2 The Visual Analysis of Data, and Current Research into the Stimuli Controlling It

Barry S. Parsonson
University of Waikato

Donald M. Baer
University of Kansas

THE CASE FOR VISUAL ANALYSIS

Visual analysis is one of the oldest forms of data analysis. It is, in essence, simply the array of one set of information relative to one or more other sets of information, so that a viewer can draw a reasonable conclusion or make a reasonable hypothesis about any relationships or lack of them among these sets. Very often, the array takes the form called a graph, especially when at least one set of the information is quantitative; in other cases, especially when all the sets are qualitative, it takes the form of essentially a picture or diagram (cf. especially maximizing dramatically the efficiency of constructing and disseminating visual analyses; these are the modern computer graphics programs. Yet in behavioral research, which is the context for this book (and this chapter), by far the prevailing mode of analysis is statistical, not visual. The multivariate analysis of variance, culminating in a table not of data numbers but of numbers testifying primarily to the probability that the data patterns could have arisen by chance, is far more frequent than a picture of the underlying data numbers themselves. Yet we can have such a picture, often by the construction of lines or other geometrical forms to show the relationship of some behaviors to the variables that may control or otherwise relate to them.

Thus, the purpose of this chapter is to state again the case for the visual analysis of behavioral relationships through graphs, and most especially for the outcome of experiments in which an ongoing, repetitive behavior is altered in its time course by the deliberate, repetitive alteration of one or more of its environmental conditions. In that context, there are at least six advantages to be gained through graphic analysis:

1. It is visual, and thereby quick to yield conclusions and hypotheses.
2. Graphs can be quick and easy to make with no more technology than grid paper, pencil, and straight edge. However, if the latest computer graphics technology is to be used, then speed and ease are recaptured only after an initial high cost of money, time, and training.
3. Graphing comprises a remarkably wide range of formats, even outside of the latest computer graphics technology.
4. Graphed messages are immediately and enduringly accessible to students at unusually diverse levels of training.

with them are in fact not functional for them. Instead, graphs invite us to make that judgment ourselves (as well as many others exemplified herein). Thereby graphs create two audiences among researchers: (a) Some researchers will see that kind of judgment as merely a personal one. They seek a science based on objective rather than subjective judgments, so they must go further than graphic analysis in their search for an apparently objective estimate of the probability. Statistical analysis will seem to offer it to them—until they have studied its workings and underlying assumptions thoroughly enough to see how many essentially personal judgments they must make in order to estimate the suitability of each of the many models of statistical analysis for their problem and its data. (b) Some other researchers will prefer to make their own judgments about the patterns and distribution of the data; they will compare those judgments with interest but not submissively to the judgments of their peers. They will be glad to have all the data as accessible as possible, especially as simultaneously accessible as possible, and will avoid any technique that makes a decision for them in ways that are exceptionally difficult to inspect in its reduction, transformation, collation, and integration of each data point into some final decision.

To illustrate many of these points, suppose that we want to know if a certain well-measured but momentarily unnamed behavior, B, is affected differently by two well-controlled but unnamed environmental conditions, Condition 1 and Condition 2. We can experimentally alternate Condition 1 and Condition 2, each for varying lengths of time, holding everything else as stable as we can, and meanwhile measuring behavior B steadily and repeatedly under each repetition of

TABLE 2.1
Hypothetical Data Gathered Under Two Conditions

Condition 1	*Condition 2*	*Condition 1*	*Condition 2*
1	5	4	11
1	6	2	10
2	6	0	11
0	8	2	9
1	7	1	9
2	9	1	8
1	8	2	7
0	11	0	7
1	10	0	9
1	10	1	8
	11	2	7
	9	1	6
	8		7
	8		5
	9		5
	8		
	10		

glance how much more of the behavior is produced by Condition 2 than Condition 1. It also suggests that the ability of Condition 2 to make this change in the behavior is perhaps decreasing by the end of its second application. That possibility, too, is perhaps not clear enough to be affirmed flatly by every viewer, but it, too, its clear enough to investigate further.

