information from a table is like extracting sunlight from a cucumber" (p. 236).

Today's proponents of various graphical techniques sound remarkably like their predecessors writing in the two previous centuries. When enumerating the advantages of graphics, they commonly refer to such attributes as mnemonic value, impact, ability to highlight patterns of relationships among variables, as well as the ability of graphs to attract attention (e.g., Chernoff, 1978; Moriarity, 1979; Cleveland, 1985). Chambers, Cleveland, Kleiner, and Tukey (1983) exemplify these views in the introduction to their recent textbook on graphical data analysis:

An enormous amount of quantitative information can be conveyed by graphs: our eye-brain system can summarize vast information quickly and extract salient features, but it is also capable of focusing on detail. Even for small sets of data, there are many patterns and relationships that are considerably easier to discern in graphical displays than by any other data analytic method (p. 1).

Although the accolades for graphical techniques have echoed, almost unaltered across the generations, the variety of displays being called "graphics" has enlarged considerably. For Playfair, the predominant graphic display was the simple line graph. This conception of "graphics" was still relatively dominant at the time of the meeting of the Joint Committee on

Standards for Graphic Presentation (American Statistical Association, 1916). Sixteen of the seventeen recommendations generated by this group were with reference to coordinate line graphs. Today, however, graphics refers not only to old standby techniques such as line graphs, bar charts, pie charts, and scatter plots, but to such new multidimensional forms as polygons, faces, trees, and chromatic contour plots, with or without animation. In addition, the term "graphics" has come to denote a whole new technology of computer-generated art, animation, and nonstandard alphanumeric displays, as well as visual representations of quantitative information (Harris, 1984).

Because of the present diversity in the use of the terms "graphics" and "graphical," the exact domain of present interest must be specified in this report. Graphics here will denote any visual representation, whether generated by hand or computer, that uses some perceptual dimension that varies in magnitude as an analog for a physically measured or derived value. These representations need not be literal, and most often they are not. For instance, the height of a bar in a vertical bar graph is not limited to the representation of skyward extent. However, by taking advantage of this built-in physical analogy, such graphs can be used very effectively for height representations. In addition, by the above definition, computer art is not considered graphics in the present sense. This exclusion is mainly functional, in that the purpose of art is rarely to communicate real-world measures. On the other hand, computer-generated alphanumeric displays, while often used for presenting such measures, are also excluded in this definition of graphics since they lack the requisite physical-analog characteristic (i.e., because of their discrete, arbitrary nature). There will be, of course, some displays with characteristics of both graphic and nongraphic forms. For instance, Tukey's (1977) stem-and-leaf display uses both a collection of specific numerals (digital representations) and the shape of this collection of numerals (analog information) in the same display. However, the present report will be concerned only with the analog aspects of such hybrids.

The emphasis of the present report, moreover, will be on "derivative" forms of graphical representation as opposed to "basic graphics" (Beniger & Robyn, 1978). Derivative graphics, the more historically recent forms, include those techniques such as pie charts, bar graphs, and bivariate point displays that are not confined to literal descriptions of space and time. Basic graphics, on the other hand, are those forms maintaining a high level of topological isomorphism with the domain that they represent--such as maps, coordinate systems used for simple geometric computations, and circuit diagrams. The next few sections of this report will be devoted to a brief history of derivative graphics and to a discussion of the present trends and issues involved with such displays. Evidence will be presented to support the thesis that the present era is a "graphical renaissance"--a revival of interest in graphics that has united a number of techniques previously restricted to either statistics or industry. This discussion will be followed by a review of the comparative graphics literature and an introduction to the hypothesis of "proximity compatibility," a theoretical framework within which further experimentation may proceed.

The Rise and Fall of Statistical Graphics

The notion of pictorial representation of quantitative information may seem at first glance to be a rather primitive one, conjuring up images of cave drawings and papyrus etchings. However, the history of derivative representational graphics is notable for its brevity, largely dating from the turn of the nineteenth century. What traces there are of a graphical history are predominantly traces of a history of statistical graphics. For a more thorough history of statistical graphics, the reader is referred to articles by Beniger and Robyn (1978), Feinberg (1979), Wainer and Thissen (1981), and Tilling (1975).

Of course, the foundations for the development of representational or derivative graphics were almost certainly laid by older, basic forms of graphics. Beniger and Robyn (1978) trace the development of graphics back to the origins of cartography, with the oldest known map dating from 3800 B.C. They further follow the evolution of early graphic forms to 1500 B.C., with the earliest graphical representations of practical geometric problems, and into the middle ages when curves were used to represent planetary orbits on a time grid. However, it was not until the seventeenth century with the development by Descartes of a coordinate system that the immediate substrate for more representational graphics is found. Here was a system for representing mathematical functions governing the behavior of objects in time and space. However, as Wainer and Thissen (1981) have observed, as well as forming the intellectual basis for derivative graphical forms, the Cartesian coordinate system may have also been responsible for an intellectual impasse of sorts. The system so dominated scientists' ideas of the function and form of graphs that nearly a century and a half passed before graphical dimensions came to be used to represent nonspatial, empirical data.

A number of graphical historians consider William Playfair (1759-1823), an English political scientist and economist, to be the Father of Statistical Graphics (Wainer & Thissen, 1981; Schmid & Schmid, 1979; Funkhouser, 1937). Although, as Beniger and Robyn (1978) note, several attempts to represent quantitative information graphically had been made somewhat earlier. For instance, a bivariate point display was used by Edmund Halley in 1686 to illustrate the relation of barometric pressure to altitude, and the line graph was used abstractly in 1724 by Nicolaus Cruquius to represent barometric observations. However, Playfair was the first person to systematically promote and experiment with many such graphic forms.

Several authors have posed the question of why experimental applications of graphics prior to the time of Playfair met with relatively little excitement. Perhaps, as Wainer and Thissen (1981) suggest, the strong Cartesian tradition of the time led the few who attempted such graphical representation of empirical data to fall into the belief that they were actually doing nothing more than Cartesian geometric analysis. Further, graphic forms derived from the coordinate approach of the French mathematicians during the early seventeenth century were not universally acclaimed. Beniger and Robyn (1978) report that graphic forms were overshadowed by the tabular approach to organizing data (die tabellen-statistik) adopted and ardently defended by a group of German social

scientists. By the late seventeenth century, the tabular approach had received some acceptance in Britain where it became known as the "political arithmetic."

When Playfair introduced his innovative graphics a century later, calling them "the lineal arithmetic," the public's acceptance was still not without its obstacles. Indeed, Playfair was still fettered by his public's expectation that graphics were literal depictions, though on a different scale, of spatially exact relations in the real world. For instance, in the third edition of his Commercial and Political Atlas (1801), Playfair introduced his lineal arithmetic by citing a concrete pictorial relation for the use of linear extent to represent income:

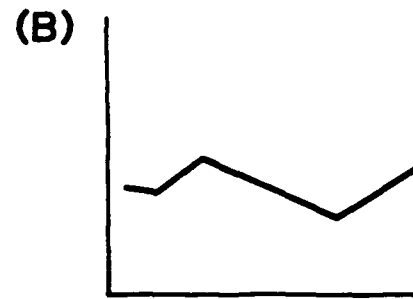
Suppose the money received by a man in trade were all in guineas, and that every evening he made a single pile of all the guineas received during the day, each pile would represent a day, and its height would be proportioned to the receipts of that day, so that by this plain operation, time, proportion, and amount, would all be physically combined. Lineal arithmetic then, it may be averred, is nothing more than these piles of guineas represented on paper and on a small scale, in which an inch (suppose) represents the thickness of five millions of guineas, as in geography it does the breadth of a river, or any other extent of country (p. 6).

Despite the trials of explaining his methods, Playfair managed to introduce in his atlases and other writings many of the most commonly used techniques found in our modern day repertoire--bar graphs, line graphs used in a nonliteral way, pie charts, and even a multivariate "object" display. Figure 1.1 presents a chronology of many of the major graphical forms found in the eighteenth and nineteenth centuries, including those developed before and after Playfair's contribution. This chronology is based primarily on the research of Beniger and Robyn (1978). As shown in Figure 1.1, such famous names as Fourier, Quetelet, and Florence Nightingale each contributed new graphic forms. By 1857, graphs had become so commonplace that the International Statistical Meeting in Vienna had an entire exhibition devoted to displays of various graphical techniques. One can easily imagine the comparisons and controversy being evoked by a collection of graphs, not unlike those shown in Figure 1.1.

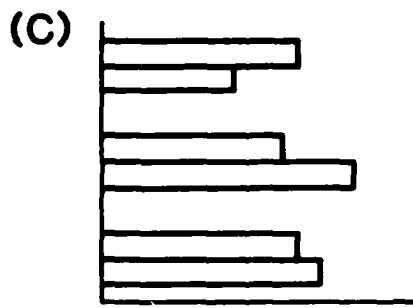
The years following the Vienna conference, from 1860 to 1890, have been called the Golden Age of Graphical Techniques in a chronology presented by Cox (1978). During this period, the earliest attempts to develop graphical standards were made (Feinberg, 1979). This concern for standardization eventually resulted in the formation of the joint committee on Standards for Graphic Presentation. In 1914, invitations were extended by the American Society of Mechanical Engineers to various other professional groups to join in the standardization process. The diversity of the members of this committee shows the wide-ranging concern for graphic standardization at this time (American Statistical Association, 1916). Included were representatives of the American Genetic Association, The American Society of Naturalists, the U.S. Census Bureau, and 14 other societies. The importance of this project to



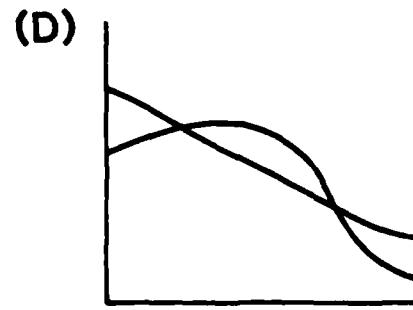
1686—Sir Edmund Halley
Bivariate point display relating altitude and barometric readings



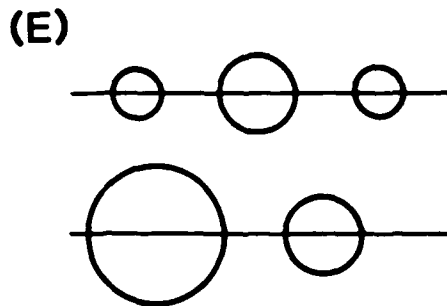
1724—Cruquius
Abstract line graph of barometric pressure



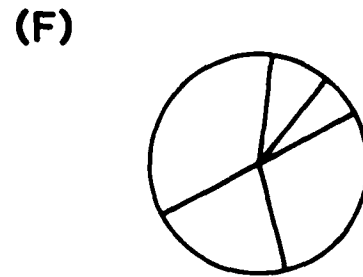
1786—Playfair
Bar chart used to display economic data



1801—Playfair
Comparative line graphs



1801—Playfair
Circle charts used area to convey population estimates

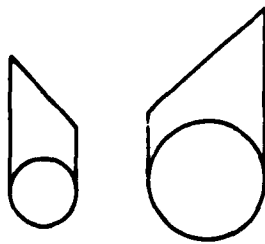


1801—Playfair
Pie chart

Figure 1.1. Chronology of major graphical formats through the nineteenth century.
(continued on next page)

(continued)

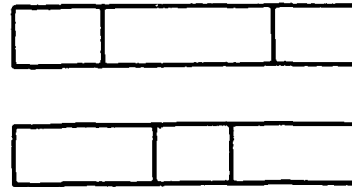
(G)



1801—Playfair

Tri-variate display used area of a circle and extent of two lines to represent three variables

(H)



1811—Von Humboldt

Subdivided bar graph used to represent proportion of part to whole (variant of pie chart)

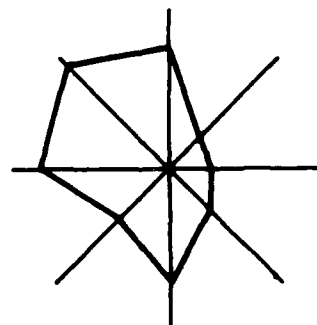
(I)



1821—Fourier

Cumulative frequency curve

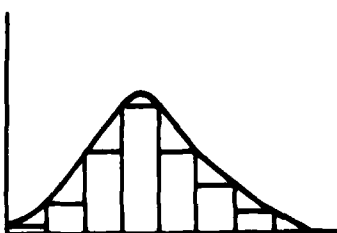
(J)



1843—Lalanne

Polar coordinate plots (polygons) used to show frequency of wind direction

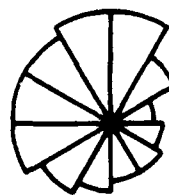
(K)



1846—Quetelet

"Curve of possibility"

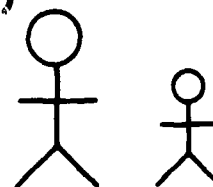
(L)



1857—Nightingale

Coxcomb chart used to show monthly fatalities during Crimean War

(M)



1884—Mulhall

Pictogram used object size to represent values

Figure 1.1. Chronology of major graphical formats through the nineteenth century.

the American Psychological Association is evident in their choice of a representative--no less a figure than E. L. Thorndike.

Feinberg (1979) traced the use of graphs in prominent statistical journals from 1920 to 1975. He reports a trend towards fewer and fewer pages devoted to charts and graphs in the Journal of the American Statistical Association and Biometrika. Beniger and Robyn (1979) likewise note a waning of interest in graphics among academic statisticians dating from World War II. These authors attribute this decline in interest to the concurrent increase in new techniques of mathematical statistics rather than to a decrease in interest in graphics per se.

Graphics and Gadgets

Parallel to the development of the static graphical techniques used to analyze and communicate statistical data runs an additional history of graphical representation. This second lineage is the story of automatic graphical recording and dynamic analog displays. Such displays resulted from attempts to automatically keep track of various natural phenomena, as well as the eventual desire to keep track of the workings of various machines. As such, the development of these displays is closely affiliated with the history of gadgetry and invention.

A review of early graphic recording by Hoff and Geddes (1959) traces early attempts to automate various counting tasks. For instance, they relate attempts by the Greco-Roman inventors to keep precise track of time with the clepsydra, and to estimate distance traveled (i.e., number of cart wheel revolutions) with the odometer. The clepsydra, or water clock, equated the passage of time with a constant-rate flow of water that slowly filled a chamber. Initial attempts to display the amount of time elapsed made use of a float on the surface of the clepsydra's rising water (third century B.C.). Projecting from the float was a pointer that moved, as the water level rose, up a carefully spaced scale, with the distance between tick marks representing particular time intervals. In a way, then, this early clock used a type of slowly rising bar graph as its time display. However, by the first century A.D., the dial face of the clock had been invented. This development arose when some ingenious tinkerer attached a cord to the rising float of the clepsydra, and then wrapped its counterbalanced end around an axle. When the axle was appropriately marked, its circular motion could be viewed as one of the hands of a modern-day analog timepiece.

Although crude analog displays were used as the output for various gadgets dating as far back as classical antiquity, the choice of a graphic rather than numeric format was often out of mechanical convenience rather than choice. For instance, by the seventeenth and eighteenth centuries, there were a plethora of mechanical recorders producing moving line graphs. However, Beniger and Robyn (1978) note that these graphs were often translated into tabular logs, the graphs themselves being considered useless for analysis.

A major breakthrough in the history of automatic graphic recordings came when Carl Ludwig developed a way of making a permanent graphic record of variations in arterial blood pressure (Hoff & Geddes, 1959). His device, the

kymograph, was developed in 1847 and involved placing a float with a writing stylus on a mercury manometer. The stylus was placed in contact with the sooted surface of a drum that was turned by a falling weight. When the manometer was attached to an artery, and the drum was set in motion, the fluctuations of arterial pressure were played out on the drum's surface. The kymograph was the predecessor of many physiological recording devices. Sophisticated versions of such monitoring devices are, of course, commonly found today in most hospitals and medical laboratories. The graphic medium, however, is more likely to be a computer-driven VDU (video display unit) rather than a sooted, rotating drum. In addition, the actual format of the display has a great deal of flexibility and is currently the focus of increasing attention. A recent article by Cole (1986) is an example of some of the considerations of the field he calls "cognitive medical graphics."

About the time statistical graphics was beginning to make its impact in the late eighteenth and early nineteenth centuries, events forced yet another function upon the available automated graphic display techniques. This new application was a result of the increased mechanization of industry. While previous techniques for graphic recording were largely used by scientists and inventors to study aspects of nature too tedious to record by hand, or unavailable to unaided human observation, the new function accorded to graphic displays was the representation of internal states of all sorts of gadgets. Industrialists realized that one of the main bottlenecks to efficient operation and production using the new technologies was the control decisions made by the human supervisors. In other words, as the importance of mechanical operation efficiency was translated into guineas and francs, the need for displays capable of representing machine function was also realized. These displays, more often than not, were analog-visual in format. An example of the early expression of the need for such graphics was the 1838 comment of a locomotive engineer regarding the maintenance of optimal steam pressure in the boiler (quoted from Hoff & Geddes, 1959):

If in the early years steam did not constantly blow off through the safety valve, the locomotive drivers believed they did not have an adequate pressure; now they let it often sink so low that the regularity of travel is influenced, which happens all the sooner, since at low pressures the driver has no means at his disposal to convince himself of the true state of the steam-pressure in his boiler (p. 16).

James Watt, the Scottish engineer, had earlier developed a means "to convince himself of the true state of steam pressure in his boilers," using glass U-tubes filled with mercury. Such devices were surely the forerunners, at least in principle, of the numerous dials, meters, and other analog-visual displays used to monitor technological systems over the ensuing years. This technology has not remained limited to steam engines, however, and the operation of machines of all sorts soon came to be displayed in various forms, often graphic. Fowkes (1984) has recently recounted the history of automotive display instrumentation. However, the mode of transportation having the greatest impact on the development of dynamic graphic displays has almost surely been aviation.

Since the hallmark flight of Doolittle in 1929, using only instruments for guidance, aircraft cockpits have provided a testing ground for new display technology. Various instruments and display strategies have been used with the aim of providing pilots with the means to direct their aircraft without dependence on visual contact with the world below. Roscoe (1968) summarizes the progression of display innovation that has been based on the objective of total instrument flight in the last half century.

In addition to transportation graphics, the twentieth century has seen an exponential increase in the complexity of many process control situations, along with the display designs that support the supervision of such processes. The epitome of these process control environments, and one in which display technology has been studied both in the name of efficiency and safety, has been control of nuclear reactors. In this environment, mercury-filled U-tubes seem remote indeed as a means of providing a picture of the process. As Sapita (1982) notes, we have come a long way from the time when pressure gauges, sight glasses, and thermometers were the major fabric of the man-machine interface. Instead, the process is viewed by means of centralized computer-assisted complexes of displays, the ultimate format of which is often as easily displayed digitally as graphically. The options for the continuous portrayal of the internal machine processes are in many ways astounding, and the choice of the appropriate modes has become a question of some importance in the last decades.

A Graphical Renaissance

There have been reports of a decline in academic interest in statistical graphics since World War II (Feinberg, 1979; Beniger & Robyn, 1978), and the fate of dynamic analog graphics has also seemed dubious with automatic digital outputs finally becoming available in the same period. However, very recent research activity in both industrial and statistical display design seems to indicate that the dimming graphic picture may have once more begun to brighten. In fact, several authors have referred to the present era, dating from the early 1970s, as a "graphical renaissance" (Kruskal, 1977; Beniger & Robyn, 1978; Chernoff, 1978; Barnett, 1981).

Just as there is a growing consensus that graphical methods of displaying information are regaining popularity, there is ample agreement that this renewed interest is related to advances in computer technology (Chernoff, 1978; Wainer & Thissen, 1981). In general, there appears to be three ways in which advances in computer science have helped revive statistical and industrial graphics. The first of these involves the increased data manipulation and analysis capabilities of computers, which has seemed to outstrip our abilities to assimilate the end-products of many of these techniques. Secondly, increased automation in aviation and industry, much of which has been made possible through advances in computer technology, has produced an ever-increasing burden on the human supervisor's ability to observe and integrate numerous sources of information. Both of these advances, thus, require innovative ways of displaying massive amounts of information--ways that frequently involve imaginative uses of graphics. Finally, advances in computer technology have made design and presentation of these innovative graphic techniques remarkably quick and easy. Each of the

three factors contributing to the graphical renaissance will now be discussed in turn.

Statistical graphics. Beginning in the 1960s and 1970s, there has been a movement called by Feinberg (1979) "the new statistical graphics." This movement has been directly related by several authors to the development of computer technology that allowed large multivariate data sets to be manipulated (Wainer & Thissen, 1981; Chernoff, 1978). However, as Barnett in the preface to his 1981 book points out, the emphasis on multivariate statistics has been one of "going back to the drawing board"-- going back to the visual representation of the data. The questions being asked about the data reflected the need to make sense of large bodies of information; that is, statisticians had to begin to acknowledge their own limits as information processors. Barnett (p. v) lists the following as typical questions that seem to be best answered with graphical assists:

What do the data really show us in the midst of their apparent chaos?"

How can we logically summarize and represent these data?

How can we reduce dimensionality and scale to a level where the message of the data is, at least informally, clear...?

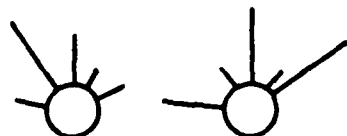
These questions, many proponents of exploratory data analysis hold, can be best addressed through the clever use of graphics. This belief has resulted in a number of graphical innovations for the representation of multivariate data within the confines of two-dimensional space, many of which are shown in Figure 1.2.