CONDITION CONDITION CONDITION CONDITION

Of course, all of those conclusions and hypotheses can be extracted from the preceding table as well—the data are identical. The reader should ask how quickly and how fully both the clear and possible relationships of B to Conditions 1 and 2 were extracted from the table, relative to how quickly and how fully they were extracted from the graph. Each method of presentation offers us all the information that we collected; the table offers it to us as an array of numbers relative to their controlling variable, and the graph offers it to us as a picture of those numbers in that same array. The picture is a transformation of the numbers, certainly, but it is a remarkably minimal transformation of them. It still leaves the judgment of whether this behavior does indeed relate differentially to Conditions 1 and 2, and if so, exactly how, entirely up to us, the viewing researchers—the people who want those answers, and have been trained as best we know how to find them out.

Another interesting question is whether the viewer can decide whether the Condition 1 distributions of B are no more different from the Condition 2 distributions of B than could reasonably have arisen by chance if Conditions 1 and 2 are not functional for B. A fair variety of statistical analyses would assure these viewers that the differences in these two distributions could hardly have arisen by chance; a few other statistical methods would suggest to them that they could. Experienced visual analysts (we daresay) will typically decide fairly quickly that these differences are certainly not chance differences, and inexperienced visual analysts will come to the same conclusion even more quickly, interestingly enough. In our experience, they tend to respond to the entire picture as a clang! phenomenon, and not go on to examine the time course of each data path, point by point, and then reconsider them as a pattern of four aggregates, two with quite different internal trend patterns (the two Condition 2s), one without (the first Condition 1), and one either without or with a very short-lived one (the second read graphs, but behave in a variety of other ways that confirm their premise. Yet most of us can see effortlessly when the lines of a graph, despite their local irregularities, as a whole are at different levels, and slant upward or downward or are horizontal, or change systematically from one of those states to another. Where did we learn that? One would think that these are easy generalizations from the facts that sighted people have already learned to appreciate–at a glance!–in the rest of their world. After all, their histories with lines are vastly more extensive than their histories with numerals: They constantly scan the world within which they move for the height, rise, fall, and stability of its contours, on which they must stand, sit, walk, ride, push, pull, drive, and place or retrieve something, and over which they will climb or jump. The "horizon" of "horizontal" is a key part of their eternal stimulus controls for where they are, where they are going, and for their balance. By contrast, their behavior with numerals is a mere occasional hobby.

On the other hand, sighted people read graphs quickly and easily only to the extent that they generalize from their extensive training in a real, three-dimensional visual world to the relatively new two-dimensional visual world of graphs; and behaviorists have learned that failures of generalization are never oddities. Of course there will be people who do not immediately and easily read in graphs everything that the graphs offer. By contrast, we learned to respond to numerals in the same domains in which they are used to represent data and their statistical analyses—on pages, chalkboards, and monitor screens. Numerals present no problem in setting generalization.

Thus, it is clear that despite the advantages of graphic analysis as a quick, easy, open, and widely accessible process, it is also true that both graph construction and graph reading are skills to be acquired. In recognition of that, a chapter like this in past years would sketch some skills from both of those skill classes.

aimed at uncovering the stimuli that currently control visual analysts' judgments of what the data show, even if not illuminating the source or the modifiability of those judgments.

Thus, research to date has focused on the extent to which visual analysts can see the major kinds of effects that can be expected from experimental interventions into an ongoing baseline of behavior that can be graphed as a time course: simple changes from baseline in the mean level of the behavior under study, usually referred to as *mean shift*; changes in which each successive data point is an interactive function of the intervention and the level of the immediately prior data point(s), yielding at least changes in trend and sometimes simultaneous changes in both level and trend, and often referred to as *serial dependency*; degrees of such changes, simple or interactive, relative to both baseline and intervention variability, thereby affecting the degree to which the ranges of the baseline and intervention data points overlap; and changes only in degree of variability. Research has also considered a few factors orthogonal to the behavior change accomplished by the intervention, such as different graphing techniques: for example, adding trend lines (calculable of course in a variety of ways) to the data path(s), or choosing logarithmic rather than arithmetic scales for the ordinate. Most of these topics will be discussed.