Many of the displays in Figure 1.2 may be called "object displays." This title is descriptive in that various attributes of a single perceptual object are used to convey the various dimensions of numeric information. These displays may be contrasted with older techniques such as the bar graph (or new variants such as "dot plots") that use the same attribute of each of several separate objects as a means to convey quantitative information. These display formats will be discussed in more detail in subsequent sections of this report.

Feinberg (1979) credits the new emphasis on statistical graphics to a group of researchers at Bell Telephone Labs and Princeton University. Particularly important to this movement has been the contributions of J.W. Tukey, whose experimentation with data presentation has made him a modern-day Playfair. By further analogy with events of the 19th century, a recent (1977) convention held in Sheffield, United Kingdom may prove to be the Vienna convention of our day. This conference, which was highly attended, has been cited by Cox (1978) as evidence of the present renaissance in graphical statistics.

Graphics in process control, management, and medicine. Computers have been responsible, not only for the production of answers to difficult-to-understand statistical multivariate problems, but also for larger quantities

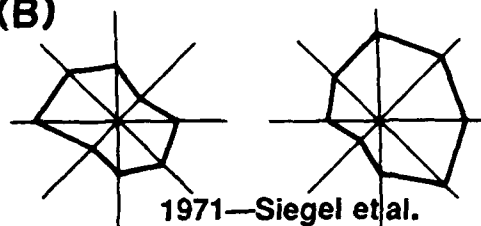
(A)



1957—Anderson

Glyphs used each "ray" to represent a different variable. Note the similarity to Playfair's 1801 multivariate display (Figure 1.1G)

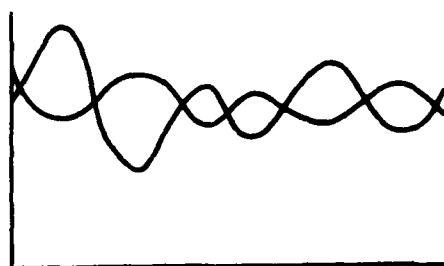
(B)



1971—Siegel et al.

Stars or polygon displays use the position of each vertex to represent variables. Similar to circular unidimensional profiles (Figure 1.1J)

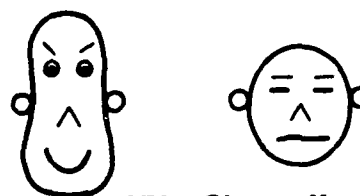
(C)



1972—Andrews

Andrews' plot uses a Fourier series to generate a plot of multivariate data

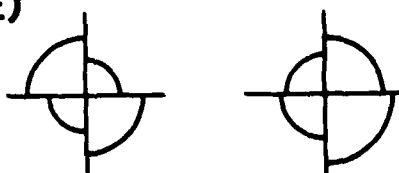
(D)



1973—Chernoff

Chernoff faces use the characteristics of facial features to represent the state of many variables. Variants have included football player displays and insect displays

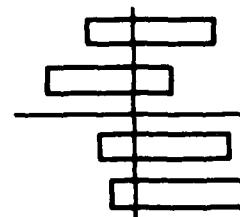
(E)



1979—Feinberg

Four-fold circular displays are used to display 3-way categorical data. Similar in form to Nightingale's coxcomb charts

(F)



1978—Wainer & Reiser

Cartesian rectangles are used for 3-way categorical data

Figure 1.2. Multivariate statistical graphs of the twentieth century.

of information being available for many applied decision-making tasks. For instance, in process control situations in industry, the function of the human in relation to the machinery has changed drastically in the last few decades. Largely because of the introduction of computers, automation of many tasks previously under the direct control of a plant worker has become commonplace. Presently, humans are finding themselves in the position of supervisory control of both the ultimate chemical or physical process and of a hierarchy of automation attending to a wide range of lower-order tasks. This position places high demands on the human to assimilate information from numerous sources, and many conventional displays have proved unsatisfactory for this purpose (Goodstein, 1981). The incident at Three Mile Island, for instance, catapulted the nuclear industry into a reevaluation of its display philosophies (Sapita, 1982). Resulting from this reevaluation has been the incorporation of innovative graphics in the preparation of an integrative safety parameter display system (SPDS). Woods, Wise, and Hanes (1981), for instance, have developed a dynamic polygon display (see Figure 1.2b for an example of a static variant) that represents nine critical safety-related values in a single form.

Another realm in which computer-related information-overload has become a problem worthy of note is in business management. Increasingly, upper-level management is relying on computer-based management information systems (MISs) to quickly provide them with relevant information and aids for organizational decision-making. Recently, the format of the information presented to these decision-makers has become of interest. DeSanctis (1984) presents a review of graphic applications in the field, and emphasizes comparisons among various graphic designs, as well as comparisons of graphics to tabular formats.

Finally, medicine has also been feeling the impact of computer-related advances, with such technologies contributing heavily to the design of diagnostic and testing equipment, patient-monitoring systems, and life-support systems. The potential importance of graphic displays to support medical diagnostic decision-making, as well as patient care, has been addressed by Siegel, Goldwyn, and Friedman (1971) and Cole (1986).

The above list is not, of course, an exhaustive overview of the domains in which new graphical techniques are presently being applied, but it does represent some areas in which graphical display support is being most vigorously researched. In addition to these areas, aviation continues its long-time leadership in the area of experimental analog displays. Reviews of aviation-related display research are available in Roscoe (1981).

Computer-generated displays. Most of the examples of new graphical applications and designs presented above would be impractical, if not impossible, without the use of computers to generate the graphs themselves. It has in many ways been this new capability, as well as the demand for more informative displays of mass amounts of data, that has triggered experimentation in techniques of graphic design. Some graphic notions that were considered decades or centuries ago, and were promptly set aside due to the difficulty of executing them by hand, have been rediscovered. For example, in a 1916 text on methods of graphical representation, Brinton suggests that a polygon display is both difficult to draw and difficult to

comprehend. Whether or not this display is in fact difficult to understand is a matter for empirical inquiry. However, the problem of producing this type of graph is presently of minimal concern.

A related influence of computer display technology has been the ability to easily produce digital rather than graphic displays for a variety of products. Graphics need no longer be used merely due to sheer mechanical convenience as was the case with early graphic recording. If one wants a digital output, one can often have it; digital clocks, digital speedometers, and digital displays for radio tuners are amongst these innovations, much to the chagrin of many a human factors engineer. However, the availability of such displays has made designers ask a very important question: Why should we use graphics at all? And the answers forthcoming have shown that graphics should be valued for some human-machine interactions above and beyond any consideration of mechanical convenience. In fact, one major automobile manufacturer, in the midst of that industry's infatuation with digitally-formatted instruments, has returned to the use of the less-fashionable analog dials. This return, presumably due to the realization that an analog display was more useful for many of the tasks people performed with these displays, was heralded in an ad campaign:

AREN'T YOU GLAD WE USED DIALS?
DON'T YOU WISH EVERYONE DID?

However, the choice to use dials, or to use any analog format rather than digital presentations, is only a starting point for today's display designer. Faced with an increasing variety of graphical formats, the designer must choose that which will be most effective for the task of the prospective human operator. The emphasis of the present report will be on specifying a framework that may be helpful to the designer faced with such choices.

CHAPTER 2

COMPARATIVE GRAPHICS

In making the decision of which graphic format to use, a display designer has little in the way of empirically derived guidelines on which to rely. There are, of course, a smattering of texts devoted to the use of graphic displays, with design caveats based largely on the personal experiences and preferences of the authors (e.g., Brinton, 1916; Schmid & Schmid, 1979; Everitt, 1978; Tufte, 1983). Although the intuitions of these experts may eventually prove to be accurate, few of their opinions have been substantiated empirically. For instance, if one graph is claimed to be superior to another in a particular situation, it seems quite reasonable to support this claim by experimentally comparing performance obtained when using each graph to perform the same task. This approach, termed "Comparative Graphics" by DeSanctis (1984), has been used infrequently.

It seems that a favorite line of discourse for reviewers of the literature on graphic displays is the scarcity of actual experiments validating claims of one display's superiority over another. Perhaps one of the most promising recent events has been the publication of Cleveland's (1985) recent text on graphic design that relies more heavily on empirical evidence than any of its predecessors. Brief reviews of the available literature can be found in DeSanctis (1984), Wainer and Thissen (1981), and Feinberg (1979). However, the most in-depth review of the field of comparative graphics comes from the work of MacDonald-Ross (1977). Even this exhaustive paper reports on fewer than two dozen studies carried out by a small handful of investigators.

The present review represents an update of the MacDonald-Ross work with an emphasis on derivative graphics. Research related to comparative cartography is excluded, as is research relevant mainly to specific graph features (e.g., verbal vs. pictorial labels, blue vs. red coding). Rather, studies comparing decidedly different formats, such as bar graphs and line graphs, will be emphasized. Finally, studies comparing graphical formats with nongraphical ones (i.e., alphanumeric) are also omitted, except where they include multiple graphs in their comparisons. For a recent review of the studies pitting graphical against alphanumeric formats, see DeSanctis (1984). DeSanctis also reviews some material relevant to experiments on specific graphical features, as does MacDonald-Ross (1977).

A review of comparative graphics could be organized in a number of ways--on the basis of subject, display, or task parameters, for example. In the following discussion, the primary means of organization is based on distinctions among the tasks that various investigators have used in their experiments. Thus, for instance, all those studies in which subjects used various graphs to make simple univariate comparisons will be treated in the same section. This organization was chosen because of its potential archival usefulness to those searching for displays to use for particular situations, and because of the general finding that graphical efficacy is task-dependent.

There has been a growing acknowledgment that no single graphical design will prove to be the best for every type of task that can possibly be required of an end-user (e.g., DeSanctis, 1984; MacDonald-Ross, 1977). The present review is organized with this consensus in mind and especially with the application aim voiced by Wainer and Reiser (1978) as its touchstone:

It seems to us that a catalog of display types could be prepared which would not only include categorizations of various displays but also some sort of parameterization indicating how good each display type is for each of a variety of purposes. The prospective user could then reach into this bag and pull out the one which most nearly fills all of his needs (p. 86).

The following discussion, although containing results of relatively few empirical studies, will be organized in the spirit of these objectives.

It is the long-term aim of comparative graphics to determine how to functionally classify different graphical formats. This aim is necessitated by the inherent difficulty, in the present age of graphic innovations, of specifying the complete range of graphical formats that are possible. And it would indeed be tedious to test each newly formulated display against all others for every single task of interest. Therefore, it is essential that of all the properties that can be used to distinguish or categorize graphical formats, those properties most relevant to graphical efficacy must be extracted and made part of a predictive framework that can be easily applied to as of yet unforeseen graphical alternatives.

First, however, if it is true that there exists no graphic technique that is always the best, without qualification of task demands, then some method must be found of categorizing all the tasks for which graphics are used. Such graphical task taxonomies have been proposed by various authors. MacDonald-Ross (1977), for instance, has proposed the following dichotomy of graphical tasks: 1) assessing general trends and comparisons and 2) finding exact numbers. Wrightstone (1936), likewise, used two major task classifications in an early experiment--those involving the localization of specific facts and those involved in the synthesis of facts. Bertin (1973), on the other hand, proposed a three-fold classification that included 1) elementary, 2) intermediate, and 3) comprehensive tasks. Elementary tasks, according to Bertin, involve the extraction of exact information, while intermediate tasks involve detection of trends. In addition, comprehensive tasks involve the comparison of entire sets of variables or structures one to another. Similarly, Washburne (1927) divides the use of graphics into identifying specific events, static trends (simple comparisons), and dynamic trends (comparing different trends or structures).

The present, tentative task taxonomy is an amalgamation of the task types presented by previous authors. It has been expanded somewhat so as to clearly include all of the present research in comparative graphics. The result is the following classification scheme describing four types of tasks for which graphic displays are commonly used:

- 1) Locating exact information
- 2) Simple (univariate) comparisons
- 3) Information synthesis
- 4) Complex (multivariate) comparisons

Experiments comparing various graphical formats in each of these task classifications will be discussed in turn. Not infrequently, results from the same experiment will be discussed in several of the following sections. This division of information from a single experiment highlights the wisdom of the experimenters who saw a need to generalize their results to more than one task scenario, but it may also be frustrating for some readers who would like to know how the results fit into the context of the individual experiment. In order to represent each of the experiments discussed in a more intact form, the various experiments are summarized in Table 2.1. These experiments are presented in chronological order, and brief summaries of the displays compared, tasks used, and results obtained are presented. Additionally, since some formats may be unfamiliar, readers may wish to refer to Figures 1.1 and 1.2 where a number of the graphical formats discussed in this section are illustrated.

Graphs for Locating Exact Information

Graphs are, as a general rule, poorly suited for the extraction of extremely exact values. The use of well-organized tables or other alphanumeric displays has often proved superior for this task (e.g., Sinclair, 1971; Zeff, 1965). However, even if graphs are infrequently used to present very detailed numerical data (the chief exception to this rule being the use of nomograms), the user is often faced with the task of locating or isolating an element of interest from the entire set of data displayed in some graphic format. Thus, the emphasis of this task category is more on locating information rather than on extracting very precise numeric values. For instance, in a graph showing the number of thunderstorms per month of the year, a person may want to know whether there were fewer than 10 storms in a given month. As another example, a pilot flying a multiengine aircraft may, upon the advice of the ground crew, be required to pay special attention to the status of one particular engine. Once again, this potential use of a display requires isolation of one piece of information (the status of one engine) from a format containing several such pieces of information (the status display for all engines).

Two early studies looked at the relative utility of several graphic formats for the location of exact information. In the first of these, Washburne (1927) compared line graphs, pictographs, bar graphs, and tabular formats. His subjects were several thousand junior high school students who were required to read passages regarding the economic history of Florence. Different groups of subjects received supplemental quantitative displays in one of the four general formats listed above. Subjects were then quizzed on the information contained in the displays. Not surprisingly, Washburne found that subjects who were given supplemental tables were able to more accurately produce specific numeric values on demand. Of the various graphic displays, however, the bar graph finished a close second to the tabular format. The pictographs (pictures of bags of money that varied in

Table 2.1
Summary of Studies Comparing Graphical Formats

Study	Graphs	Tasks	Results
Eells, 1926	Pie charts Segmented bars	Estimate percent of component parts	Pie charts more accurate, faster to use, and user preferred
Washburne, 1927	Bar graphs Pictographs Line graphs Numeric tables	Recall for specific facts, static comparisons, and dynamic comparisons	Bar graph best for static comparisons. Line graphs best for dynamic comparisons, and numbers best for specific facts
Croxton, 1927	Pie charts Segmented bars	Estimate ratio of component parts	Bars better for absolute accuracy
Croxton & Stryker, 1927	Pie charts Segmented bars	Estimate percent of component parts	Pie charts better in most instances
Croxton & Stein, 1932	Bar graph Squares Circles Cubes	Estimate percent smaller object's size is of larger's	Bars best, then squares and circles, then cubes
Wrightstone, 1936	Pictographs Bar graphs Line graphs Circles	Locating facts, synthesizing facts, immediate recall, delayed recall	Pictograph best for locating facts. No difference between pictograph and others for synthesizing facts and immediate recall. Pictograph best for delayed recall
Culbertson & Powers, 1959	Horizontal bars Vertical bars Grouped graphs Segmented graphs Pie charts Segmented bars	Answering comparative questions	Bars better than lines. Vertical bars superior to horizontal. Grouped better than segmented. No difference in pie charts and segmented bars

(continued on next page)

Table 2.1 (continued)

Study	Graphs	Tasks	Results
Schutz, 1961a	Horizontal bars Vertical bars Line graphs	Estimation of trend probability	Line graph faster, more accurate, more preferred
Schutz, 1961b	Multiline, single graph. Single line, multigraph	Point reading, comparisons	No difference for point reading. Single graph better for comparisons
Jacob, Egeth, & Bevan, 1976	Faces Polygons Glyphs Numbers	Subjective clustering, Paired-associate learning	Faces clustered most accurately, Faces and polygons clustered more quickly. Faces learned most quickly
Mezzich & Worthington, 1978	Faces Polygons Line graphs Polar and linear Fourier plots	Subjective clustering	Best to worst: Polar Fourier plots, linear Fourier plots, faces, line graphs, polygons
Wainer & Reiser, 1978	Segmented bars Cartesian rectangles Floating 4-fold circular displays	Sentence verification of comparative statements	Cartesian rectangles best, then segmented bars, and floating 4-fold circles
Wainer, 1980	Nightingale petals Bar graphs Line graphs	Extract exact information, detect trends and compare complex structures	Bars better for exact numbers. Lines best for complex comparisons and trend detection
Goldsmith & Schvaneveldt, 1984	Rectangles Triangles Bars	Multicue Probability Learning	Rectangles better than 2 bars, and triangles better than 3 bars

(continued on next page)

Table 2.1 (continued)

Study	Graphs	Tasks	Results
Petersen, Banks, & Gertman, 1981	Bars Stars Meters	Failure detection, Failure localization, Parameter recognition	Stars and bars best best for failure detections, with stars slightly better than bars. Meters tend to be better for para- meter recognition and failure localization
Wilkinson, 1981	Faces Stars Castles Blobs	Similarity rankings	Faces best in reliability of rankings, accuracy of recovering data structure
Brown, 1985	Andrews' plot Faces 3-D box plots	Subjective clustering	Faces outperform Andrews' plot and box plots
Cleveland & McGill, 1984	Points on common scale Points on common nonaligned scale Length Angles Circles Blobs	Estimate percent smaller object relative to larger	In order of per- formance: position on common- aligned scale, position on common- nonaligned scale, length, angles, circles, blobs

size to represent varying amounts of currency) were used much less successfully than the bars, but it was the line graphs that finished last. In sum, bars fared better than either pictographs or line graphs, but tables were slightly superior to bars.

It is important to note that the "location" of the specific information in Washburne's experiment was not made while subjects had access to the actual graphs. Instead, the ability to extract such information was based on recall of the graphs. Wrightstone (1936), however, performed a similar comparison in which over eight hundred students in grades 7 to 12 located information from graphs. Four graphical displays were used in this study, including pictographs, circular graphs, line graphs, and bar graphs. Wrightstone concludes that pictographs resulted in the best performance when used for locating facts. Unfortunately, the separate data for the remaining three display types (bars, circles, and lines) were not given. All comparisons were made between pictographs versus all other formats combined. However, given the results of Washburne's (1927) study, it would not be surprising if the performance obtained with the three nonpictorial displays was markedly different from one another. For instance, Washburne's line graph was found to be a quite poor instrument for locating exact facts, but the bar graphs were found to almost match performance obtained with tabular presentations. When data are summarized over formats associated with such different performance levels, it is difficult to draw any final conclusions about the relative merit of the pictograph. One will recall that the pictograph fared poorly in Washburne's research. However, the preeminence of this format in the Wrightstone experiment is not conclusive without a breakdown of performance in the other graphic formats. Furthermore, the pictograph used by Wrightstone differed significantly from that used by Washburne. Instead of using the size of an object to represent amount, Wrightstone used the number of pictures laid side by side to represent such a quantity. This technique is the preferred technique for constructing pictographs (see Brinton, 1916; Neurath, 1944) because it avoids the difficulties of estimating different sizes from irregularly shaped patterns. Instead, users can focus on the length of the line of objects (or on the number of objects) as an index of amount. In this case, the pictograph is merely a stylized bar graph, with linear extent being the dimension used to convey information. Thus, the different pictographic techniques used by Wrightstone and Washburne may contribute to the potential disagreement found between these two early studies.

At least one finding on which the Wrightstone (1936) and Washburne (1927) studies do agree is that it seems to be relatively more difficult to isolate information using line graphs. Though the details of her study are somewhat vague, Vernon (1952) seems to provide further data supporting the difficulty of extracting specific information using line graphs. Schutz (1961b) undertook the comparison of two quite different types of line graphs to see if one type might yield better point-reading performance than another. His line graphs were categorized as 1) single-graph, multiple-line displays and 2) multiple-graphs, single-line displays. In the first case, several lines were presented in the same frame, each superimposed on the others. In multiple-graph displays, each line graph is presented in a different frame. Schutz asked his subjects (adult professionals) to read

the value of particular points on the displays. He found that for reaction time there was no discriminable difference between the two formats. However, the multiple-graph format was more highly preferred by subjects than was the single-graph, multiple-line display for this particular task.

Wainer (1980), in testing the "graphicacy" (as opposed to literacy) of children in grades 3 to 5, included questions involving the extraction of specific information from line graphs, bar charts, Nightingale petals, and tables. The petal chart was a sort of modified pie chart. Instead of the angle of a "pie slice" varying to represent numeric values, however, the slices are equally divided (angle is held constant). What varies, instead, are the radii describing the various slices. Those slices representing larger values of some variables simply jut further off the "pie plate" than the others (see Figure 1.1L for an example). Wainer found, once again, that the line graphs were associated with the worst performance in locating information. The bar chart seemed to be associated with an intermediate level of performance, with the tables and petal charts producing the best localization.