Serial Dependency

A pioneering study by Jones, Weinrott, and Vaught (1978) found that mean agreement between visual and time-series analyses was both low (60%) and, what little there was of it, inversely related to the degree of serial dependency in the data—that serial dependency decreased agreement between the two methods of analysis. This study also pioneered in two frequently unexamined, sometimes implicit assumptions: (1) That it is relevant to compare the seemingly ment about whether the current experimental intervention has had an effect on the behavior under study. Perhaps that is done in honor of the frequency with which science has been advanced by the disagreements of a few data analysts with the prevailing conventional wisdom about what the current data mean. One or both of these typically unexamined assumptions operates in virtually every study that will be reviewed here.

For example, to evaluate different types of serial dependency, Rojahn and Schulze (1985) used computer-generated AB-design graphs showing no, weak, or strong serial dependencies, some of the moving-average type and some of the autoregressive type; in addition, the graphs were constructed to represent five different significance levels of treatment effect (p = .50, .10, .05, .01, .001). These researchers then asked judges to rate 70 such graphs on a 5-point scale made up of just those probability values. The results did not support those of Jones et al. (1978), in that serial dependency did not much affect agreement between visual and statistical analyses. Rather, it was found that the more pronounced the moving-average and autoregressive effects, the greater the agreement between the two modes of analysis. The moving-average and autoregressive processes affected this agreement somewhat differently; strong autoregressive processes in particular led judges to overestimate treatment effects relative to statistical analysis.

Studies by Ottenbacher (1986) and Gibson and Ottenbacher (1988) investigated the effects of six graphical characteristics, including serial dependency, on interjudge agreement. Ottenbacher (1986) asked 46 occupational therapists to indicate if change had occurred between phases on five specially devised AB graphs. Gibson and Ottenbacher (1988) obtained ratings (0–5) of significance of submitted for judgment to 108 editorial board members and guest reviewers of two behavior-analytic journals. Their graphs represented three patterns of mean shift: "ideal," in which the phase means changed as "expected" with "experimental" conditions; "inconsistent," in which the mean shifted only in the final B phase; and "irreversible," in which the mean shifted only in the first B phase and remained stable thereafter. In addition, three degrees of mean shift were displayed, and while this varied from graph to graph, it was held constant within graphs. In this study, the pattern of mean shift proved to be critical, on the average, in that mean ratings of "experimental control" were high only for at least some of the "ideal" graphs showing large degrees of mean shift. A statistical analysis revealed that pattern and degree of mean shift were highly significant main effects, yet together accounted for only a small proportion of the total variance. (It might be interesting, and certainly would be democratic, if similar analytic effort were invested to determine the extent to which highly significant *p* levels and proportion of variance accounted for are controlling variables in the conclusions of given populations of statistical-analysis consumers, especially in that typical publication greatly emphasizes the former over the latter.)

Knapp (1983) used AB graphs with a mean value of 5 in baseline and a range of nine mean values between 2 and 8 in the B phase, to see how that would affect the judgments of three different groups: that much-studied subpopulation, the editorial-board members of two behavior-analytic journals, as well as graduate behavior-analysis students and undergraduate psychology majors. A statistical analysis revealed significant main effects for mean shift, graph type, and their interaction. Extreme (e.g., 5-to-8 and 5-to-2) or zero (5-to-5) mean shifts were judged similarly regardless of graphing technique. At more moderate levels of mean shift (e.g., 5-to-3.5 or 5-to-6.5), graphing technique became critical. Judgments generally were comparable for those types of graphs commonly evaluated

Ottenbacher, 1988; Ottenbacher, 1986). Mean shift will be present when there is no change in level or trend and no intervention effect between phases in the case of an upward/downward baseline trend continuing into an intervention attempt. In this case, a visual analyst relying only on mean shift as an indicator of an intervention effect will report change quite frivolously. Experience suggests that this misjudgment is more likely to be made if variability in the data paths masks perception of the absence of change in level and/or trend, or if phase means lines, emphasizing the shift, have been added to the graph. One way of minimizing such errors of judgment is to use trend lines (Baer & Parsonson, 1981).