Most recently, a study by Petersen, Banks, and Gertman (1981) compared three alternative formats for a safety parameter display system in a nuclear power plant control room. For the nine parameters displayed, they configured a panel of nine separate meters, a nine-element bar graph, and a nine-sided polygon or star display. The star display represented, basically, a line graph wrapped around a central point--a polar profile (see Figures 1.1J and 1.2B for examples). Subjects were a mixture of engineers and control room monitors, and their task was to report on a requested parameter, indicating whether it was in a normal or abnormal state. Both accuracy and latency to respond were collected. Main effects were found for both dependent variables. However, none of the planned comparisons yielded reliable results. The trend in these data suggest, however, that subjects were able to respond fastest with the separate meters, and that they were least accurate with the polygon displays. Similarly, when subjects were required to check and localize each of the nine parameters as part of a diagnosis task, Petersen et al. found the same trend of display superiority. Once again, the separate meters seemed to be at an advantage.

In summarizing the results from the above studies, it generally appears that line-type graphs, whether linear or circular, are less effective for locating specific information sources than are other graphical forms. Those graphs that appear more effective--bar graphs, pictographs, petal charts, separate meters--all have as a common characteristic a greater degree of segregation among display elements. This distinction is further supported by the subjective preference of subjects who chose the single-line multiple graph format over the multiple-line, single-graph format for point-reading in the Schutz (1961b) study.

A general limitation of the present studies is that only two of the five have been carried out with adults. Further, the two studies that used adult subjects were not sufficiently similar to the remaining studies in terms of display sample to allow any degree of generalization. The results of further studies on adult populations need to be considered.

Graphs for Simple Comparisons

The second major category of graphical tasks has been the target of more extensive empirical inquiry than any of the other categories. Not only were these tasks the earliest to receive study (e.g., Eells, 1926), but they have also been the recent target of one of the most extensive research programs in the comparative graphics literature (Cleveland, 1985). This basic task can be formulated to the subject in several different ways, and is often embedded as a requirement for other, more complex decision-making tasks. However, the studies described below generally assess the usefulness of graphs in one of two different variants of the simple comparison. They will either require the subject to make specific percentage estimates of the value of one variable compared to another, or they will require a somewhat more ordinal form of judgment such as specifying which of two values is the greater. These judgments are all subsumed under the category of "simple" comparisons because they do not require the comparison of whole data structures. That is, the comparisons are all univariate. In most instances, the subject must only locate and compare two data points. More complex varieties of comparisons will be described in a later section.

Specific estimates of relative size. The 1920s saw a heated controversy develop, predominantly in the pages of statistical journals, regarding the relative worth of two commonly used subdivided graphic forms. These forms--the pie chart (sometimes called a circle diagram) and the subdivided bar chart--were commonly used to represent the proportion of some variable or class of events relative to the whole of such events. Eells (1926) began the debate by breaking two traditions in graphical design. First, he suggested that experimental tests rather than the opinion of authorities should be used in determining which of several competing graphic forms was superior. Secondly, on the basis of his experimental results, he rejected current wisdom that favored the segmented bars over pie charts.

Eells asked subjects to estimate what percent of the whole was represented by the subdivisions of the two graph types. He found that pie charts could be used as rapidly, and even more accurately, than subdivided bars. Croxton (1927) performed a similar study, and failed to find evidence to support Eells' claim that pie charts were "a compliment to man's intelligence." In this experiment, subjects were asked to estimate the ratio of one part of a figure to another. He found a larger number of subjects made correct estimates with the subdivided bar graphs rather than the pie charts. However, in a subsequent study, Croxton (Croxton & Stryker, 1927) repeated his experiment with a larger sample of stimuli and subjects and with the requirement that subjects estimate percentages rather than ratios. With the majority of stimuli used, the pie chart was found to yield more accurate estimates. There were, however, some exceptions to this rule in the two-division graphs, notably when the charts compared divisions other than 50-50 and 25-75. However, with three- and four-part displays, the pie charts nearly always equaled or surpassed the bars.

Croxton and Stein (1932) culminated this program of research with an analysis of the relative merits of bars, squares, circles, and cubes. Subjects were shown two objects of a kind (e.g., two bars, two circles,

etc.) and were asked to judge the size of the smaller relative to the larger. This experiment differs from the previous ones in that two objects are compared, as opposed to two or more parts of a single object. Estimates based on bar charts were more accurate than estimates based on squares, circles, or cubes. Cubes, on the other hand, were the least accurately used.

Culbertson and Powers (1959), in what is perhaps a final and fitting tribute to the battle of bars and pies, compared the ability of subjects to use each form in making comparisons related to agricultural data. They concluded that segmented bars and pie charts were equally useful in comparisons of component values. However, these authors studied several other types of graphs that did differ significantly in their ability to support performance in simple comparison tasks. Their data indicate that both horizontal and vertical bar graphs surpassed line graphs, and that "grouped" graphs were superior to "segmented" graphs. The grouped graphs used in this study consisted of bar graphs grouped by variable and line graphs composed of several lines each sharing a common baseline. The segmented graphs, on the other hand, were simply segmented bars and line graphs that used one line as the baseline for the next. Thus, grouped graphs originated from common baselines, and segmented graphs did not.

Finally, a recent program of research carried out by Cleveland and colleagues (e.g., Cleveland, 1985; Cleveland & McGill, 1984, 1985, 1986; Cleveland, Harris, & McGill, 1983) has followed in the tradition of the early studies on statistical graphics, but in a relatively more theory-driven fashion. Cleveland proposed that much of what accounts for differences in the effectiveness of different graphics is the ease and accuracy with which the preattentive visual system can assess relative magnitudes. Given this assumption, the ability of subjects to make judgments of relative magnitude for various graphical elements is considered of central importance in predicting the efficacy of any graph. Thus, subjects were asked to perform comparative judgment tasks using an impressive array of commonly used graphical attributes. Tentatively, Cleveland has ranked the graphical elements, from most accurately to least accurately judged, as: position on common scale, position on nonaligned common scales, length, angle, slope, circle area, and blob area. The work of the display designer is, then, to use the graphical elements as far to the front of this list as possible.

The work of Cleveland compares well with the earlier work of Croxton (Croxton & Stein, 1932). Croxton's comparisons between pairs of bars, circles, squares, and cubes showed that magnitude judgment of position on common scale (aligned bars) was indeed superior to any of the area judgments. Note that these two tasks are ranked far apart on Cleveland's list of perceptual elements. In addition, the early work on component graphics (segmented bars vs. pie charts) is also echoed in Cleveland's work. If the segmented bar graph is used by making a length judgment, and the pie charts are used by making judgments of angles, then the conflicting results of the early studies are to be expected. These two types of judgments lie

in close proximity on Cleveland's list. Any differences found in the two displays are likely to be negligible, or are due to differences in labeling and scale construction.

In general, these studies converge on the notion that it is desirable, when possible, to use position on a common scale when making a simple comparison of a few variables. On the other hand, one should always avoid having to compare volumes or areas of any sort. Length and angle judgments as used in segmented bar charts and pie charts, respectively, yield few strongly consistent differences.

As cautioned regarding tasks involving isolation and extraction of exact amounts, the use of graphs to obtain highly exact numerical information about specific comparisons, as with the tasks presented above, may be somewhat better served by numeric displays, at least when relatively few values need to be displayed. However, the use of graphs to answer such ordinal questions as "are the two values the same or different?" and "is the second value greater than the first?" may more truly capitalize on the special properties of graphics.

Simple ordinal magnitude judgments. Washburne (1927) was an early student of comparative graphics for simple ordinal comparison tasks. While the studies of comparisons using various graphical techniques in statistics tended to emphasize exact magnitude estimates, Washburne was asking his sample of junior high school students such questions as "which merchants had a higher income in A.D. 1100?" When subjects answered such questions from memory, bar graphs were found to yield best performance, pictographs and lines yielded intermediate performance, and tables seemed least suited for such a task.

Schutz (1961b), in a study reported earlier, compared multigraph single-line displays with single-graph, multiline displays. The subjects were asked a series of questions that required them to determine which of two values was greater at a particular point on the abscissa. Thus, in effect, the multigraph format required the subjects to make position judgment on nonaligned but common scales while the single-graph situation allowed them to make position judgments on aligned scales. In accord with the work of Cleveland (e.g., Cleveland & McGill, 1985), the aligned scale (single graph) display produced superior performance. Subjects also showed strong preference for this format.

More recently, Wainer and Reiser (1978) asked subjects to verify "greater than, less than" statements using three graphical formats. These formats included a standard, segmented line graph, a Cartesian rectangle (a bar graph with grouping dictated by the four quadrants of a Cartesian graph), and a floating four-fold circular display (FCD). The FCD used here was a variant of the Nightingale petals previously described, this example having only four petals (see Figure 1.2E). All displays presented count data categorized by three variables. Once again the grouped (aligned) bar graphs proved superior to the other formats, this time in terms of reaction time. Accuracy estimates were not reported.

The data gathered in ordinal comparison tasks strongly suggest the use of aligned position judgments (i.e., grouped bar graphs or multiline trend displays) over other types of judgments (e.g., length judgments in the segmented bars or FCD). As a general note, all of the information presented to subjects in these three studies required subjects to locate and isolate two or more values from a larger data set in order to compare them. Thus, for simple comparisons embedded in larger data sets, the rules governing data localization may also be relevant. The next task situation to be discussed will focus on data sets where all the values presented must be used or integrated in generating a response.

Graphs for Information Synthesis

This category of graphics-supported tasks departs from the previous two categories in several ways. First of all, these tasks are defined, in part, by their relative freedom from locating or isolating specific data elements in the display. Instead, users must integrate all or most of the available information in order to, for example, determine the probability of a particular trend, choose the best of a number of alternatives, predict a particular outcome, or diagnose a particular "syndrome." As indicated by these variants of information synthesis, this classification includes many situations relevant to the use of graphics in a variety of professional, military, and industrial situations. As a result, much of this research has made use of adult, professional populations as subjects rather than the grade-school children used in the two previous task classifications. A further distinction is that in some cases subjects must deal with multivariate as well as univariate data sets. The graphs they use are generally generated by a computer; thus, not surprisingly, this research has taken place almost exclusively in the last few decades--in the Age of Electronic Graphics.

The earliest of these studies was performed by Schutz (1961a) and compared both horizontal and vertical bar graphs with line graphs. Subjects were professional-level corporate employees who were asked to detect trends in a data set containing up to 18 data points. Before testing, subjects were taught a set of arbitrary rules for detecting a trend and estimating its probability (e.g., six consecutive decreasing points represent a 90 percent chance of a downward trend in the smaller data sets.). After training, reaction times and accuracy measures were obtained in experimental trials with each graphic format. Schutz found an advantage for the line graph over either type of bar graph on both measures.

Schutz also manipulated the amount of missing data in the data sample. With some missing data, the line graph appeared as a discontinuous line and the bar graphs simply had fewer bars. In this condition, the advantage of the line graph vanished.

Jacob, Egeth, and Bevan (1976) conducted a study in which the subjects' task was to reliably assign names to particular sets of data. This task, in many ways, represents many real-world situations in which specialists have to recognize various states or syndromes based on interrelations among a number of variables. Jacob et al. asked subjects to identify twelve such

distinct states, using one of several graphic formats. Comparisons were made amongst Chernoff face displays, upside down face displays, polygons, glyphs, and tables of numbers.

When subjects had to learn to recognize sets of nine independent variables, upright faces provided for best performance. That is, subjects needed fewer trials to reach criterion in this condition. Glyphs, on the other hand, were the least useful of the graphics tested. The polygons were a close second to the faces. In a second condition, where subjects had to learn to classify nine highly interrelated variables, upright faces came in second to the polygon display. Inverted faces, glyphs, and numeric tabular displays were all poorly used. Finally, in a condition where only three different values were varied, the faces tied with tables for best performance and the polygons (triangles in this case) finished last behind the glyphs and inverted faces.

Jacob et al. credited the general superiority of the faces and nine-dimensional polygons to their perceptual integrality. That is, they are configurations that do not as readily allow selective attention to their individual parts, but rather are perceived in a more holistic manner--as a perceptual unit. Goldsmith and Schvaneveldt (1984), also used this distinction in making comparisons between display formats. These authors studied multicue probability learning in which subjects received multiple information cues and then estimated a criterion value associated with that specific combination of cues. In one experiment, the authors used two cues to predict a criterion and these were displayed either with a bar graph (two bars) or with the height and width of a single rectangle, a simple object display. The more integral rectangle was found to facilitate performance. And when criterion prediction was based on three variables in a later experiment (three bars vs. one triangle), the more integral triangle display was once again found to be superior.

Finally, two studies have evaluated several displays in tasks that are perhaps the most characteristic of real-world information synthesis tasks. Zmud (1978) studied subjects' preferences for various displays as used in a management decision scenario. Line graphs, bar graphs, and tables were compared. Overall, subjects preferred the line graphs, rating them as being more relevant, accurate, readable, and as presenting a larger quantity of data.

Petersen, Banks, and Gertman (1981) studied bar graphs, separate meters, and a polygon display for presenting information about nuclear power plant failures. Subjects were required to respond if any of nine safety parameters departed from normal conditions (note that they were not required to indicate which parameter failed, and thus localization was not required). Using a signal detection paradigm, these authors reported greater sensitivity to abnormal conditions with the polygon and bar displays. The polygon appeared, in addition, to be somewhat better than the bars, but this trend did not reach statistical significance. Meters were much less useful than either bars or polygons for making failure detections.

These studies, involving decidedly different scenarios, but all involving multivariate decision tasks, converge with respect to several findings. Primarily, the bar graph no longer reigns supreme in task after task, as was the case with simple comparisons and localization. Instead, when compared to line graphs, bar graphs were always inferior. Lines were found to facilitate performance for detecting trends (Schutz, 1961a), were preferred over bars in a managerial decision task (Zmud, 1978), and in the form of polygons (polar line profiles) were used more effectively than either bars or meters for a failure detection task (Petersen et al., 1981). With regard to these tasks, at least one theoretical display dimension was tabbed as a predictor of display utility--display integrality. Thus, the more unitary or holistic the graphical form appeared, the better suited it was to communicate complex data structures (Goldsmith & Schvaneveldt, 1984; Jacob et al., 1976).

Graphs for Complex Comparisons

The last of the four major categories of tasks used to compare graphic formats involved comparisons of two or more sets of variables. Thus, entire data structures must be compared in some way. Typical of such tasks are subjective clustering of data points defined by multiple variables, as well as similarity judgments and same/different judgments of such points. A major difference between this and the previous task category, both of which involve multivariate data, is that in the present tasks both sets of data are physically present for perceptual comparison. In the previous section, implicit comparisons may well take place; however, these comparisons are of necessity with some prototype or other representation in memory. This distinction between tasks is similar to the distinction between absolute judgments (information synthesis) and relative judgments (complex comparisons).

Washburne (1927) referred to the present task category as "dynamic comparisons." Specifically, he was referring to questions requiring subjects to compare values in two or more categories at two or more levels of another variable. Washburne tested junior high school students' recall of such information from tables, bar graphs, pictographs, and lines. He found that dynamic comparisons were more easily made with line graphs and were most difficult with tables of numbers.

Wainer (1980) used even younger children (students in grades 3 - 5) to test for different levels of efficacy amongst line graphs, bar graphs, Nightingale petals, and tables. He asked students to compare whole data structures and found that the line graph far outstripped performance using the other displays.

Four studies have looked at the ability of subjects to recover the structure in artificially generated multivariate point clusters. A typical subjective clustering experiment was conducted by Jacob, Egeth, and Bevan (1976) who studied face displays, polygons, and tables. Fifty data sets were generated, each consisting of nine values. The nine-dimensional vectors were constructed so as to fall into five distinct clusters generated as permutations of five equidistant prototypes. The five prototypes, one representing each cluster, were presented to subjects. These subjects were

then required to categorize the fifty data sets as belonging to one of the five categories represented by the prototypes. Results indicated that face displays were sorted more accurately than polygons, which in turn, were sorted more accurately than tables. With respect to total sorting time, both faces and polygons were sorted into groups more quickly than tables. There was no significant difference in the two object displays, however.

In a similar investigation, Mezzich and Worthington (1978) had 11 experienced psychiatrists each describe a prototypical psychiatric patient in four diagnostic classes--manic-depressive manic, manic-depressive depressed, simple schizophrenic, and paranoid schizophrenic. These 44 imaginary patients were described by the psychiatrists using ratings on a 17-variable diagnostic rating scale. Subjects were given the entire forty-four 17-dimensional vectors and were asked to sort them into four equal groups based on similarity. A different set of subjects each used one of seven graphical forms for the sorting task: linear profiles (line graphs), circular profiles (polygons), Chernoff faces, linear Fourier representations (Andrews' plots), polar Fourier representation (blobs), two-dimensional bivariate point displays generated by factor scores, and point displays of a two-dimensional multidimensional scaling solution. These authors found that the data reduced to two dimensions (the multidimensional scaling solution and factor scores) were classified much more accurately than the full-dimensional graphic forms. The best accuracy scores for the remaining displays, in order, were polar Fourier plots, linear Fourier plots, faces, linear profiles, and circular profiles. Preference scores seemed to follow the same ordering, with the reduced data sets being most preferred, and the profile methods being least preferred. Some individual differences were, however, noted. Those subjects who had the most overall difficulty with the task seemed to benefit more by use of the faces than did other subjects.

Brown (1985) has studied three graphic forms for complex comparisons in even more detail using the subjective clustering of computer-simulated data. Studied were Andrews' plots, faces, and three-dimensional box plots. Simulated data clusters were generated in four and eight dimensions, with the clusters having both low and high Euclidian proximity. Subjects were able to more accurately use faces in all cases except when there were both few dimensions (four) and the clusters were close. The greatest advantage for the faces came when subjects had to cluster data points with both a high dimensionality and low proximity.

In a variant of the subjective clustering paradigm, Wilkinson (1981) had subjects make individual similarity judgments for all possible pairs of eight 20-dimensional data vectors. These vectors were presented in four formats: faces, castles (a variant of Kleiner-Hartigan trees), blobs (circular Fourier plots), and polygons (stars). The actual distance among the eight vectors was most accurately recovered by the face displays, followed by the polygons, castles, and blobs. Further, in a test-retest situation, faces were most reliably used in making the similarity judgments.

These studies of complex comparisons seem to yield a general benefit for faces. In general, the faces' advantage tends to be attenuated in situations where the task requires fewer dimensions to be varied. This

notion is consistent with the work of Naveh-Benjamin and Pachella (1982) who found that speeded classifications were made more quickly in face displays that had more varying features. As a general rule, bar-type displays (bar graphs, glyphs, castles) were not used as well by subjects performing these tasks.

An exception to these statements is the results found by Mezzich and Worthington (1978) showing that methods that reduce the higher-dimensional data to a lower dimensionality (as with bivariate plots of factor scores and MDS solutions in two dimensions) or which emphasize the first several principle components (as in Fourier techniques such as Andrews' plots and blobs) are used better than faces. However, blobs were used less effectively than faces in Wilkinson's (1981) study, and faces were superior to Andrews' plots as used by Brown (1985). It is possible that the structure may have been such in the psychiatrist-produced data sample as to allow two variables to carry the weight of the discriminations. With less intercorrelated data (such as that generated in the simulated studies), the face display may be, in fact, more useful. Of those three techniques that did use and emphasize all 17 data variables in this study, the face display was superior to the polygon and line display. Further, the face display was particularly useful in cases where subjects were having difficulty performing the task.

Mediating Variables in Comparative Graphics

Although the present discussion of studies comparing various graphic forms has focused on task variables as potential mediators in display superiority effects, other kinds of variables almost certainly come into play. These may include such attributes of the subjects as age, education, experience, and motivation, as well as such attributes of the information to be presented as number of inputs and intercorrelation amongst variables. Finally, at the heart of comparative graphics, are the factors specific to the displays themselves that result in low or high levels of performance given a particular task, subject, and set of information. There may well be some display factors that provided for better performance in almost any condition, while other factors may be more volatile, interacting with such factors as task demand. Although admittedly based on limited experimental evidence, a few tentative suggestions can be made regarding these factors.

Task factors. The organization of the preceding review on the basis of the tasks' characteristics serves to highlight what must certainly be the major finding of comparative graphics: the efficacy of any graphic format is task specific. As a further illustration of this point, one need only browse through the contents of Table 2.1. With almost no exception, each study that used multiple tasks for testing alternative graphic forms found interactions between task requirements and preferred display format. While bar graphs might dominate simple comparisons, performance with line graphs might prove superior to bars when subjects were asked to compare whole structures (e.g., Wrightstone, 1936; Wainer, 1980). In general, bar-type graphs tended to dominate the first two categories discussed--locating specific information and making simple comparisons. On the other hand, line-type graphs or object displays tended to yield better performance when used for more complex comparisons and with information synthesis tasks.

A number of the displays discussed cannot be directly compared across the various task categories because some have not been used in more than one or two task scenarios. In fact, some of the displays would probably not be considered for use by anyone but the most sadistic of designers. For instance, using the face display for simple comparisons such that one compared the value of the mouth and the nose seems all but impossible. In this case, as Garner (1981) has concluded with regard to another topic, more may often be learned by taking careful account of the studies we know better than to conduct.

However, there are cases where surprising gaps in the choice of displays have occurred. For instance, in the simple comparisons in which only two variables are displayed (that is, no data isolation is required), it is surprising that no comparison has been made of a bar graph and a line graph. Bar graphs seem to perform better in tasks where comparisons must be extracted from a background of extraneous variables (e.g., Wrightstone, 1936; Wainer & Reiser, 1978), while lines seem to perform better in a situation where complex comparisons must be made, but where no extraneous variables must be ignored (e.g. Schutz, 1961a; Wainer, 1980). It would be interesting to compare the two in a task calling for simple comparisons but requiring little in the way of focusing. However, at present such data do not exist.

A more critical need in this research area is the better delineation of task descriptors. The present classification was fostered more by the conventions of description within the present literature base, and less on the basis of underlying theory (e.g. Bertin, 1973; Wrightstone, 1936; MacDonald-Ross, 1977). Other divisions are certainly possible, and may be more productive in formulating a model to predict performance of particular displays. However, few of the studies reviewed provided sufficient detail about the tasks actually used. Thus, the broad and somewhat nebulous categories presently used are partly a function of the lack of specificity commonly encountered in the older literature.

Additional task variables pertain to the nature of the information being presented. These task variables can vary within any of the task categories described, and include such factors as the number of information channels presented, and the degree of correlation among these channels. Such factors have been manipulated in a number of comparative graphic studies but have been of particular interest with regard to information synthesis and complex comparisons.

The general reason for the inclusion of number of channels as a factor in these experiments has been articulated by Bertin (1973). Bertin argues that comparisons of displays with only a small number of variables are tantamount to making no comparison at all. In short, almost anything can be used to present simple data sets. So, in order to truly test the potential benefits of various formats, those formats must be given a rigorous examination under conditions of high information content.

The display that has undoubtedly received the most attention with regard to data set size has been the Chernoff face display. As a general rule, whenever more variables are varied between faces (i.e., when a greater

number of facial features are manipulated), performance is enhanced (e.g., Brown, 1985; Jacob et al., 1976; Naveh-Benjamin & Pachella, 1982). Of particular interest is the Jacob et al. study, since there were three informational load conditions. Subjects were required to learn to recognize faces when either nine facial features were varied from target to target, when only three features were varied, and when nine features were varied in an intercorrelated fashion. Subjects performed best in the nine-variable condition. Both the three-variable and nine-variable correlated data sets took much longer to learn. These data are supported by the work of Naveh-Benjamin and Pachella (1982) who found that common irrelevant features of a caricature face did not influence similarity judgments supposedly based on only a few relevant features. However, distinctive irrelevant features enhanced ratings of dissimilarity based on the same relevant features. Thus, the overall distinctiveness of faces is a function of the number of dissimilar features. Goldsmith and Schvaneveldt (1984) also found in a study comparing geometric object displays and bars in a multicue probability learning task, that the benefits of the object display were greater in the three-variable than in the two-variable condition. However, in comparing line graphs and bars, Schutz (1961a) found no interaction between set size and display benefit.

Subject variables. In addition to task variables, subject variables may also play a part in determining which graph is more readily used. DeSanctis (1984) discussed two subject variables that may be related to the relative effectiveness of graphs as opposed to alphanumeric displays. These factors include the cognitive style of the subject and his or her experience with a particular graphic form. However, little work in the area of individual differences has been performed strictly with application to comparative graphics.

Jacob et al. (1976) addressed indirectly the issue of experience with regard to one particular graphic form. These authors argued along with Chernoff (1973) that it was the familiarity individuals have with the appearance of human faces that gives the face display its advantage. When they compared faces and upside down faces in a paired associates learning task, they found upright faces to be the superior display. Reasoning that the upright and rotated faces were similar in their integrality and complexity, they suggested that it was the greater familiarity people have with upright faces that accounts for their superiority.

Wainer and Reiser (1978) also noted anecdotally an initial advantage for the most familiar of the three graphical devices they tested in a simple comparison task. However, once the subjects had received more experience, a more innovative technique, the Cartesian rectangles, became the favored form. This observation points to the importance of allowing the subjects some practice with each of the displays to be compared in a given experiment. Certainly, there are situations where one wants to know the ease with which a person can immediately grasp the meaning of a particular graph, however many other situations require a subject to use particular forms repeatedly. In the latter case, knowledge of performance beyond initial practice is desirable.

Surprisingly, of the comparative studies reviewed, only that of Goldsmith and Schvaneveldt (1984) actually used level of training as an independent variable. Their conclusions, on the basis of a comparison of bars and triangles in a multicue probability learning task, was that the integral triangular display showed the largest advantage during "periods of significant learning."

Wainer (1980) deals with the issue of subject variables through the use of his concept of "graphicacy"--the general ability to use analog representations. Just as a person who has mastered some level of command of written language is considered literate, a person who has mastered the use of graphic forms is "graphicate." He found that on a test of graphicacy, there was much improvement from third to fourth grade, but little improvement from fourth to fifth grade. Since his test was developed to assess comprehension of graphs "that a literate adult would be expected to deal with in a day-to-day existence," he concluded that substantial graphicacy was achieved by the end of the elementary school years.

Other authors have studied the relation between graphicacy and aptitudes (Culbertson & Powers, 1959). These authors found a moderate correlation between number of correct items in their graphical comprehension test and tests of nonverbal, verbal, and abstract reasoning. They also found that when correlational analysis was performed on particular subsets of the data defined on the basis of particular graphical attributes (e.g., bars vs. lines), there was no difference in level of interrelation between aptitude and graphicacy. Vernon (1952), however, reported that less intelligent or well-educated individuals preferred pictorial graphics compared to other forms (e.g., bars or lines). Furthermore, Mezzich and Worthington (1978) found that their subjects who performed most poorly in a subjective classification task tended to benefit more from faces than did more able subjects. However, Casey and Wickens (1986) found no relation between spatial ability and graphical preferences.

Thus, the degree of interaction between display format superiority and subject variables remains in question. It seems likely, as Vernon (1952) suggests, that iconic forms may be less intimidating to the less able user, but this remains fairly speculative.

Display factors. In the study of comparative graphics, the lack of all but the most tentative functional classifications of display formats is surprising. Thus, in trying to generalize from the effects of various task manipulations, it is only possible to say that in one situation "bar-type" graphs seem desirable and in others "line-type" graphs are preferred. This is said without clearly being able to define what makes several graphical formats line-like rather than bar-like. And even if one could divide all graphical forms into bar vs. line graphs, would this distinction ultimately prove meaningful? Could any new graphical technique be easily classified? Would the classification ultimately prove to be useful in predicting graphical efficacy?

Several authors have, at least, addressed this issue of "functional" distinctions between graphic forms. Some early investigators seemed to

think that the most important functional dimension was whether the information channels were part of iconic or more abstract forms. Thus, Wrightstone (1936) and Vernon (1952) compared pictographs vs. a pool of other graphic forms. MacDonald-Ross (1977) also maintained this distinction of "abstract" vs. "pictorial" graphics in his reviews of the comparative graphics literature. Other authors (e.g., Jacob et al., 1976; Goldsmith & Schvaneveldt, 1984) have suggested that it might be the integrality of a graph that is an important determinant of how well that graph will support a particular task. That is, it may be important how unitarily the dimensions can be processed, or how separably they can be used, in predicting any graph's efficacy.

Tufte (1983) has suggested that the underlying variable that separates good from bad graphs is the "data-ink" ratio. His notion is that the higher the amount of data to the amount of ink used, the better the graph will be. On the other hand, excess ink, or a low data-ink ratio, will almost uniformly result in a poor graphic device. As an example of this, we have already noted that faces are used more poorly when few of their features are varied (i.e., there are large numbers of irrelevant features or "ink" relative to the actual data).

Finally, Cleveland and associates (Cleveland, 1985; Cleveland & McGill, 1985; Cleveland & McGill, 1984; Cleveland, Harris, & McGill, 1983) have argued that it is important to classify graphs according to the "elementary perceptual tasks" they require for use. These perceptual elements or tasks are divided up into requirements to judge position along common aligned scales, position on common nonaligned scales, lengths, angles, slopes, and so on. Although some limited experimentation has been conducted on these display distinctions, the usefulness of these various categories has yet to be applied to anything other than simple estimates of relative magnitude.

Interactions. The main conclusion that can be reached from these data is that graphical efficacy is almost certainly a function of interactions among a number of factors. The results of experiments comparing various graphs show that they are influenced by the task being performed, and may be influenced by the age, specific aptitudes, and experience or familiarity of the user. However, before these issues can be resolved, it seems essential to study further what elements of the displays themselves may be responsible for better or worsened user performance in any given task, with any given population of subjects.

In addition to this fine-grain analysis of display attributes in particular task domains, one aim of comparative research should be to establish a larger-scale theory to explain and predict interactions between the various subject, information, and task variables discussed and the important display attributes that are yet to be uncovered. That is, a general theoretical framework is needed to both unify some of the scattered research findings in this field, and to direct further research in a way that will foster future generalizations from specific studies. At the present time, one such theory exists in the general field of display formatting, and this theory seems promising in regard to its applicability to comparative graphics in particular.

CHAPTER 3

THE PROXIMITY COMPATIBILITY HYPOTHESIS

The Concept of Compatibility

The concept of "compatibility" has a rich history in information processing psychology and in the field of engineering psychology in particular. In its most general usage, the notion has come to mean any combination of task interface variables (e.g., display format, response form, etc.) that maximizes performance on a given task. But particular theories of compatibility have been proposed that specify more exact reasons why, for instance, different responses are better suited to specific stimulus modalities (e.g., Greenwald, 1979), or why the spatial arrangement of responses must map onto the spatial arrangement of stimuli (e.g., Fitts & Seeger, 1953). Both of these examples involve stimulus-response compatibility. In general, theories of compatibility have invoked the notion of minimizing the number of transformations that must be made on information enroute from input to output. Within such a framework, the compatibility between central processing codes and both stimuli and responses has also been considered relevant to performance.

An example of one such notion of stimulus-central processing-response compatibility is that proposed by Wickens and his colleagues (e.g., Wickens, Vidulich, & Sandry-Garza, 1984; Wickens, Sandry, & Vidulich, 1983), and is particularly applicable for predicting when analog display forms (including graphics) are likely to be used to their best advantage. These authors have formulated one answer to the question "Why use graphics at all?" Their research suggested that graphic displays should be used with those tasks thought to involve spatial codes of working memory and/or manual responses. Alphanumeric displays, on the other hand, were more compatible with verbal working memory and vocal responses. These findings, and the S-C-R compatibility hypothesis, support earlier recommendations that tasks requiring exact responses (generally associated with discrete/verbal working memory codes) are better served by numeric displays. Relative judgments, on the other hand, imply spatial processes and are thus better served by analog displays.

More recently, Wickens has introduced a new notion of compatibility to integrate findings regarding the benefits and disadvantages of displaying multiple sources of information in similar or proximal ways (Wickens et al., 1985; Polson, Wickens, Colle, & Klapp, 1986). This hypothesis suggests that the variable to consider in determining how distantly or proximally to display multiple information sources is the degree to which the task requires similar processing of information provided by these displays. In extreme instances, if a large number of variables must all be taken into account before a required response can be executed, then the task is one requiring information integration and involves a large degree of information processing proximity. That is, the inputs cannot proceed independently through the organism and still yield correct responses. In this instance, according to the proximity compatibility hypothesis, the various input sources should be displayed in proximity. At the other extreme, if several

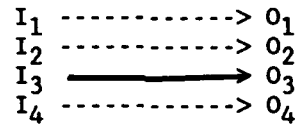
information sources are to be used in several completely independent information processing tasks, each with its own response, then task proximity is low and display of the elements should emphasize their separability through low proximity display manipulations.

These two examples of task proximity represent two endpoints on a continuum of information processing proximity from total independence to complete information integration. There are, of course, some tasks that may involve both types of processing to a greater or lesser degree. Some of the various task situations that can be specified using the present classification scheme are diagrammed in Figure 3.1A-D. These diagrams are taken from Wickens et al. (1985) and formally describe several types of task proximity relations. Figure 3.1A represents a focusing task in which a number of inputs are present, but the value of only one (or some subset) is relevant for making the correct response. Note that many of the tasks subsumed under the task classification of "extracting exact information" are well-represented by this diagram. The focusing task is a nonintegration task, as is the task represented in Figure 3.1B. Here, each of several variables is associated with its own response, and each information source is independent of the others. An example of this task can also be drawn from the "extraction of specific information" category. In the case of Petersen et al. (1981), subjects had to locate or diagnose failures in a system of nine variables. This task can be conceptualized as nine stimuli each associated with a go, no-go response. Only if a particular variable were to take on a value that was out of bounds should its associated response be made. Thus, the first category of tasks reviewed in the previous section may be classified as nonintegration tasks.

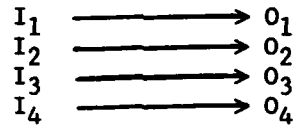
The simple comparison tasks of the previous section presents a compromise between the nonintegration situation and the integration task. Here, there is a need to focus attention on a limited number of variables, but these variables must then be integrated in order to yield the appropriate response. This situation is shown in Figure 3.1C.

Finally, both the complex comparison and synthesis tasks represent true information integration tasks as schematized in Figure 3.1D. Here, several pieces of information must be taken into account in order for a single response to be executed. There is no way that a response can be made based on a single variable or subset of variables. To the degree that a subset of the presented information can be used to perform the task, the subject can choose to focus on a limited number of variables to perform the task, thus making it more of a focusing, nonintegration task than a proximal integration task. An instance in which such a strategy may be used by the subject is when the input variables are highly correlated.

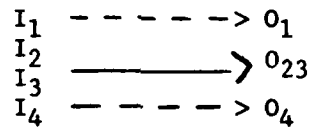
All in all, those tasks involving extraction of specific information may be considered nonintegration tasks, while complex comparisons and data synthesis tasks may be categorized as integration tasks. Simple comparisons, when nested in a larger data set, are a compromise between the two. According to the proximity compatibility hypothesis, then, the extraction of specific information should be facilitated by the less proximal or similar displays, while the synthesis of information and complex comparisons should be



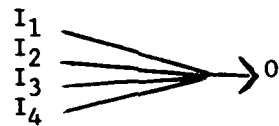
A. Focusing task where subject only responds to I_3 (low proximity).



B. Four concurrent tasks, each with its own input and output (low proximity).



C. Combination focusing and integration task (intermediate proximity).



D. Total integration of four inputs (high proximity).

Figure 3.1. Some examples of mapping proximity.

facilitated by display proximity. Simple comparisons, on the other hand, may be facilitated by relatively similar or proximal displays only to the degree that data isolation is not required.

Having defined task proximity, the next step required for applying the proximity compatibility hypothesis is finding a satisfactory definition for display proximity. Wickens has described display proximity in terms of physical/spatial proximity. Thus, the closer two information sources are in space, the greater the degree of "display proximity." In addition, other forms of proximity may refer to such Gestalt characteristics as whether two or more features form perceptual groups or units. Thus, two features that form part of two separate groups, units, or objects are less proximal than two features that form a part of a single group. The integral-separable dimension that Jacob et al. (1976) and Goldsmith and Schvaneveldt (1984) suggested as being influential fall into this category of proximity measures. Integral display dimensions, those treated by the organism as one rather than several dimensions, are thus more proximal than separable dimensions.

Given these admittedly crude measures of task proximity (integration vs. nonintegration) and display proximity (physical proximity and integrality vs. physical distance and separability), the conclusions drawn from the review of comparative graphics can be restated. In situations where exact data extraction is required (nonintegration task), those displays with less proximity between elements will result in better performance relative to the more integral displays. Thus, several separate bars (or pictures, in the pictorial graphs) provide for better performance than do lines that connect several points into a single unified contour. In the situations where simple comparisons must be made while extracting these from a larger data set, results are likely to be more ambiguous. Thus, for the most part, separable bar graphs served to an advantage. However, proximity in one situation also showed an advantage when in Schutz (1961b) comparisons were found to benefit from superimposed line graphs relative to graphs presented in separate frames. Note that no difference between the formats was found for simple point reading.

For those tasks requiring information synthesis and complex comparisons (integration tasks), the more integral or proximal stimuli seemed to yield superior performance. Face displays tended to be better than, for instance, separate bars. Line graphs, in this context, also outperformed bars. Once again, Schutz (1961a) sheds some light on the present distinction. The superiority found for the line graph in a trend detection task was lost when missing data were included in the sample data set. This condition effectively switched the continuous line to a discontinuous one, in which case it showed no advantage over the separable bar graphs against which it was compared. Thus, the configuration that presented data in the most unified way tended to enhance performance with these tasks.

The reviewed literature seems to fit quite readily into the framework of the proximity-compatibility notion. However, some studies could not be fit into the framework, predominantly because the notion of display proximity has not been specified well enough to differentiate, for example, the proximity of face displays relative to blobs or castles. However,

several further studies exist that were designed to directly test the notion of proximity compatibility between tasks and displays.

Experiments on Proximity Compatibility

Integration tasks. Carswell and Wickens (1987a) have studied display proximity in a simulated process control failure detection paradigm. Their task involved subjects observing displayed input and output variables for hypothetical systems. Subjects were instructed to detect any discrepancy from particular input-output relationships. Both a separable bar graph and an object display (triangles) were used to perform the task. Performance using the more integral triangles surpassed that obtained with bar graphs.

Studies by Jones and Wickens (in press) and Casey and Wickens (1986) also compared integral object displays and separable bar graphs in information integration tasks typical of the process control environment. Jones and Wickens (in press) had subjects use either pentagon displays or staggered bar graphs to perform a task that required the integration of five values to yield a reading of "average system state." In this scenario, the pentagons were found to be superior to the bar graphs. Casey and Wickens (1986), however, failed to find any display advantage for a failure detection task that required subjects to indicate when any of five values departed from their normal correlated structure. The displays compared in this experiment included bars, faces, and pentagons.

A pair of studies was performed by Goettl, Kramer, and Wickens (1986) on the ability of subjects to extrapolate from multivariate data sets. In the first of these studies, subjects were shown concocted results from two different conditions in a fabricated experiment. The results of each condition consisted of two dependent variables, and the subjects were required to estimate what the value of a third condition might be based on the results they had seen. Subjects were shown either bar graphs or bivariate point displays to represent the data, with the point display being considered the more proximate display of the two. As predicted, subjects were better able to extrapolate to a third set of values when the point displays were used. However, in a second experiment, Goettl et al. (1986) found no display advantage for a triangular object display over a three-bar graph when three cues had to be used to predict a criterion value.

Finally, Barnett and Wickens (1988) have studied the ability of subjects to integrate probabilistic information from a number of sources. In their study, they represented each multivariate data source as either bar graphs or rectangles. The rectangles were also of two types, either being spatially distant, or being contiguous with one another. Thus, three levels of proximity were used to represent the data--bars (low proximity), rectangles (moderate proximity), and contiguous rectangles (high proximity). Results indicated that both rectangular displays were superior to the bar graphs. In addition, a nonsignificant trend favored the contiguous rectangles over the distinct rectangles.

Nonintegration tasks. Fewer studies have been aimed specifically at the situation in which information does not require integration. Carswell

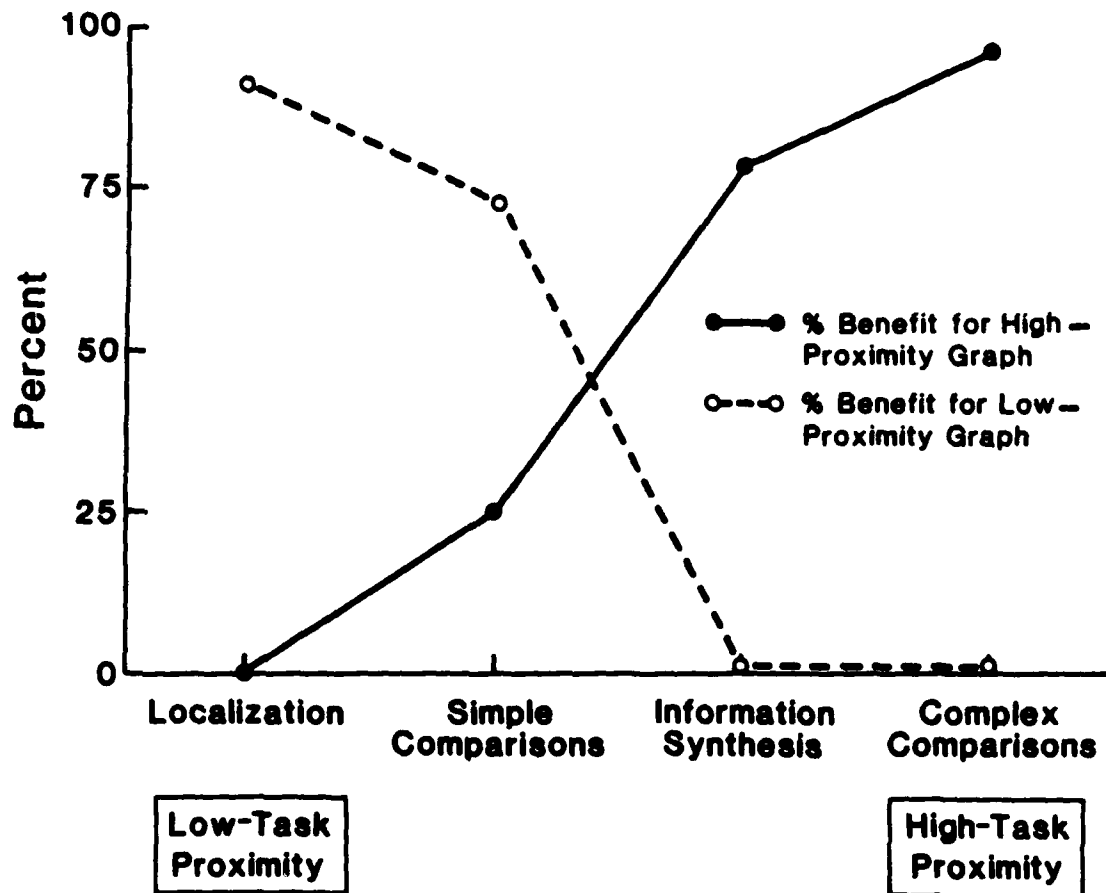
and Wickens (1987a), in a second experiment, found that the display proximity advantage they had found using a triangle display disappeared when the requirement to integrate multiple sources was dropped. When subjects were required to process each of six information sources independently in six separate detection tasks, the bar graphs proved to be the better format. Casey and Wickens (1986) also found an advantage for bar graphs in a task that required localization of a failed unit from amongst a larger set of variables, a focusing task. In another focusing task, Goettl et al. (1986) found that when one of three cues had to be ignored in order to make a correct estimate of a criterion variable, bar graphs provided for superior performance compared to the more proximal triangle displays. These experiments thus support the notion that when focusing or independent multitask processing is required, more separable forms of information representation should be chosen.