Level and Trend

Half of the graphs generated by DeProspero and Cohen (1979) had a 30° upward trend, the remainder had zero slope. Their statistical analysis did not identify slope as a significant variable, although the data in their Table 1 reveal that judges' ratings almost always indicated less "experimental control" in graphs with that slope than in equivalent graphs without it. This points to some visual analysts responding to absence of trend as indicative of control. Indeed, it was the criterion most frequently mentioned by the judges.

In a study of teachers' abilities to discern trends, Munger, Snell, and Loyd (1989) investigated the effects of ascending, descending, flat, and flat but variable data paths, and four different weekly frequencies of probe-data collection (1, 2, 3, or 5 days per week). The teachers rated the degree of student progress in reading accuracy that they could see in the graphs, and also made mock decisions on program continuation. The graphs were derived from student records representing each of the data-path trends, modified to show different probe frequenchanges in level, trend, and both level and trend. They transformed these graphs in three ways (referred to as standard, scaling, and variation transformations) to modify the visual appearance of these data paths. The judges were a group of graduate students studying single-subject research methodology, a group of students studying multivariate statistics (Wampold & Furlong, 1981), and the inevitable sample of editorial-board members of a behavior-analytic journal (JABA) (Furlong & Wampold, 1982). These judges were asked first to sort 36 graphs into whatever number of groups were justified as showing "similar effects" different from the effects of the other groups; and then to identify the common feature(s) of the graphs in each group. The single-subject students often responded primarily to the absolute size of the change from A to B; they were influenced most by those scaling transformations that enhanced both the variation and the size of the intervention effect. The multivariate students were mainly influenced by changes in level and/or trend, which often were not discriminated as such. The JABA reviewers typically did discriminate changes in level, trend, and level plus trend, and also attended frequently to absolute size of effect. The researchers suggested that the single-subject students and JABA reviewers were so strongly influenced by size of effect, rather than by more subtle relative variations in the data, because the identification of large changes has been emphasized as crucial to the character of applied behavior analysis (e.g., by Baer, 1977). (However, they drew this conclusion without evidence that their judges had ever even read those arguments, let alone agreed with them.) Ottenbacher (1986) also found that the detection of trend changes as such was made moderately difficult by the judges' ideas about clinically significant changes; the relevant correlation was 0.59. This finding was replicated by Gibson and Ottenbacher (1988), who obtained a similar correlation of 0.64, and also found raters' confidence in their ... trend changes across phases; they conby the work of Cleveland and his colleagues (Cleveland, 1985; Cleveland & McGill, 1987a; Cleveland, McGill, & McGill, 1988). In what may be a recapitulation of the Weber–Fechner Law, their subjects' judgments of slope depended heavily on the magnitude of the relative angle that the data paths create when graphed, and that how steep an angle differential is needed to be seen as such varies widely across judges. Thus, research into ways to increase accurate discrimination of relative angle differential, independent of their absolute magnitudes, is relevant. The use of trend lines for that purpose is reported later.

So far, the investigation of trend-change detection skills has focused on abrupt, sustained changes between phases. The detection of delayed or temporary changes, especially within-phase changes, which also are relevant to visual analysis of behavioral data (Parsonson & Baer, 1978, 1986), remains unanalyzed. Similarly, the study of level-change detection has investigated either change in overall phase level (Furlong & Wampold, 1982; Wampold & Furlong, 1981) or abrupt change between the last data point in one phase and the first data point in the succeeding phase (Gibson & Ottenbacher, 1988). Delayed and temporary within-phase level changes (Parsonson & Baer, 1978) are also important real-world processes, but would not be caught (even if present) by the change-of-level judgments required in the studies reviewed so far. This is of course not a criticism of those studies, only a reminder that research examining judgments of within-phase changes, and their interaction with between-phase changes, is needed for a more complete understanding of visual analysis.

Variability and Overlap

The effect of variability in the data path has seen frequent study. DeProspero and