However, two experiments present something of a puzzle at the present time. These experiments deal with the recall of specific information, or focusing in memory. In two such tasks where subjects were periodically cued to recall information presented as part of an integral object display or as part of more separable bar graphs (Barnett & Wickens, 1988; Carswell & Wickens, 1987b), no disadvantage for the integral displays was found. In the Carswell and Wickens study, the display that provided best memory support was dependent on the primary task the subject was required to perform at the time of recall. These findings are in conflict with the earlier work of Washburne (1927) who found a decided advantage for bar graphs in the recall of specific values. However, the present studies required immediate recall and the Washburne study focused on relatively long-term recall performance.

Evidence for the Proximity Compatibility Hypothesis

Some general conclusions from the research specifically aimed at testing the notion of proximity compatibility and the research on general comparative graphics are presented in Figure 3.2. Almost all studies from the comparative graphics literature previously reviewed are included in this summary graph. However, three studies (Mezzich & Worthington, 1978; Wilkinson, 1981; and Wainer & Reiser, 1978) were excluded because the displays they used could not, under the present crude definition of display proximity, be judged as more or less proximal with reference to one another. The remaining data do support the notion that when integration tasks are performed, the display proximity advantage is much more likely to occur than when a nonintegration task is required of the subject.

In addition to the present work on proximity within the graphical format, the proximity compatibility hypothesis has been applied to situations involving the mixing of graphic and alphanumeric displays. In this situation (Boles & Wickens, 1983) when subjects were required to integrate information, performance was fostered by having all the information displayed either graphically or numerically. On the other hand, if independent tasks were performed upon two information sources, the task benefited from more dissimilar displays. Thus, the subjects were better able to use one numerical and one graphical display simultaneously.



- Petersen, Banks, & Gertman, 1981
 - Wainer, 1980
 0 Schutz, 1961b
 - Wrightstone, 1936
 - Washburne, 1927
 - Casey & Wickens, 1986

- Wainer, 1980
 + Schutz, 1961a
 - Culbertson & Powers, 1959
 - Washburne, 1927

+ Jones, 1986
 + Petersen et al., 1981
 + Goldsmith & Schvaneveldt, 1981
 + Jacob et al., 1976
 0 Wrightstone, 1936
 + Carswell & Wickens, 1987a

+ Wainer, 1980
 + Jacob et al., 1976
 + Washburne, 1927

STUDIES: 0 = No difference
 + = Proximity advantage
 - = Proximity disadvantage

Figure 3.2. Advantages of low- and high-proximity graphs used in four types of tasks.

With the wide applicability of the proximity compatibility hypothesis to display formatting, and the present weight of the empirical evidence in its favor, the present report will retain the proximity compatibility hypothesis as its working hypothesis. The aim of future research, then, should be to refine the concept of display proximity by studying and comparing various alternative definitions over a relatively large number of displays. With this aim in mind, we will now turn to a discussion of the psychological underpinning of the concept of display proximity and to several alternative methods of determining degree of proximity.

CHAPTER 4

DEFINITIONS OF GRAPHICAL PROXIMITY

In the previous chapters, the authors have considered experiments designed to compare graphic formats. As a general rule, use of different graphical formats tended to result in differential performance on many tasks; however, no single format was found to be superior over all or even the majority of the tasks studied. While informational and subject factors both showed some evidence of mediating display superiority effects, it seemed that the most consistent determinant of display superiority was the nature of the task itself.

The proximity compatibility hypothesis (Wickens et al., 1985; Polson et al., 1986) was introduced as a framework for describing the pattern of interactions between stimulus (displays) and central processing requirements (task demands). For the most part, the conclusions of many comparative graphics studies were accurately described by the proximity compatibility hypothesis (see Figure 3.1). Studies using tasks that demanded integration of the information from numerous information channels, such as information synthesis and complex comparisons, tended to be performed better with those displays possessing greater "display proximity" (i.e., any display manipulation that increases the similarity or unitariness of the physical dimensions used to present information). Conversely, when the task requirements emphasized the independence of information channels, high proximity displays demonstrated less of an advantage or were even detrimental to performance.

Table 4.1 provides an outline of the critical relationships between task and graphical proximity as predicted by the proximity compatibility hypothesis. The upper left and bottom right quadrants of the figure represent those task-graphical display combinations that should be most compatible. That is, when high graphical proximity is paired with high task proximity, or when both task and graphical proximity are low, performance should be relatively more efficient. Although this relation may seem quite straightforward, its immediate application to the comparative graphics literature is not without its problems.

A major obstacle to application is the imprecise definition of graphical proximity that we presently use. It is essential that we establish proximity measures that are sensitive enough to make fine discriminations between various graphic forms and at the same time are easily applied objectively. Certainly, as a starting point, discriminations of proximity can be made in terms of a physical (spatial) distance measure or in terms of whether the channels displayed are part of a single object or are distributed over several forms. However, these tentative definitions quickly run into problems when an attempt is made to apply them to the literature. For instance, most researchers have made some attempt to equate the size of their displays (one possible measure of spatial proximity); yet, most studies found reliable format differences. Further, in some studies, the object/nonobject distinction proved useless since all the tested displays were object displays (e.g., Wilkinson, 1981; Mezzich & Worthington, 1978). Once again, reliable

Table 4.1

Compatible and Incompatible Matches of Task and Display Proximity

	High-Task Proximity	Low-Task Proximity
High-Display Proximity		INCOMPATIBLE
Low-Display Proximity	INCOMPATIBLE	

differences were found amongst these formats, thus indicating that the object-nonobject distinction, like spatial proximity, is probably not a sufficient discriminant of display efficacy for a given task.

Further, in some cases, different potential indexes of proximity were in apparent conflict. For instance, the face display may constitute a single object display. However, the parts of the face used to represent the various information sources require quite different discriminations on the part of the user. The tilt of the eyebrow, size of the head, and color of the eyes may all be used as information sources. Therefore, even though the face display may be proximal in terms of its "objectness," there is a great deal of heterogeneity in the features used to display the information. By comparison, a bar graph is made up of several perceptual objects rather than just one. However, the discrimination required for each variable is the same--height of a rectangle. Which of these measures of proximity should be considered dominant--objectness or dimensional homogeneity? Are the proximity effects additive, or do they interact in some important way? Is there some useful composite measure of proximity that includes both factors? These questions are indicative of some of the uncertainties surrounding the meaning of display proximity in graphic design.

Task Proximity

Before discussing some alternative ways of defining stimulus proximity, some observations regarding task proximity should be made. Essentially, there are two ways in which task or information processing proximity have been described by Wickens (e.g., Polson et al.)--mapping proximity and cognitive proximity. These two definitions, while not mutually exclusive, may have slightly different implications for describing task requirements within the context of the proximity compatibility hypothesis.

The first type of information processing proximity borrows heavily from the definitions of perceptual independence outlined by Garner and Morton (1969). In this context, independence is analogous to low task proximity and nonindependence is analogous to high task proximity. To determine the degree of processing independence or nonindependence required by a particular task, one looks at the optimal mapping of stimuli to responses, hence the name "mapping proximity."

In general, mapping proximity deals with the degree to which several inputs and one or more responses can really be considered a single task rather than several independent tasks. To illustrate, suppose there are two inputs, I1 and I2, and two outputs, O1 and O2. To qualify as having low mapping proximity, the variation in O1 should reflect the variance in I1, but should not reflect any of the variation in I2. The same result should hold for O2 and I2, with O2 being exempt from the variation of I1. Thus, for each response, the associated input is sufficient; other information is irrelevant. Such a description indicates that there are multiple tasks to be performed, that independence between different stimulus-response pairs should optimally exist. On the other hand, if the variance in either response is jointly

determined by the variance in I1 and I2, then high mapping proximity is indicated. In this case, the many-to-one mapping of several stimuli to a single response may be characterized as a single integration task.

Referring back to Figure 3.1, lines connecting stimuli to responses indicate some degree of correlation between that input channel and response. Visually, those tasks that appear crossed, or have many-to-one mappings of stimuli to responses, are the more proximal tasks. Those tasks, on the other hand, that are characterized by more parallel mappings are the independent, low proximity tasks. An extremely high level of proximity is exemplified by the total integration task, in which information from four channels is required for a single response. Lower mapping proximity is present in both the multitask and focusing examples. It should be noted that the research in selective attention commonly distinguishes between focusing tasks and divided attention tasks. In general, focusing refers to the selective use of stimuli from a multidimensional display, while divided attention tasks emphasize the concurrent use of the multiple dimensions available. In the present classification of tasks, all focusing tasks have relatively low mapping proximity, but only some divided attention tasks may be so characterized. Some divided attention tasks may be more adequately characterized as integrative, high proximity tasks.

A second type of task proximity is defined in terms of hypothesized central processing requirements. Thus, while the mapping definition of task proximity relies on a specification of $s - r$ covariation between multiple stimuli and responses, cognitive proximity is indirectly defined in terms of hypothesized central processing constructs. An example of cognitive proximity might be same versus different code of processing, processing of conceptually related concepts, or use of same versus different internalized scale.

Cognitive proximity will not be used for the initial development of predictions based on hypothesized relations between graphical and task proximity. However, it should be acknowledged that, like graphical proximity, more than one type of task proximity may be specified and applied via the proximity compatibility hypothesis. For instance, using the cognitive conceptualization of task proximity, Harwood, Kramer, Wickens, Clay, and Liu (1986) found that subjects performed a complex identification task best when units of each of two conceptual groups were displayed in proximity. Graphical proximity in this case was defined as having either all the information about a single conceptual unit spatially proximal or similar in color. This general benefit for stimulus proximity of conceptually related inputs was maintained regardless of whether mapping proximity dictated information integration within or between conceptual units. In the present review of graphical proximity concepts, mapping proximity will be used to describe task proximity. This choice has been made because of the relative ease of applying the more objective mapping distinctions compared to the somewhat more ambiguous distinctions required for statements of cognitive task proximity. However, the potential importance of such task proximity concepts, emphasized long ago as the imperative for the development of one of the oldest three-dimensional object displays (Playfair, 1801), must certainly be acknowledged.

Graphical Proximity: An Overview

There are a number of ways to define proximity in a graphic display. Some of the distinctions previously used by researchers in comparative graphics will form a starting point. For instance, dimensions used to represent two or more information sources may be more or less integral (e.g., Goldsmith & Schvaneveldt, 1984; Jacob, Egeth, & Bevan, 1976). Dimensions can be closer together in space (Schutz, 1961b). The variables may be displayed as part of the same perceptual object, or as different objects (Carswell & Wickens, 1987a,b; Casey & Wickens, 1986; Barnett & Wickens, 1988). In addition, the dimensions along which task-relevant variation occurs may require either similar or different discriminations (e.g., two orientation discriminations vs. one color and one orientation discrimination). Each of these variables suggests different ways of measuring graphical proximity.

The present review of basic research relevant to these tentative measures of proximity will be divided into three sections. The first of these will look at dimensional dependence. This category of proximity measures includes those theoretical concepts related to the psychological unitariness of physically manipulable display parameters. For example, the concepts of dimensional integrality and configurality will be discussed in this section. The following section will be devoted to dimensional homogeneity. This category will include a discussion of the impact of similarity between the dimensions that are varied to present relative magnitudes of variables. For example, what is the implication of using height of two identical bars to represent values on two variables rather than the height of one bar and the color of another? Finally, a third category relates to the objectness of display parameters, high proximity being defined as information presented within a single object rather than divided amongst several. This proximity distinction is particularly relevant to the usefulness of object displays or iconic graphics. Further, this category is likely to combine much of the background of the other two types of proximity. That is, objects are likely to be more perceptually unitary, to involve more dimensional dependencies, than attributes of several separate objects; however, they are likely to use different attributes to present various information sources, for example the height of the "trunk" of a tree versus the angle of a "branch." Thus, proximity in terms of both dimensional dependencies and dimensional homogeneity may be relevant to discussions of object displays.

The discussions of each of these three candidate definitions for graphical proximity will follow a similar pattern. A background discussion will include pertinent theory and research suggesting the role of relevant display variables on information processing outcomes. This theoretical background will be followed by a discussion of the methods used to define or measure the type of proximity under consideration; and finally, the candidate proximity measure will be used with the proximity compatibility hypothesis to make some general predictions regarding the usefulness of different graphic formats for tasks that vary in degree of mapping proximity.

Dimensional Dependence

The first class of proximity measures to be discussed pertains to the degree of unitariness or sameness (in the sense of being part of the same thing) in any collection of physically variable dimensions. The issue is whether or not two dimensions are perceived predominantly as a singular source of variation, or as separate sources. This distinction takes into consideration the possibility that the dimensional structure set forth by the experimenter for his or her stimulus set may not be the dimensional structure the subject uses, or even perceives.

Various researchers have used different names to refer to stimulus dimensions that seem relatively unitary as opposed to more distinct. For instance, Shepard (1964) contrasted unitary with analyzable dimensions. Lockhead (1966) coined the terms integral and nonintegral to refer to similar concepts. Recently, Cheng and Pachella (1984) have referred to relatively inseparable dimensions as nonpsychological, while more separable dimensions are said to correspond to psychological dimensions. However, the most commonly used terminology comes from the framework for dimensional relations established by Garner (1970, 1974). In this framework, dimensions are either integral or separable.

Integrality: Theoretical background. Garner's (1970) first major statement of the distinction between integral and separable dimensions came in a plea to information processing psychologists to pay more attention to stimulus variables in their experiments. He cites, as an example of this oversight, the studies on parallel versus serial processing of multidimensional stimuli. In these cases, he argues, psychologists rarely pay attention to the question of whether the stimuli used are truly multidimensional to the subject (i.e., made up of separable dimensions). Thus, the distinction between integrality and separability must be made prior to any distinction between parallel and serial processing.

By noting that some of the discrepant results in the information processing literature became interpretable when the concepts of integrality and separability were applied, Garner demonstrated the utility of these stimulus variables. In addition to reviewing previous research, Garner also demonstrated in his own experiments the differential effects of integral and separable dimensions (Garner, 1970, 1974, 1976; Garner & Felfoldy, 1970; Gottwald & Garner, 1975). That is, Garner's approach was to look for tasks in which there was some evidence of stimulus-specific outcomes. Then, he looked for a convergence in the results observed with particular types of stimuli (i.e., integral vs. separable) over the various tasks. Similarity judgments, free classification, restricted classification, absolute judgments, concept learning, choice processes, and speeded classifications were among the tasks either reviewed or directly tested by Garner and his coworkers.

One of the first tasks reviewed was scaling of direct similarity (or dissimilarity) judgments. The discrepancy in the results from similarity judgments involved the type of distance relation that best characterized any particular set of dimensions. For some pairs of dimensions, a simple addition of the relevant unidimensional dissimilarities was sufficient to estimate the

perceived multidimensional dissimilarity of stimulus pairs. This method of calculating dissimilarities or "distances" was termed the "city-block" metric. The city block metric seemed adequate to describe the multidimensional dissimilarities of such stimuli as the brightness and size of a single form (Torgerson, 1958), the size of a circle and orientation of its radius (Shepard 1964), the color and shape of a single form (Handel & Imai, 1972), and the brightness of one color chip and the saturation of another (Hyman & Well, 1967, 1968). However, for other multidimensional stimuli, this simple additive definition of dissimilarity tended to overestimate the perceived multidimensional dissimilarity between stimuli. These stimulus dimensions were best fit by a Euclidean metric. Saturation and brightness of a single color chip (Torgerson, 1958) represent such a dimensional pair. Shepard (1964) suggested that those stimuli that were fit well with the city-block metric were relatively analyzable. Lockhead (1966) referred to stimuli that were fit by the Euclidean metric as integral.

Other tasks that seem to show integrality effects included restricted and free classification. In these tasks, subjects are presented with a subset of stimuli taken from a set defined by two or more dimensions. The subject is asked to sort the stimuli into a specific number of categories (in restricted classification) or into any number of categories (free classification). The particular classification chosen by the subject is then analyzed to see if it is based on the experimentally manipulated dimensions. If such a classification is frequently chosen by subjects, it is supposedly indicative of the salience of the individual dimensions, and hence of separability. Accordingly, Handel and Imai (1972) found that size and brightness tended to yield such classification. On the other hand, they found that brightness and saturation of a single form tended to yield classifications that were best described by interstimulus similarity in Euclidean space. These results were attributed to stimulus integrality.

Although the results from direct similarity scaling, restricted classification, and free classification tasks showed some convergence with regard to the integral versus separable concepts, the task most closely associated with the distinction is the speeded classification paradigm (e.g., Garner & Flowers, 1969; Garner & Felfoldy, 1970). This task, even with the more recent usage of computer-based stimulus presentation, is sometimes called "card sorting." The subject is required to indicate into which of two categories each of a series of stimuli belongs, with mean sorting time per stimulus set being the major dependent variable. The stimulus set is usually formed of two dichotomous dimensions combined orthogonally (i.e., four possible stimuli in all). Three types of tasks are performed with series of stimuli constructed in this manner. In control tasks, only two stimuli are used in the test series. Subjects must sort stimuli on the basis of the value of only one of the two dimensions; the second dimension is always held constant. Likewise, in the redundancy condition, only one dimension is formally defined to be used in distinguishing category membership, but the irrelevant dimension varies redundantly with the relevant dimension. Finally, in the orthogonal classification set, the subject makes classifications on the basis of only one dimension while the irrelevant dimension varies randomly from trial to trial.

The general pattern of results obtained with integral dimensions is that, relative to control classification tasks, classification of the relevant dimension is facilitated when the irrelevant dimension is varied redundantly. However, when the irrelevant dimension is varied orthogonally, sorting speed is impaired. With separable dimensions, on the other hand, neither of these results occurs. That is, sorting time is roughly equivalent regardless of whether the irrelevant dimension is fixed, varied orthogonally, or varied redundantly. With separable dimensions, responses to the relevant dimension are independent of the irrelevant dimension.

The separable pattern of results has been found with brightness of one color chip and saturation of another, with size and line orientation of circles, and with color and form. On the other hand, redundancy gains and orthogonal decrements (integral patterns) have been found with saturation and brightness of a single color chip, vertical and horizontal position of a dot, and auditory pitch and loudness of a monosyllable. In general, those stimulus dimensions that, using other paradigms, resulted in dimensional classification and city-block metrics, were those associated with no facilitation or interference. Those stimulus pairs associated with both redundancy gain and orthogonal interference were associated with similarity classifications and Euclidean metrics.

To summarize, Garner (1970, 1974, 1976) demonstrated that seeming inconsistencies in data from several different paradigms showed some cohesiveness when the stimulus concepts of integrality and separability were invoked. Other authors have gone on to add more tasks to the list whose information processing outcomes show some dependence on the presumed integrality or separability of the dimensions used to convey multiple information sources. For instance, Boer and Keuss (1981) studied interference from orthogonally varying irrelevant dimensions in two-stimulus matching tasks. Their findings indicated that with three-dimensional pairings, the ranks of interference effects were the same as those obtained with speeded classification of orthogonal sets. In addition, Garner (1976) has also cited evidence for integrality effects in concept formation and choice decision tasks.

Most recently, Treisman's theory of feature integration has generated several new diagnostics for integral and separable dimensions (e.g., Treisman, Sykes, & Gelade, 1977; Treisman & Gelade, 1980; Treisman & Schmidt, 1982). According to this conception, attention acts to conjoin different features. Treisman suggested that values along two integral dimensions should behave as a single feature. Thus, such dimensions should allow parallel search, could form the basis for texture segregation, and could allow identification without localization. On the other hand, values along two separable dimensions should behave as two different features. In order for such values to be integrated, selective attention is required. Thus, conjunctions of separable features should require serial search, should show little difference in identification and location times, and should not form the basis of texture segregation.

These additional studies, along with the initial work of Garner and colleagues, emphasize the status of dimensional integrality as a stimulus variable with implications for a wide range of information processing tasks.

However, Garner's system for describing dimensional dependence has also received some criticism. The arguments leveled against the system can be roughly divided into two types. First, the degree to which the integrality of a set of stimulus dimensions is impervious to organismic factors such as task strategy and practice has been questioned. Second, the not infrequent failure of various performance results to converge perfectly has worried some observers.

Garner (1970, 1974) has stated that integrality is a mandatory property of the stimulus; neither the strategy adopted by the organism nor the amount of practice can make the dimensions behave separably. However, recent research has revealed that some subject variables do indeed influence predicted performance outcomes with ostensibly integral dimensions. Dykes (1981) and Dunn (1983) have demonstrated that integrality can be attenuated by subject strategies. Further, Dykes (1979) has shown that subjects can, with extended practice, selectively attend to dimensions that at first produced results indicative of integrality. Lorch, Anderson, and Well (1984), on the other hand, found that practice in a speeded classification task was needed even with separable dimensions before orthogonal interference effects were eliminated. Additionally, Smith and Kemler (1977) and Ward (1980) have found developmental differences in the way individuals classify multidimensional stimuli, with adults being more likely to make the dimensional classifications characteristic of separability.

Many of these findings are consistent with the distinction Garner (1974) makes between mandatory and optional processes. While integrality forces a mandatory distribution of attention to both dimensions, separability between dimensions allows some processing options. That is, separable dimensions can be characterized as having mandatory or optional selection. Strictly speaking, separable dimensions may be truly separable (in that separation is an option), or they may be separate (nonoptional separation of dimensions). "Configural dimensions" are an example of dimensions that might be considered separable in this sense. Although configurality will be discussed more thoroughly in the next section, an example of configural dimensions is the height and width of a rectangle (Pomerantz, 1981). An example of completely separate dimensions, on the other hand, might be the height of one rectangle and the width of another. Since several of the studies showing strategy or practice effects used height and width of rectangles as their "integral" dimensions (Dykes, 1979, 1981), these results can be accounted for in Garner's (1976) expanded framework.

Dimensions may then be, among other choices, mandatorily separate, separable, integral, or configural. This more intricate description of dimensional relations leads to the criticisms raised by Cheng and Pachella (1984) and Pachella, Somers, and Hardzinski (1981). Namely, these researchers have objected that the converging operations defining integral dimensions have often failed to converge. As a result, the dimensional taxonomy proposed by Garner and coworkers has seemed to be an ever-expanding one, a taxonomy that has come to include degrees of integrality, asymmetric separability (i.e., where one dimension can be processed separately from a second dimension, but not vice versa), degrees of asymmetry, and the like. Thus, they argue, the explanatory and predictive power of the integrality concept is being constantly diluted. Garner (1970), however, has maintained for some time that

integrality is likely to be a continuum. That is, some pairs of dimensions are going to be all but impossible to extract independently from one another, no matter how much additional (secondary) processing is provided. Alternatively, other pairs will only marginally benefit from redundant variation and will be only weakly disrupted by orthogonal variation in selective attention tasks. Phenomenologically, this postulated continuum of integrality may seem reasonable. However, the notion of degrees of integrality makes a qualitative, logical definition of the sort proposed by Garner (1970) somewhat difficult. Instead, it argues for the use of quantitative metrics.

The present research will take as its starting point the large body of literature that shows some convergence in results leading to a general construct of stimulus integrality. That integrality is primarily a stimulus concept will also be retained, although the possibility that some organismic concepts can intervene in some cases (e.g., optional strategies with separable dimensions) must be acknowledged. Further, the notion of integrality as a continuum will be a working hypothesis, the adequacy of which will be compared to categorical a priori definitions of integrality and their ability to predict graphical efficacy in a variety of tasks.

Definitions of integrality. In recent years, definitions of dimensional dependencies have mainly been operational. In particular, definitions based on the convergence of performance in several tasks, like those described above, have been used to delineate various classes of dimensional relationships. The favorite operational definition of integrality, it seems, makes use of the results of the speeded classification tasks with correlated and orthogonal variation between dimensions. That is, integrality is associated with performance gains when dimensions are correlated, but suffer from decrements associated with orthogonal variations. These definitions, however, are not the only ones used to define integrality.

Monahan and Lockhead (1977) have reviewed the various definitions--operational, phenomenological, and logical--that have been proposed for integral dimensions. For instance, a phenomenological definition given by Lockhead (1966) can be seen as the logical predecessor of the operational definitions summarized above. He originally stated that integral dimensions were those with which we have difficulty attending to one aspect or dimension without being aware of the other aspects. This, of course, translates into the predictions of an influence of irrelevant integral dimensions in focused attention paradigms such as speeded classifications.

Logical definitions of integral dimensions, based on physical relationships of stimulus attributes, have also been proposed. The most noteworthy among these was Garner's (1970) a priori definition of integral dimensions: In order for one dimension of a two-dimensional integral stimulus to exist, the stimulus must have some value on the other dimension. Monahan and Lockhead (1977) suggest a modified version of this logical definition. They suggest that two dimensions of a stimulus are integral if removal of a physical aspect renders the other aspect unspecifiable or if removal of an aspect removes relational aspects of the stimulus. This latter logical definition, however, may encompass other types of relationships between stimulus dimensions other than integrality (at least as defined in the Garner

system). This possibility will be discussed in the next section when the concepts of configural dimensions and emergent features are introduced.

Configurality: Theoretical background. Garner has argued that some separable dimensions are optionally so (Garner, 1974). This implies that such dimensions can also be processed in a more unitary fashion. Such perceptual unity may be achieved if the two dimensions in combination produce a third dimension, particularly if this additional dimension is more salient than either of the original dimensions. This possibility has motivated much of the work of Pomerantz on dimensional configurality and emergent features (e.g., Pomerantz, 1981; Pomerantz & Schwaitsberg, 1975; Pomerantz & Garner, 1973).

As an example of configural dimensions, Pomerantz and Garner (1973) studied a parenthesis pair. A four-alternative stimulus set was formed by presenting each parenthesis opening to the right or to the left. Pairs of such parentheses clearly did not fit Garner's a priori, logical definition of integrality since the presence of both parentheses was not a requirement to determine whether one of them was left- or right-facing. However, the two parentheses did not seem truly independent either, since certain combinations seemed to form nominally distinctive stimuli. Thus, when each parenthesis opened inward on the other, an oval configuration was formed; when both parentheses opened away from each other, an hour-glass configuration was formed.

Pomerantz hypothesized that such interactions between dimensions should have special information processing consequences that distinguish them from either integral or separable dimensions. This idea was substantiated with the parenthesis pairs. Speeded classification tasks were not performed more quickly when there was redundancy between the two stimuli, thus pointing to potential separability. However, unlike separable dimensions, orthogonal variation in the nontarget parenthesis was associated with decrements in classification speed of the target parenthesis. The lack of redundancy gain was attributed to the notion that the parentheses (or configural parts) are dimensions only to the experimenter. To the subject, they do not function as such and thus the fact that the stimuli are physically correlated should have little effect. The filtering decrement, likewise, is said to arise not because the subject cannot exclude the irrelevant dimension from attention, but rather because each stimulus pair is processed or judged categorically. Thus, the task is not performed as a filtering task at all, but is instead treated by the subject as a grouping task where there are two possible stimuli mapped to each of the two possible responses.

Pomerantz (1981; Pomerantz & Pristach, 1987) has recently argued that the notion of emergent features may aptly describe the process responsible for configural effects in information processing tasks. Emergent features are defined as aspects of the novel perceptual wholes that result from configuration. These features, according to Garner (1981) and Pomerantz and Pristach (1987) are available in addition to the various parts or dimensions that make up the stimulus. That is, emergent features do not destroy any parts or make them less perceptible. Instead, subjects may opt to use emergent features for performing various tasks, when possible, due to their relative salience compared to the individual parts. Besides their

potential involvement in the configural effects manifested in speeded classification tasks, several authors have suggested a mediating role for emergent features in the object-superiority effect (Lanze, Maguire, & Weisstein, 1985; Pomerantz, Sager, & Stoever, 1977).

Definitions of configurality. As for integrality, performance measures have been widely used as a diagnostic for dimensional configurality. A task that has been used extensively for this purpose is the condensation task (Garner, 1981; Pomerantz & Pristach, 1987). In this variant of the speeded classification task, subjects are asked to make classifications that depend on their using both dimensions of the stimuli. With configural dimensions, this task is performed almost as easily as classifications based on only one dimension (when the irrelevant dimension is held constant). This is interpreted as indicating that the subject can divide attention over both dimensions as easily as he can attend to one dimension. However, Pomerantz (1981; Pomerantz & Pristach, 1987) has argued that this diagnostic for configural effects will only work so long as an emergent feature can be used to distinguish the dimensional pairs associated with one response from those of the other in the condensation task.

Other indications of configurality reviewed by Pomerantz and Pristach (1987) include the typical orthogonal interference with no redundancy gain (e.g., Pomerantz & Garner, 1973). In addition, there tend to be large differences in performance associated with different redundant pairings of dimensions, a finding less typical of integral stimuli. Treisman and Paterson (1984) have also added further performance diagnostics based on the feature integration theory of attention. They suggest that emergent features should behave like a separable feature in their paradigms, and thus should show parallel search, texture segregation, and should form illusory conjunctions with other features. However, emergent features should not result from the illusory conjunctions of other separable features.

Unlike dimensional integrality, no logical, a priori definitions exist for configurality. One might argue that the destruction of relational aspects with the removal of one dimension, as per Monahan and Lockhead's (1977) definition of integrality, might seem to fill the bill. However, the specification of what relational aspects are actually useful to the observer cannot always be made without consideration of the particular task required. That is, the relevance of relational aspects is task dependent. However, to the extent that a number of such relations exist between two dimensions, one might be able to argue that the dimensional set is more likely to provide for performance diagnostic of configurality.

Integrality, configurality, and the proximity compatibility hypothesis. As candidates to fill the role of "graphical proximity" in the proximity compatibility hypothesis, dimensional integrality and configurality must both be considered strong contenders. These concepts of dimensional unitariness have an intuitive appeal as proximity definitions because they both describe perceptual interactions of potential graphical elements whose effects are dependent on the task being performed. Table 4.2 shows the performance outcomes predicted by the proximity compatibility hypothesis, this time with integrality and configurality representing high display proximity, and with

separability representing low display proximity. In short, this table emphasizes the following predictions: for integration tasks, integral or configural graphical dimensions will be most compatible, while separable dimensions will be more compatible with the demands of focusing and multitask scenarios.

In addition to the presentation of the overall proximity compatibility hypothesis predictions, Table 4.2 also outlines some of the tentative reasons for expecting such compatibility effects. For example, under the heading of high task proximity, reasons for the expected compatibility effects with dimensional integrality and configurality are given, along with complementary explanations of the predicted incompatibility between high proximity tasks and separable dimensions. Jacob, Egeth and Bevan (1976) summarized this interaction when they suggest that integral dimensions may be useful in tasks requiring information integration because this mandatory perceptual integration of physical dimensions may replace the more effortful task of logically relating information from several different sources. Thus, a relatively quick, automatic perceptual process may be used to replace an attention-demanding, logical one. Similarly, with configural dimensions, direct processing of an emergent feature may sometimes be used to circumvent additional cognitive processing (Pomerantz, 1981; Carswell & Wickens, 1987a,b). Or, in the simplest case, a reduced number of functional perceptual discriminations may be required when integral or configural dimensions are used to perform the task. However, these potential perceptual shortcuts provided by both integrality and configurality will be absent when separable dimensions are used to present information that must be integrated. Thus, the incompatibility of separable information sources and integration demands results from the mandatory logical processing required to compare, contrast, or otherwise integrate the pertinent information.

The right half of Table 4.2 summarizes the potential relations of integrality, configurality, and separability of information sources in nonintegration tasks. Integrality and configurality may be incompatible with either focusing tasks or independent tasks for several reasons. For instance, if two dimensions are integral and information about the individual value of either is required, the subject may have to resort to additional (i.e., secondary) processing to "disintegrate" the perceptually united information sources. Both Garner (1974) and Lockhead and King (1977) have suggested additional stages of processing to account for the longer reaction times obtained when integral dimensions are used for focusing tasks. Pomerantz and Pristach (1987) have also suggested that additional processing may account for delayed reaction times when configural dimensions are used in focusing tasks. In this case, irrelevant emergent features may be more salient than the parts from which they are formed, thus resulting in initial misallocations of attention. The relative compatibility of separable dimensions for independent or focused attention tasks, then, lies in the ready correspondence of the physically manipulated dimensions to the functional information channels used for the multiple tasks; no secondary perceptual processing is required. Additionally, Dykes (1981) has suggested that separable dimensions are generally processed serially, thus the deleterious effects of intrusion and confusion errors in parallel processing may be attenuated.

Table 4.2

Dimensional Dependencies Used as Graphical Proximity Estimates
in the Proximity Compatibility Hypothesis

	High-Task Proximity	Low-Task Proximity
Display Integrality/ Configurality	COMPATIBLE • Automatic, perceptual integration of physical dimensions • Use of emergent features	INCOMPATIBLE • Requires secondary processing • Misallocation of attention
Display Separability	INCOMPATIBLE • No perceptual "Shortcuts" to avoid resource-demanding, logical operations	COMPATIBLE • Correspondence of physical dimensions to functional information channels

Research issues. The central question to be answered regarding dimensional unitariness is whether differences in graphical displays based on this measure are associated with the performance differences predicted by the proximity compatibility hypothesis. A graph showing hypothetically "ideal" data is presented in Figure 4.1. For two different tasks--one requiring integration and one requiring independent processing--the effect of unitariness is a monotonically increasing or decreasing function, respectively. To what degree do any of the available measures of dimensional dependencies approximate these ideal functions?

As a starting point, the abscissa in Figure 4.1 may be defined as a continuum with integral and configural graphs at one end, and more separable graphs at the other. Or, operationally, graphs may be ordered by the degree to which their dimensions produce orthogonal interference in speeded classification tasks. Since both integrality and configurality share this performance outcome, the more proximal or unitary displays may be either integral or configural. This composite description of proximity is consistent with the notion of integrality proposed by Monahan and Lockhead (1977). These authors have argued that both the dimensional syndromes of integrality and configurality may be the result of comparable similarity relations among stimuli in multidimensional psychological space. Thus, the term integrality is retained to denote both concepts. As a limiting case for such integrality, Lockhead (1966) has suggested that integral stimuli (i.e., integral or configural dimensions) must be both temporally and spatially proximal. In a similar vein, Garner (1976) has suggested that the degree to which dimensions are included in a single object as opposed to several different forms increases the likelihood of integral and configural relations. He further suggested that inclusion of dimensions in a single form might be sufficient for predicting stimulus-related differences in concept-learning and choice performance.

Alternatively, the differentiation of configurality and integrality effects on performance may be crucial. Thus, only those graphs with dimensions showing, for instance, redundancy gain in speeded classification might be associated with proximity advantages in integration tasks. That is, only stimulus integrality may act to produce display proximity advantages. Or, configurality rather than integrality might be necessary for such integration benefits, making the condensation performance diagnostic the more important measure of graphical proximity. The possibility that configurality rather than integrality among dimensions might be responsible for a number of performance outcomes with object displays has been noted by Wickens and Carswell (1987). However, most writers attribute object display effects to integrality (e.g., Jacob, et al., 1976; Goldsmith & Schvaneveldt, 1984).

A further issue in regard to specifying a dimensional dependence definition of proximity is whether a dependence continuum is necessary, or whether a discrete classification of dimensional relations will suffice. For instance, is it feasible to use Garner's (1970) a priori definition of integrality to describe the abscissa of Figure 4.1. From an applied standpoint, it would be preferable to be able to use a purely logical definition. This would circumvent the problem of having to derive performance measures such as those resulting from speeded classification to judge the

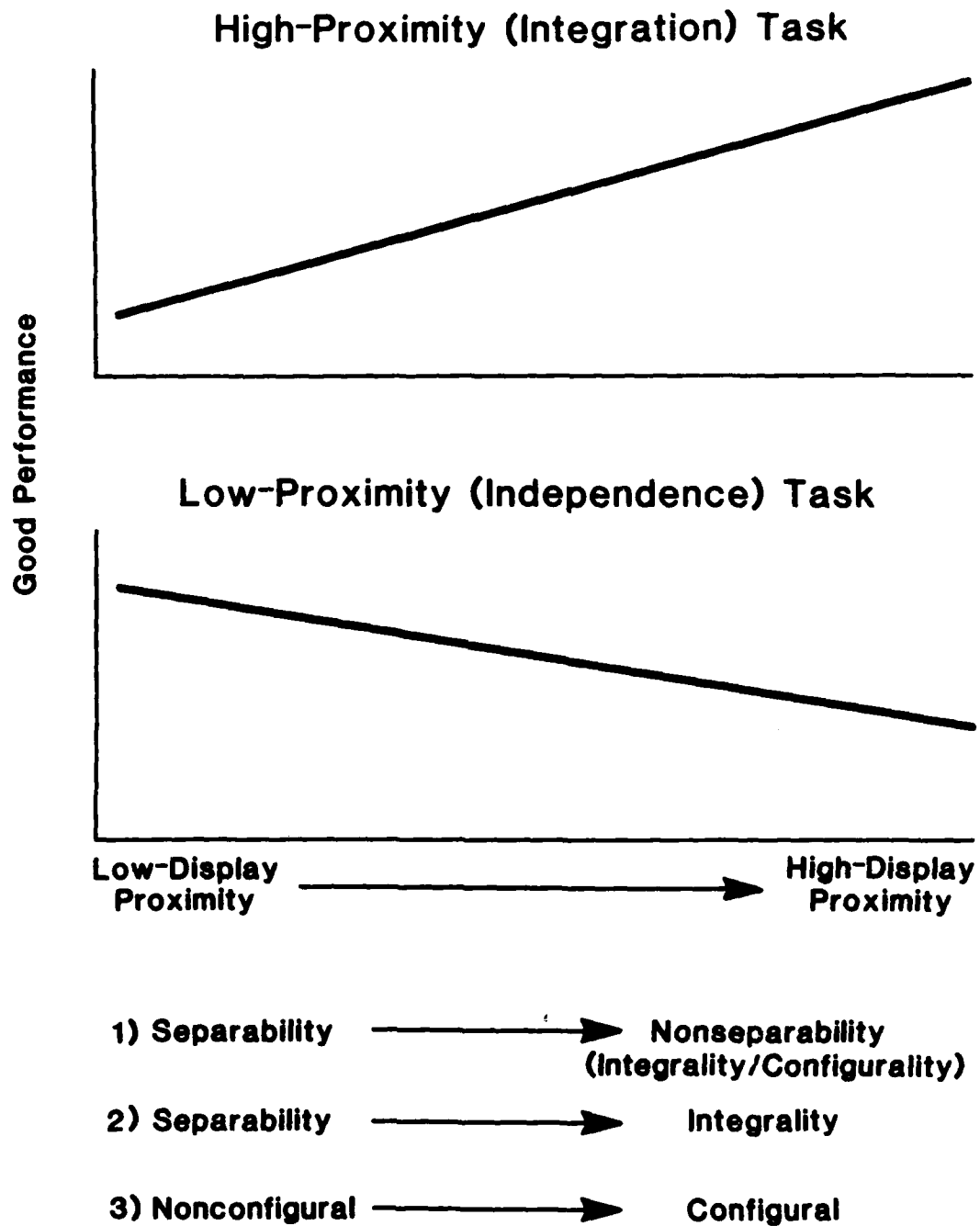


Figure 4.1. Hypothesized effect of display proximity on performance for an integration task and an independent processing task.

degree of dependence between each and every dimensional pair of potential interest to display designers. However, such a definition may not be adequate, and this can only be determined by comparing the two approaches and seeing how much information is gained by the use of the operationally derived continuum. The only previous comparative graphics study to use a continuum of integrality to predict performance was Jacob et al. (1976). However, their continuum was based primarily on the intuitions of the researchers, and their results did not show a strongly consistent ordering relating this integrality continuum to performance. The utility of a performance-based continuum to the problem of predicting graphical efficacy has yet to be studied.

Dimensional Homogeneity

The thrust of the research on integral versus separable dimensions, related above, deals with empirical attempts to derive what the basic, functionally independent units of perception might be. Thus, those experimentally manipulable dimensions that failed to meet the tests of true separability were said to constitute one class of graphical display elements that could certainly be considered more "proximal" since they seem not to be used independently by the visual system. Integrality and configurality, therefore, imply an extreme form of proximity.

However, other proximity metrics must surely exist that relate functionally independent perceptual dimensions to one another. For instance, if two information sources are needed for a particular display, the two dimensions chosen for the purpose can either require similar or different judgments. For example, both information sources may require a determination of height (as with two bars in a bar graph). Alternatively, the display could require one judgment of color and another of height. This distinction may constitute an additional definition of graphical proximity. To the degree that the same perceptual dimensions are used for multiple information sources, display proximity as determined by dimensional homogeneity exists.

Of course, other sorts of proximity may exist, even when the two dimensions used are the same. For instance, the two dimensions may use overlapping (proximal) or nonoverlapping (distant) sets of features for each of the two information sources. Additionally, the way that the two dimensions are displayed to the subject may form a sort of proximity-distance metric. The two homogeneous dimensions may be made less proximal, for instance, by increasing the physical distance between them, by displaying them at different orientations, or in different colors. The crucial aspect of this type of proximity manipulation is that it is a consistent value of the display (i.e., it transmits no information regarding the value to be extracted from either information source; rather, it is involved in labeling or identifying the source). Because the distinctions that serve to separate two or more homogeneous dimensions are usually different values or features along some additional dimension, these factors will be called feature homogeneity.

This section will describe in some detail the proximity compatibility hypothesis based on dimensional homogeneity. However, research appropriate to feature homogeneity will not be totally ignored. Since several of the feature homogeneity issues are intimately connected with perceptual grouping issues,

they will receive somewhat more detailed review in the discussions on object proximity.

Background. What is the evidence that two or more information sources can be processed efficiently when different dimensions are used to present the information for each task? Much of the theoretical impetus for these comparisons comes from the notion that functionally independent (separable) dimensions are served by separate "analyzers" (Treisman, 1969). To the degree that there is competition for the use of a particular analyzer (i.e., when multiple instances of the same dimension must be interpreted), there will be some interference in processing. However, to the extent that separate analyzers can be used (i.e., different dimensions are used to perform the task), then interference should be minimized.

Treisman's evidence comes mainly from the literature on auditory perception. However, some evidence from the visual perception literature exists. The work of Allport (1971; Wing & Allport, 1972) has focused on the ability of subjects to report two aspects of a briefly presented display. In one study (Allport, 1971) subjects were required to report selected information about sets of three-dimensional stimuli. Each stimulus was defined by an outline shape, an inscribed number, and a color. Compared to conditions when report of only one dimension was required, subjects were able to maintain performance when required to report both color and shape or color and a number. However, they were unable to maintain baseline performance when asked to report both shape and numbers. Allport suggested that this was because the overlap in dimensions used for numeral and shape identification. Thus, there was interference when two shape discriminations were required. To further explore the possibility that subjects could divide attention over different dimensions, Wing and Allport (1972) constructed stimuli out of spatial gratings that varied in size, orientation, and the orientation of a superimposed "break." As expected, subjects had difficulty reporting both the orientation of the break and the grating, but were able to report grating density and orientation without substantial performance decrements.

Definitions of dimensional homogeneity. The relation between heterogeneous dimensions and divided attention, as outlined above, is the major reason for the selection of dimensional homogeneity as a tentative descriptor of graphical proximity. To be able to use this concept in a test of the proximity compatibility hypothesis, however, definitions for such terms as "dimensions" and "features" must be derived. Treisman and Gelade (1980) have suggested the following distinctions. They use the term dimensions to refer to the complete range of variation that is separately analyzed by some functionally independent perceptual subsystem. Features are simply particular values along a dimension.

The way in which researchers have gone about determining what constitutes "functionally independent perceptual subsystems" has involved many different paradigms. For example, performance definitions, such as those discussed for separability of dimensions have been used (Garner, 1974; Treisman & Gelade, 1980). In addition, other performance measures such as those taken from adaptation studies have contributed to the search. Physiological studies involving single-unit recordings from the cortex of various animals have also

revealed that cells particularly responsive to some physically manipulable dimensions are organized into distinct retinotopic maps (e.g., Zeki, 1978). The findings from these various types of studies do not always converge (e.g., Houck & Hoffman, 1986), but for some tentative subsystems, such as color and orientation, the evidence for functional independence is perhaps stronger than for others.

Dimensional homogeneity and the proximity compatibility hypothesis. The proximity compatibility hypothesis would predict that multiple information sources represented by homogeneous dimensions should support performance in tasks requiring integration of various information sources. However, such homogeneity should harm performance in tasks requiring independent concurrent processing of the multiple sources, or focused attention on a subset of the sources. On the other hand, if the information is represented by heterogeneous dimensions, superior independent processing and focusing should result, while less efficient integration performance will be found. These predictions are presented in Table 4.3, along with some tentative reasons for expecting such findings.

The primary reason for the poorer performance suspected with use of homogeneous dimensions to display different information sources is the competition for resources that may occur within any particular analyzer (i.e., subsystem). This proposal was described earlier with examples from the work of Allport (1971; Wing & Allport, 1972).

However, if use of the same dimension results in degraded performance due to interference within a particular dimensional analyzer, then why should dimensional homogeneity ever be expected to facilitate performance when integration is required? One possibility arises from the work of Pomerantz (1981) and Prinzmetal (1981) on perceptual grouping. These authors relate evidence implicating similarity as a force in deciding what elements in the visual field will group. The majority of the studies reviewed by Pomerantz involve similarity among numerous elements in a large field. And, as Pomerantz notes, these findings relating element similarity to texture segregation may not generalize to results containing only two- or three-element displays, the types of displays that may presumably be more important in graphical presentations. One study that does show the effect of similarity on grouping in a two item display is presented by Garner (1981). In this experiment, subjects performed constrained classification of parentheses pairs, as well as with a parenthesis-bracket combination. Garner notes that the typical diagnostics of grouping (i.e., failure of selective attention and relatively successful condensation performance) that are apparent with the parentheses vanish when one parenthesis is replaced by a bracket.

If similarity between the dimensions being processed does, in fact, create a greater likelihood of perceptual grouping, then emergent features are also more likely to result. If such features are present, salient, and represent some useful combined value of the variables represented, then performance may be facilitated in integration tasks that demand the use of such combined values. Conversely, when different dimensions are used, emergent features may be much less likely to result, and thus performance in integration tasks is likely to be inefficient compared to the case when homogeneous dimensions are

Table 4.3

Dimensional Homogeneity Used as a Measure of Display Proximity
in the Proximity Compatibility Hypothesis

	High-Task Proximity	Low-Task Proximity
Homogeneous Displays	COMPATIBLE • Increased chance of perceptual grouping and emergent features	INCOMPATIBLE • Competition for analyzers
Heterogeneous Displays	INCOMPATIBLE • Emergent features may lead to misallocation of attention • Make comparisons difficult	COMPATIBLE • Use separate analyzers and avoids competition

used. In addition, Cleveland and McGill (1984) argue that the heterogeneous display used in an integration task may be at an additional disadvantage since these tasks often require comparisons. Comparisons of magnitude on different dimensions (e.g., is the brightness relatively more bright than the height is tall?) may prove to be particularly difficult tasks.

Research issues. Once again, the main issue to be tackled experimentally is whether or not this measure of graphical proximity--dimensional homogeneity--is useful for predicting graphical efficacy in independent and integration tasks. However, there are several more specific questions that need to be addressed, particularly if dimensional homogeneity does appear to be an important variable for graphic design.

The first of these issues involves the generality of homogeneity benefits for integration (and heterogeneity benefits for independent tasks) over different kinds of dimensional combinations. For example, do combinations of form dimensions such as linear extent and orientation give rise to results similar to combinations of form and nonform dimensions? In short, to what degree does the nature of the dimensions in the heterogeneous (nonproximal) displays alter the predictions of the proximity compatibility hypothesis?

A further issue is whether the use of dimensional homogeneity as a proximity metric is equally applicable to different types of integration tasks. This issue, in particular, is related to both the claims of Cleveland and McGill (1984) and to the commonsense idea that judging the relative magnitudes of two or more variables strongly requires displays constructed with homogeneous dimensions. Thus, comparative integration tasks may be particularly susceptible to this form of proximity manipulation. Other types of integration tasks, such as conjunctive tasks requiring a detection response when specific levels of each of several variables are present, may be less susceptible to homogeneity manipulations.

Finally, it may be interesting to determine whether dimensional homogeneity is equally applicable to cases where the relevant dimensions are displayed in separate rather than the same perceptual object. This issue will serve to introduce the last type of graphic proximity to be considered--object proximity.

Object Proximity

The final candidate for a descriptor of graphical proximity is one that may encompass some of the previously described measures of proximity, although in an imprecise way. In particular, a high degree of proximity may be assumed for variables displayed as dimensions of a single, unitary object. Low proximity, on the other hand, would be ascribed to dimensions that are parts of different perceptual objects. This category of proximity pits object displays such as faces, polygons, glyphs, and trees against multiobject displays such as bar graphs, dot charts, most pictographs, and banks of separate meters.

Theoretical background. Recent interest in the object concept in information processing has been the result of the work of Kahneman and

colleagues on the "object file" model of attention (e.g., Kahneman & Treisman, 1984; Kahneman & Henik, 1981). Kahneman argues that we are limited in our ability to divide attention between separate objects. However, divided attention among the parts of an object is relatively effective. That is, allocation of attention to an object facilitates the processing of all of its properties.

One implication of the object file model of attention is that information processing of multiple dimensions should be facilitated whenever they are incorporated into a single object. That is, there should be a general benefit to performance of both independent dual tasks and integration tasks when relevant information is presented in an object display. Evidence for such benefits has come from several sources. Perhaps one of the earliest examples of an object display benefit came from Lappin (1967). He found that subjects could more accurately report three different attributes of a single briefly presented object than they could either the same dimension or different dimensions of three separate objects. Treisman, Kahneman, and Burkell (1983) also found that the detection rate for both words and position of a gap in a rectangle was higher when the rectangle surrounded the word than when it was beside the word. This was found to be the case even though the distance from the gap to the word was the same in both conditions. Treisman et al. (1983) suggest that when the rectangle surrounds the word, the display is seen as a single object. When the rectangle is to one side of the word, two objects are seen. Thus, the poorer performance obtained with the latter case results from the increased difficulty of dividing attention over two objects rather than one.

Duncan (1984) reported similar results for reports of briefly presented displays. In order to control for spatial differences that tend to confound single and multiple object displays, Duncan created displays where one object was superimposed on another. He asked subjects to report a dimension from each of the two objects or two dimensions from only one of the two. His data support the notion that reporting dimensions from the same object can be made more efficiently. Once again, the difficulty of dividing attention over two objects was supported.

Using a slightly different approach, Kramer, Wickens, and Donchin (1985) found that allocation of attention to one task leads to increases of resources invested in a secondary task, a concurrence benefit. This benefit was obtained, however, only when the stimuli for the two tasks were presented in a single object. When attention was divided between objects, there was a cost of concurrence, where attentional resources dedicated to one task are reduced and allocated elsewhere when the need arises.

Definitions of object proximity. So far, the discussion of object proximity has avoided the issue of how perceptual objects may be defined. There has been some controversy, for instance, over whether a printed word is an object in the same sense that a geometric figure may be considered an object (Duncan, 1985) or whether surrounding a word by a contour defines both as part of the same object. Duncan suggests two alternative views for consolidating results using these two different types of configurations. First, he suggests that a simple continuum may exist from configurations that

may be grouped less strongly (such as letters of a word) to those that are grouped very strongly (such as the height and color of a triangle). Alternatively, he suggests, one can assume an hierarchical organization of visual information. Within this hierarchy, there may be various levels of objects--that is, more global and more local patterns.

Pomerantz (1981, 1983) has suggested two different types of configurations--P configurations and N configurations. The first of these are placeholders, configurations such as the stars that form a constellation, or the hierarchical letter stimuli used for many part-whole experiments (e.g., a large H formed out of the arrangement of smaller "X's"). In this class of configurations, the form of the individual elements is not essential to the form of the more global stimulus; only their relative placement is crucial. On the other hand, there may be configurations that are determined by the actual formal properties of the parts and their relationships such as the angles between various lines that form a triangle. These are N configurations. Pomerantz warns that different types of attentional effects may be obtained by using these various types of configurations.

Other attempts to define objects involve the degree to which certain stimuli possess features assumed to be more common in perceptual objects. An example of such a fuzzy definition is that given by Wickens (1984, Wickens & Carswell, 1987). He suggests that there are several properties that tend to describe most perceptual objects, but that none of these entirely defines such a concept. Among these properties are presence of contours, spatial proximity, and correlation of attributes. Another notion related to the objectness of a stimulus is Garner's (1974) concept of pattern goodness. Some patterns--those that yield few alternatives when rotated or reflected about various axes--may be said to have various processing advantages. Maybe these are also more essentially object-like.

All of these conceptions of objects (or configurations or patterns) lead to a notion of degrees of objectness. In the studies cited regarding the object file notion, definitions of objects versus multiobject stimuli were made predominantly in an either/or fashion. When subjects were themselves asked to rate the objectness of stimuli (e.g., Duncan, 1984), they were evidently given the option of a stimulus either being one object or two objects. The notion of one display being more subjectively object-like than another, the notion of subjective degree-of-objectness, has not been rigorously studied with regard to the object file notion. Thus, the present use of the term object, will imply a continuum, and agreement among display users will constitute a measure of this degree of objectness. It may be that these subjective measures reflect some combination of stimulus properties that can be measured via performance or physical observations--such as degree of integrality, configurality, and spatial proximity. Since the notion of objectness has depended heavily on subjective estimates of perceptual unity, the systematic study of these judgments in a given stimulus set should be undertaken. Thus, the present experiment will study the utility of subjective estimates of degree of objectness in predicting performance in integration versus nonintegration tasks. The question of what such subjective definitions

contribute to performance prediction, beyond what can be predicted by dimensional homogeneity and dimensional dependence will thus be critically evaluated.

Objects and the proximity compatibility hypothesis. Table 4.4 presents an overview of the proximity compatibility hypothesis when objectness is used as the measure of proximity compatibility. In brief, when variables are displayed within a single object (a high proximity display), integration performance (in high proximity tasks) should benefit, and independent, multitask and focusing performance should suffer. However, relative to the object display, multiobject displays should aid independent multitask performance and should detract from integration performance. Thus, the high compatibility conditions are those using object displays for integration tasks and those using multiobject displays for independent processing and focused attention tasks.

Garner (1976) used the object/nonobject distinction to predict performance in both concept learning and choice tasks. He argued that inclusion in a single perceptual object was probably a necessary, though not sufficient condition, for integrality of two or more dimensions. Further, dimensions of two separate objects were almost certainly separable. Similarly, Lockhead (1966) has argued that integrality depends on multiple dimensions coexisting at the same place and time--a requirement satisfied by most elements of a single object. Thus, to the extent that the dimensions of a single object are more likely to be integral than are those of different objects, the benefits likely to accrue with use of integral dimension to display information are more likely to be found in object displays than in nonobject displays. Thus, object displays should be better in integration tasks than should multi-object displays. However, when such integration is not desired, the additional processing that may be required to analyze each of two integral dimensions (Garner, 1970, 1974) is likely to reduce performance efficiency. Object displays are, therefore, less well-suited for nonintegration tasks.

A related benefit of object displays may be their greater tendency to yield emergent features. That is, objects involve not only parts, but relations amongst parts that may be directly perceived (Pomerantz, 1981; Pomerantz & Pristach, 1987). To the extent that such relations are directly perceived, and to the extent that they are directly related to task-relevant responses, object displays containing such features should benefit performance. Since the type of response or decision that might require use of such emergent features is one that requires recognition of relations amongst variables, this means that the use of emergent properties is especially suited to integration tasks. Thus, once again, the object display may be more suited for integration tasks than are nonobject displays. However, when the individual values of the various dimensions are required, as in focusing are independent processing tasks, then object displays containing especially salient but irrelevant emergent features may promote inefficient attention allocation or filtering decrements.

Kahneman's object file model of attention seems to be somewhat at odds with the predictions of the proximity compatibility hypothesis, particularly the outcomes predicted for the upper right cell in Table 4.4. The object file

Table 4.4

Object Proximity Used as a Measure of Display Proximity
in the Proximity Compatibility Hypothesis

	High-Task Proximity	Low-Task Proximity
Object Displays	COMPATIBLE • Benefits of integrality/configurality • Added benefits of "Object"-induced parallel processing	INCOMPATIBLE • Disadvantage of integrality/configurality • Response conflict
Multiobject Displays	INCOMPATIBLE • Forces logical integration of information	COMPATIBLE • Reduces response conflict

model implies that there is a general benefit to using the attributes of a single object for any task involving multiple variables. The proximity compatibility hypothesis, alternatively, suggests that this may be true with integration tasks, but that object displays may be more likely to cause problems when independent tasks are concerned. The rationale for the relatively greater advantage of the object display for integration as opposed to nonintegration tasks involves the increased probability of response conflict occurring in the latter case. When multiple information sources must be integrated to produce a single response, the chance for response competition is rather slim. However, when multiple responses are necessary for different inputs, or when some inputs must be ignored, the possibility of the wrong input actuating the response becomes more likely.

Perhaps the prototypical example of response competition in a focused attention task is the Stroop phenomenon (Stroop, 1935; Dyer, 1973). Subjects are asked to name the color ink in which each of a list of words is written. When the words refer to colors that are inconsistent with the actual ink color in which they are written, color naming becomes quite effortful. Subjects usually respond correctly, but after some delay compared to naming the color of neutral words. Thus, as proposed by the object file model, both the relevant ink color and the irrelevant semantic content of the word are processed; however, performance is disrupted because two conflicting responses are associated with the two stimulus attributes. Kahneman and Henik (1981) tested the proposition that the Stroop effect would be diluted if the irrelevant color name appeared in another word (object) rather than in the word whose ink-color was to be identified. Their results indicated a dramatic decrease in the amount of response conflict with such stimuli. Thus, for a nonintegration (focusing) task, performance was best when conflicting irrelevant attributes were displayed in a multiobject display. The proximity compatibility hypothesis would, however, predict the reverse finding if an integration task were required of Stroop stimuli. For example, if the task were to respond "yes" when the ink was in the color specified by the color name, and to respond "no" when the two attributes were in conflict, then proximity compatibility would predict that the single object condition would be associated with superior performance compared to the matching of one word and a separate colored object.

A line of research relevant to both the object file concept and the proximity compatibility hypothesis is that dealing with the effect of spatial proximity on performance in focusing and divided attention tasks (e.g., Eriksen & Hoffman, 1973; Eriksen & Yeh, 1985). The general premise of these studies is that attention acts as a spotlight or, more precisely, as a zoom lens. Thus, attention allows processing of units within its spatial focus, with a narrower focus resulting in a concentration of attentional resources on a limited spatial location and with wider focus resulting in distribution of resources over many units in the field. To the extent that several aspects of a single object are more likely to be in greater spatial proximity than are attributes of several different objects, then the zoom lens model may be used somewhat interchangeably for objectness in the proximity compatibility framework. That is, when two or more dimensions are spatially proximal (in the same object) they are likely to promote integration task performance and may produce response competition in focused attention or independent

multitasks. Thus, performance should be increased in focused attention tasks to the extent that two displays are separated spatially, and this is more likely in the case where two different objects are used. Eriksen and Hoffman (1973), for instance, have demonstrated that interference from an irrelevant letter was greatest in a focusing task when that letter was more physically proximal than when it was further removed. Presumably, this is due to the increased probability that the focus of attention can capture only the relevant target letter when there is more distance between the relevant and irrelevant material.

Although the probability of dimensional interaction (dimensional proximity) is greater within than between objects, as is spatial proximity, dimensional homogeneity is not so clearly related to the object versus nonobject description of proximity. In the previous section, proximity in terms of homogeneity of stimulus dimensions was proposed. Integration performance should be superior when two judgments are required for two examples of the same dimension, and multi-task and focused attention performance should be relatively better under conditions of dimensional heterogeneity. This distinction runs more nearly orthogonal to the object/nonobject distinction than do integrality or configurality. Thus, it is possible to choose heterogeneous dimensions that are either contained in a single object or are divided between objects to display relevant information. Will the effects of dimensional heterogeneity be independent of those of objectness, or might they interact in some way? In short, are these two descriptions of proximity--object proximity and dimensional homogeneity--actually additive?

Research issues. Given that the notion of objectness includes many of the previously discussed measures of proximity, such as integrality and configurality, the main issue to be addressed is whether any additional information can be obtained by using objectness as the basis for categorizing graphical displays. On the other hand, are such measures as dimensional heterogeneity, integrality, or configurality sufficient to predict graphical efficacy?

In addition, is the notion of a subjective continuum of "objectness" a useful one for analyzing graphical forms? Prior experiments have used the object/nonobject dichotomy, but is this a sufficient way of describing one's impressions of the cohesiveness of stimulus attributes? Further, it will be interesting to determine whether logical definitions of objectness, such as the presence of contours and spatial proximity, are a useful heuristic for predicting subjective perceptions of objectness, as well as for predicting performance.

CHAPTER 5

SUMMARY OF APPLIED AND EXPERIMENTAL ISSUES IN GRAPHIC DESIGN

Based on the historical analysis presented in Chapter 1, we have argued that there is growing demand for good graphical representations of quantitative information. This need has been created largely by the widespread use of computers and has been reflected in what might be called a "graphical renaissance." This rebirth of interest in developing novel graphical formats may be seen in statistics, industrial process control, medicine, and business, as well as in aviation. In each of these areas of application, there is the growing recognition that the amount of information available to operators is quickly reaching unmanageable proportions. One potential way to alleviate some of this burden may be through the use of well-constructed graphics.

In the second chapter, we reviewed previous attempts to determine what constitutes a well-constructed graph. Studies that compared the ability of subjects to use different formats--comparative graphics--were discussed. The cumulative findings of such research indicate that there is no single best graphic format. Instead, graphical efficacy seems to be very task dependent. Therefore, any comprehensive psychological model that seeks to predict graphical efficacy must focus on the nature of the interaction between task and display characteristics.

The proximity-compatibility hypothesis was described in Chapter 3 as one potential framework for studying graphical alternatives. According to this model, the operator's ability to use a display will be maximized to the extent that "proximal" tasks are matched with "proximal" displays. A proximal task was described as one that involves the mapping of information from several channels onto fewer responses. Less proximal tasks would include multiple-task situations in which there is independence in the utility of information from different sources. The research in comparative graphics seemed to agree with this model and, in addition, experiments designed as direct tests of its predictions have been fairly successful. However, the proximity compatibility hypothesis, as presently applied to graphic design, is mainly a qualitative, heuristic model. Future research should be aimed at determining the degree to which the model can be used to make more precise, quantitative predictions regarding graphical efficacy in different task environments. A number of experimental issues are associated with this aim:

1. How should graphical proximity be defined and measured? The fourth chapter proposed three candidate definitions of graphical proximity--dimensional dependence, dimensional homogeneity, and objectness. Basic research implicating these three proximity measures was reviewed.
2. When graphs varying on the three candidate measures of proximity are used to perform varying types of tasks, to what degree is the proximity compatibility hypothesis supported?

3. Is any one of the proximity definitions sufficient to predict graphical efficacy, or is some composite of these measures required?
4. In addition, what is the form of the relationship between proximity measures and display efficacy for any given task? Is it necessary to view proximity as a continuum for graphic design purposes? Or, may categorical definitions of proximity suffice?
5. In addition to the three proposed proximity measures, what other descriptors of graphic displays might be useful in determining overall graphical efficacy? How does graphical proximity compare to such factors as the "data-ink ratio" or the "basic graphical elements" chosen for a particular design?
6. If task-display interactions are obtained, but the proposed measures of proximity are not adequate for predicting the form of these interactions, what alternative display variables may be responsible?

In an ongoing program of research, many of these issues are being addressed using, as a starting point, bivariate graphs. This program consists of two phases. In a preliminary descriptive phase, each of a number of candidate graphs was tested, using the traditional speeded classification paradigm, for evidence of dimensional dependencies. This phase provides an important set of behavioral scaling data for one type of proximity. Multidimensional scaling of subjective objectness is also an important part of this phase. Phase 2 of this experimental program involves use of the graphs selected in phase 1 to perform experimental tasks representative of different levels of task proximity. This phase represents a critical test of the proximity compatibility hypothesis and will provide data to determine which types of display proximity may be most relevant to display design.

REFERENCES

- Allport, D.A. (1971). Parallel encoding within and between elementary stimulus dimensions. Perception and Psychophysics, 10, 104-108.
- American Statistical Association. (1916). Joint committee on standards for graphic presentation. Publications of ASA, 14, 790-797.
- Anderson, E. (1957). A semigraphical method for the analysis of complex problems. Proceedings of the National Academy of Science, 13, 923-927.
- Andrews, D.F. (1972). Plots of high dimensional data. Biometrics, 28, 125-135.
- Barnett, B.J., & Wickens, C.D. (1988). Display proximity in multicue information integration: The benefits of boxes. Human Factors, 30, 15-24.
- Barnett, V. (1981). Interpreting multivariate data. Chichester: John Wiley and Sons.
- Beniger, J.R., & Robyn, D.L. (1978). Quantitative graphics in statistics: A brief history. The American Statistician, 32, 1-11.
- Bertin, J. (1973). La semiologie graphique. The Hague, Netherlands: Mouton-Gautier.
- Boles, D.B., & Wickens, C.D. (1983). A comparison of homogeneous and heterogeneous display formats in information integration and nonintegration tasks (Technical Report EPL-83-6/ONR-83-6). Urbana, IL: Engineering-Psychology Research Laboratory, University of Illinois at Urbana-Champaign.
- Brinton, W.C. (1916). Graphic methods of presenting facts. New York: Engineering Magazine.
- Brown, R.L. (1985). Methods for the graphic representation of systems simulated data. Ergonomics, 28, 1439-1454.
- Carswell, C.M., & Wickens, C.D. (1987a). Information integration and the object display: An interaction of task demands and display superiority. Ergonomics, 30, 511-528.
- Carswell, C.M., & Wickens, C.D. (1987b). Why not the object display? Proceedings of the 3rd Mid-Central Human Factors/Ergonomics Conference, Oxford, OH.
- Casey, E.J., & Wickens, C.D. (1986). Visual display representation of multidimensional systems (Technical Report CPL-86-2/MDA903-83-K-0255). Champaign, IL: Cognitive Psychophysiology Laboratory, University of Illinois at Urbana-Champaign.

- Chambers, J.M., Cleveland, W.S., Kleiner, B., & Tukey, P.A. (1983). Graphical methods for data analysis. Blemant, CA: Wadsworth.
- Cheng, P.W., & Pachella, R.G. (1984). A psychophysical approach to dimensional separability. Cognitive Psychology, 16, 279-304.
- Chernoff, H. (1973). The use of faces to represent points in K-dimensional space graphically. Journal of the American Statistical Association, 68, 361-368.
- Chernoff, H. (1978). Graphical representation as a discipline. In P.C.C. Wang (Ed.) Graphical representation of multivariate data. New York: Academic Press, 1-12.
- Cleveland, W.S. (1985). The elements of graphing data. Monterey, CA: Wadsworth.
- Cleveland, W.S., Harris, C.S., & McGill, R. (1983). Experiments on quantitative judgments of graphs and maps. The Bell System Technical Journal, 62, 1659-1674.
- Cleveland, W.S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphic methods. Journal of the American Statistical Association, 70, 531-554.
- Cleveland, W.S., & McGill, R. (1985). Graphical perception and graphical methods for analyzing scientific data. Science, 229, 828-833.
- Cleveland, W.S., & McGill, R. (1986). An experiment in graphical perception. International Journal of Man-Machine Studies, 25, 491-500.
- Cole, W.G. (1986). Medical cognitive graphics. Proceedings of the ACM, SIGCHI '86.
- Cox, D.R. (1978). Some remarks on the role in statistics of graphical methods. Applied Statistics, 27, 4-9.
- Croxton, F.E. (1927). Further studies in the graphic use of circles and bars: Some additional data. Journal of the American Statistical Association, 22, 36-39.
- Croxton, F.E., & Stein, H. (1932). Graphic comparisons by bars, squares, circles, and cubes. Journal of the American Statistical Association, 27, 54-60.
- Croxton, F.E., & Stryker, R.E. (1927). Bar charts versus circle diagrams, Journal of the American Statistical Association, 22, 475-482.
- Culbertson, H.M., & Powers, R.D. (1959). A study of graph comprehension difficulties. AV Communication Review, 7, 97-110.

- DeSanctis, G. (1984). Computer graphics as decision aids: Directions for research. Decision Sciences, 15, 463-487.
- Duncan, J. (1984). Selective attention and the organization of visual information. Journal of Experimental Psychology: General, 113, 501-517.
- Dunn, J.C. (1983). Spatial metrics of integral and separable dimensions. Journal of Experimental Psychology: Human Perception and Performance, 9, 242-257.
- Dyer, F.N. (1973). The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. Memory & Cognition, 2, 106-120.
- Dykes, J.R., Jr., & Cooper, R.G., Jr. (1978). An investigation of the perceptual basis of redundancy gain and orthogonal interference for integral dimensions. Perception and Psychophysics, 23, 36-42.
- Dykes, J.R., Jr. (1979). A demonstration of selection of analyzers for integral dimensions. Journal of Experimental Psychology: Human Perception and Performance, 4, 734-745.
- Dykes, J.R., Jr. (1981). Perceptual encoding and decision strategies for integral dimensions. Journal of Experimental Psychology: Human Perception and Performance, 7, 56-70.
- Eells, W.C. (1926). The relative merits of circles and bars for representing component parts. Journal of the American Statistical Association, 21, 119-132.
- Eriksen, C.W., & Hoffman, J.E. (1973). The extent of processing of noise elements during selective encoding from visual displays. Perception and Psychophysics, 14, 155-160.
- Eriksen, C.W., & Yeh, Y.Y. (1985). Allocation of attention in the visual field. Journal of Experimental Psychology: Human Perception and Performance, 11, 583-597.
- Everitt, B. (1978). Graphical techniques for multivariate data. New York: North-Holland.
- Feinberg, S.E. (1979). Graphical methods in statistics. The American Statistician, 33, 165-178.
- Fitts, P.M., & Seeger, C.M. (1953). S-R compatibility: spatial characteristics of stimulus and response codes. Journal of Experimental Psychology, 46, 199-210.
- Fowkes, M. (1984). Presenting information to the driver. Displays, 9, 215-223.
- Funkhouser, H.G. (1937). Historical development of the graphical representation of statistical data. Osiris, 3, 269-404.

- Garner, W.R. (1970). The stimulus in information processing. American Psychologist, 25, 350-358.
- Garner, W.R. (1974). The processing of information and structure. Hillsdale, NJ: Erlbaum Associates.
- Garner, W.R. (1976). Interaction of stimulus dimensions in concept and choice processes. Cognitive Psychology, 8, 98-123.
- Garner, W.R. (1981). The analysis of unanalyzed perceptions. In M. Kubovy & J.R. Pomerantz (Eds.), Perceptual organization (pp. 119-139). Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Garner, W.R., & Felfoldy, G.L. (1970). Integrality of stimulus dimensions in various types of information processing. Cognitive Psychology, 1, 225-241.
- Garner, W.R., & Flowers, J.H. (1969). The effect of redundant stimulus elements on visual discrimination as a function of element heterogeneity, equal discriminability, and position uncertainty. Perception and Psychophysics, 6, 216-220.
- Garner, W.R., & Morton, J. (1969). Perceptual independence: Definitions, models, and experimental paradigms. Psychological Bulletin, 72, 233-259.
- Goettl, B.P., Kramer, A.F., & Wickens, C.D. (1986). Display format and the perception of numerical data. Proceedings of the 30th Annual Human Factors Society Meeting (pp. 450-454). Santa Monica, CA: Human Factors Society.
- Goldsmith, T.E., & Schvaneveldt, R.W. (1984). Facilitating multi-cue judgments with integral information displays. In J. Thomas & M. Schneider (Eds.), Human factors in computer systems (pp. 243-270). Norwood, NJ: Ablex.
- Goodstein, L.P. (1981). Discriminative display support for process operators. In J. Rasmussen & W.B. Rouse (Eds.), Human detection and diagnosis in system failures. New York: Plenum Press.
- Gottwald, R.L., & Garner, W.R. (1975). Filtering and condensation tasks with integral and separable dimensions. Perception and Psychophysics, 18, 26-28.
- Greenwald, A.G. (1979). Time-sharing, idiomotor compatibility and automaticity. In C. Bense (Ed.), Proceedings of the 23rd Annual Human Factors Society Meeting. Santa Monica, CA: Human Factors Society.
- Handel, S., & Imai, S. (1972). The free classification of analyzable and unanalyzable stimuli. Perception and Psychophysics, 12, 108-115.
- Harris, D. (1984). Computer graphics and applications. London: Chapman and Hall.

- Harwood, K., Wickens, C., Kramer, A., Clay, D., & Liu, Y. (1986). Effects of display proximity and memory demands on the understanding of dynamic multidimensional information. Proceedings of the 30th Annual Human Factors Society Meeting. Santa Monica, CA: Human Factors Society.
- Hoff, H.E., & Geddes, L.A. (1959). Graphic recording before Carl Ludwig: An historical summary. Archives Internationales d'Histoire des Sciences, 12, 3-25.
- Houck, M.R., & Hoffman, J.E. (1986). Conjunction of color and form without attention: Evidence from an orientation-contingent color aftereffect. Journal of Experimental Psychology: Human Perception and Performance 12, 186-199.
- Hyman, R., & Well, A. (1967). Judgments of similarity and spatial models. Perception and Psychophysics, 2, 233-248.
- Hyman, R., & Well, A. (1968). Perceptual separability and spatial models. Perception and Psychophysics, 3, 161-165.
- Jacob, R.J.K., Egeth, H.E., & Bevan, W. (1976). The face as a data display. Human Factors, 18, 189-200.
- Jones, P. (1986). The display of multivariate information: the effects of auto- and cross-correlation, reliability, and heterogeneity. Unpublished manuscript, University of Illinois at Urbana-Champaign.
- Jones, P. M., & Wickens, C.D. (in press). The display of multivariate information: The effects of auto- and cross-correlation, reliability, and heterogeneity. Human Performance.
- Kahneman, D., & Chajczyk, D. (1983). Tests of the automaticity of reading: Dilution of stroop effects by color-irrelevant stimuli. Journal of Experimental Psychology: Human Perception and Performance, 9, 497-501.
- Kahneman, D., & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J.R. Pomerantz (Eds.), Perceptual organization (pp. 181-209). Hillsdale, NJ: Erlbaum Associates.
- Kahneman, D., & Treisman, A.M. (1984). Changing views of attention and automaticity. In R. Parasuraman, R. Davies, & J. Beatty (Eds.), Varieties of attention. New York: Academic Press.
- Kramer, A.F., Wickens, C.D., & Donchin, E. (1985). Processing of stimulus properties: Evidence for dual-task integrality. Journal of Experimental Psychology: Human Perception and Performance 11, 393-408.
- Kruskal, W. (1977). Visions of maps and graphs. Proceedings of the International Symposium on Computer-generated Cartography, Auto-Carlo II, 27-36.

- Lanze, M., Maguire, W., & Weisstein, N. (1985). Emergent features: A new factor in the object-superiority effect? Perception and Psychophysics, 38, 438-442.
- Lappin, J.S. (1967). Attention in the identification of stimuli in complex displays. Journal of Experimental Psychology, 75, 321-328.
- Lockhead, G.R. (1966). Effects of dimensional redundancy on visual discrimination. Journal of Experimental Psychology, 72, 93-104.
- Lockhead, G.R., & King, M.C. (1977). Classifying integral stimuli. Journal of Experimental Psychology: Human Perception and Performance, 3, 436-443.
- Lorch, E.P., Anderson, D.R., & Well, A.D. (1984). Effects of irrelevant information on speeded classification tasks: Interference is reduced by habituation. Journal of Experimental Psychology: Human Perception and Performance, 10, 850-864.
- MacDonald-Ross, M. (1977). How numbers are shown: A review of research on the presentation of quantitative data in texts. AV Communication Review, 25, 359-409.
- Mezzich, J.E., & Worthington, D.R.L. (1978). A comparison of graphical representations of multidimensional psychiatric diagnosis data. In P.C.C. Wang (Ed.), Graphical representation of multivariate data (pp. 123-142). New York: Academic Press.
- Monahan, J.S., & Lockhead, G.R. (1977). Identification of integral stimuli. Journal of Experimental Psychology: General, 106, 94-110.
- Moriarty, S. (1979). Communicating financial information through multidimensional graphics. Journal of Accounting Research, 17, 205-224.
- Naveh-Benjamin M., & Pachella, R.G. (1982). The effect of complexity on interpreting Chernoff faces. Human Factors, 24, 11-18.
- Neurath, O. (1944). Visual aids and arguing. New Era, 25, 51-61.
- Pachella, R.G., & Somers, P. (1978). The development of integrated multidimensional displays. Naval Research Review, 31, 19-32.
- Pachella, R.G., Somers, P., & Hardzinski, M. (1981). A psychophysical approach to dimensional integrality. In D. Getty & J. Howard (Eds.), Visual and auditory pattern recognition. Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.
- Petersen, R.J., Banks, W.W., & Gertman, D.I. (1981). Performance-based evaluation of graphic displays for nuclear power plant control rooms. In Proceedings of the Conference on Human Factors in Computing Systems (pp. 182-189). Gaithersburg, MD: ACM.

- Playfair, W. (1786). Commercial and political atlas (1st ed.). London: Corry.
- Playfair, W. (1801). Commercial and political atlas (3rd ed.). London: Stockdale.
- Polson, M., Wickens, C., Colle, H., & Klapp, H. (1986). Interactive processes: Task interference, independence and integration. In M. Chagnall & P. Hancock (Eds.), Adaptive human-machine interfaces. New York: Wiley.
- Pomerantz, J.R., & Garner, W.R. (1973). Stimulus configuration in selective attention tasks. Perception and Psychophysics, 14, 565-569.
- Pomerantz, J.R., & Pristach, E.A. (1987). Emergent features, attention, and perceptual glue. Manuscript submitted for publication.
- Pomerantz, J.R., & Schweitzerberg, S.D. (1975). Grouping by proximity: Selective attention measures. Perception and Psychophysics, 18, 355-361.
- Pomerantz, J.R., Sager, L.C., & Stoever, R.J. (1977). Perception of wholes and of their component parts: Some configural superiority effects. Journal of Experimental Psychology: Human Perception and Performance, 3, 422-435.
- Pomerantz, J.R. (1981). Perceptual organization in information processing. In M. Kubovy & J.R. Pomerantz (Eds.), Perceptual organization (pp. 141-180). Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Pomerantz, J.R. (1983). Global and local precedence: Selective attention in form and motion perception. Journal of Experimental Psychology: General, 112, 516-540.
- Prinzmetal, W. (1981). Principles of feature integration in visual perception. Perception and Psychophysics, 30, 330-340.
- Rasmussen, J., & Jensen, A. (1974). Mental procedures in real life tasks: A case study of electronic trouble-shooting. Ergonomics, 17, 193-207.
- Roscoe, S.N. (1968). Airborne displays for flight and navigation. Human Factors, 10, 321-332.
- Roscoe, S.N. (1981). Aviation psychology. Iowa City: University of Iowa Press.
- Sapita, R.F. (1982). The interface: Man-machine or man-process? Process Automation, 75-84.
- Schmid, C.F., & Schmid, S.F. (1979). Handbook of graphic presentation. New York: Ronald Press.

- Schutz, H.G. (1961a). An evaluation of formats for graphic trend displays. Human Factors, 3, 95-107.
- Schutz, H.G. (1961b). An evaluation of methods for presentation of graphic multiple trends. Human Factors, 3, 108-119.
- Shepard, R.N. (1964). Attention and the metric structure of the stimulus space. Journal of Mathematical Psychology, 1, 54-87.
- Siegel, J.H., Goldwyn, R.M., & Friedman, H.P. (1971). Pattern and process in the evolution of human septic shock. Surgery, 70, 232-243.
- Sinclair, H.J. (1971). Digital versus conventional clocks: A review. Applied Ergonomics, 2, 178-181.
- Smith, L.B., & Kemler, D.G. (1977). Developmental trends in free classification: Evidence for a new conceptualization of perceptual development. Journal of Experimental Child Psychology, 24, 279-298.
- Stroop, J.R. (1935). Studies of interference in serial verbal reaction. Journal of Experimental Psychology, 18, 643-662.
- Torgerson, W.S. (1958). Theory and methods of scaling. New York: Wiley.
- Treisman, A.M. (1969). Strategies and models of selective attention. Psychological Review, 76, 282-299.
- Treisman, A., Kahneman, D., & Burkell, J. (1983). Perceptual objects and the cost of filtering. Perception and Psychophysics, 33, 327-332.
- Treisman, A.M., & Gelade, G. (1980). A feature-integration theory of attention. Cognitive Psychology, 12, 97-136.
- Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. Cognitive Psychology, 14, 107-141.
- Treisman, A., Sykes, M., & Gelade, G. (1977). Selective attention and stimulus integration. In S. Dornic (Ed.), Attention and performance VI (pp. 333-361). Hillsdale, N.J.: Lawrence Erlbaum.
- Treisman, A., & Paterson, R. (1984). Emergent features, attention, and object perception. Journal of Experimental Psychology: Human Perception and Performance, 10, 12-31.
- Tilling, L. (1975). Early experimental graphs. The British Journal for the History of Science, 8, 193-213.
- Tufte, E.R. (1983). The visual display of quantitative information. Cheshire, CT: Graphics Press.
- Tukey, J.W. (1977). Exploratory data analysis. Reading, MA: Addison-Wesley.

- Vernon, M.D. (1952). The use and value of graphical methods of presenting quantitative data. Occupational Psychology, 26, 22-34.
- Wainer, H. (1980). A test of graphicacy in children. Applied Psychological Measurement, 4, 331-340.
- Wainer, H., & Reiser, M. (1978). Assessing the efficacy of visual displays. In Graphic Presentation of Statistical Information (Technical Paper #43). Washington, D.C.: Census Bureau.
- Wainer, H., & Thissen, D. (1981). Graphical data analysis. Annual Review of Psychology, 32, 191-241.
- Ward, T.B. (1980). Separable and integral responding by children and adults to the dimensions of length and density. Child Development, 51, 676-684.
- Washburne, J.N. (1927). An experimental study of various graphs, tabular and textual methods of presenting quantitative material. Journal of Educational Psychology, 18, 361-376 (part one); 465-476 (part two).
- Wickens, C.D. (1984). Engineering psychology and human performance. Glenview, IL: Scott, Foresmen Pub. Co.
- Wickens, C.D., & Carswell, C.M. (1987). The object display: Principles and a review of experimental findings. Unpublished manuscript, University of Illinois at Urbana-Champaign, Aviation Research Lab.
- Wickens, C.D., Kramer, A., Barnett, B., Carswell, M., Fracker, L., Goettl, B., & Harwood, K. (1985). Display/cognitive interface: The effect of information integration requirements on display formatting for C³ displays (Technical Report EPL-85-2/RADC-85-1). Urbana, IL: Engineering-Psychology Research Laboratory, University of Illinois at Urbana-Champaign.
- Wickens, C.D., Sandry, D., & Vidulich, M. (1983). Compatibility and resource competition between modality of input, output, and central processing: Testing a model of complex task performance. Human Factors, 25, 227-248.
- Wickens, C.D., Vidulich, M., & Sandry-Garza, D. (1984). Principles of S-C-R compatibility with spatial and verbal tasks: The role of display-control location and voice-interactive display-control interfacing. Human Factors, 26, 533-543.
- Wilkinson, L. (1981). An experimental evaluation of multivariate graphical point representation. In Proceedings of the Conference on Human Factors in Computing Systems. Gaithersburg, MD: ACM.
- Wing, A., & Allport, D.A. (1972). Multidimensional encoding of visual form. Perception and Psychophysics, 12, 474-476.

- Woods, D., Wise, J., & Hanes, L. (1981) An evaluation of nuclear power plant safety parameter display systems. In R. C. Sugarman (Ed.), Proceedings of the 25th Annual Human Factors Society Meeting. Santa Monica, CA: Human Factors Society.
- Wrightstone, J.W. (1936). Conventional versus pictorial graphs. Progressive Education, 8, 460-462.
- Zeff, C. (1965). Comparison of convention and digital time displays. Ergonomics, 339-345.
- Zeki, S.M. (1978). Functional specialization in the visual cortex of the rhesus monkey. Nature, 274, 423-428.
- Zmud, R.W. (1978). An empirical investigation of the dimensionality of the concept of information. Decision Sciences, 9, 187-195.