CHAPTER 9

VISUAL ANALYSIS IN SINGLE-CASE RESEARCH

Jason C. Bourret and Cynthia J. Pietras

The visual analysis, or inspection, of graphs showing the relation between environmental (independent) variables and behavior is the principal method of analyzing data in behavior analysis. This chapter is an introduction to such visual analysis. We begin by describing the components, and construction, of some common types of graphs used in behavior analysis. We then describe some techniques for analyzing graphic data, including graphs from common single-subject experimental designs.

TYPES OF GRAPHS AND THEIR CONSTRUCTION

Of the many ways to graph data (see Harris, 1996), the graph types most frequently used by behavior analysts are cumulative frequency graphs, bar graphs, line graphs, and scatterplots. Each of these is described in more detail in this section.

Cumulative Frequency Graphs

Cumulative frequency graphs show the cumulative number of responses across observation periods. The earliest, and most common, such graph used by behavior analysts is the cumulative record. In the other graph types discussed in this section, measures of behavior during an observation period are collapsed into a single quantity (e.g., mean response rate during a session) that is represented on a graph by a single data point. By contrast, cumulative records show each response and when it occurred during an observation period. Thus, cumulative records provide a detailed picture of within-session behavior patterns (see Ferster & Skinner, 1957/1997). An example of a cumulative record is shown in Figure 9.1. On a cumulative record, equal horizontal distances represent equal lengths of time; equal vertical distances represent equal numbers of responses. The slope of the curve in a cumulative record indicates the rate of responding. Researchers sometimes include an inset scale on cumulative records to indicate the rate of responding represented by different slopes, although usually more precise calculations of rate are also provided. Small vertical lines oblique to the prevailing slope of the line, traditionally called *pips*, typically indicate reinforcer deliveries. When the response pen reaches the top of the page, it resets to the bottom, producing a straight vertical line. Researchers may also program the response pen to reset at designated times (e.g., when a schedule change occurs), to visually separate data collected under different conditions. Cumulative records were traditionally generated by now-obsolete, specially designed machines (cumulative recorders). More contemporarily, computer software programs that record and plot each response as it occurs have been used to generate these records. Cumulative records can also be constructed after data collection is complete, if the time of occurrence of each response and all other relevant events during a session are recorded.

Although the cumulative record was one of the most commonly used graphs in the early years of the experimental analysis of behavior, it has since fallen out of favor as researchers increasingly present data averaged across a single or multiple sessions. It is

DOI: 10.1037/13937-009
APA Handbook of Behavior Analysis: Vol. 1. Methods and Principles, G. J. Madden (Editor-in-Chief)

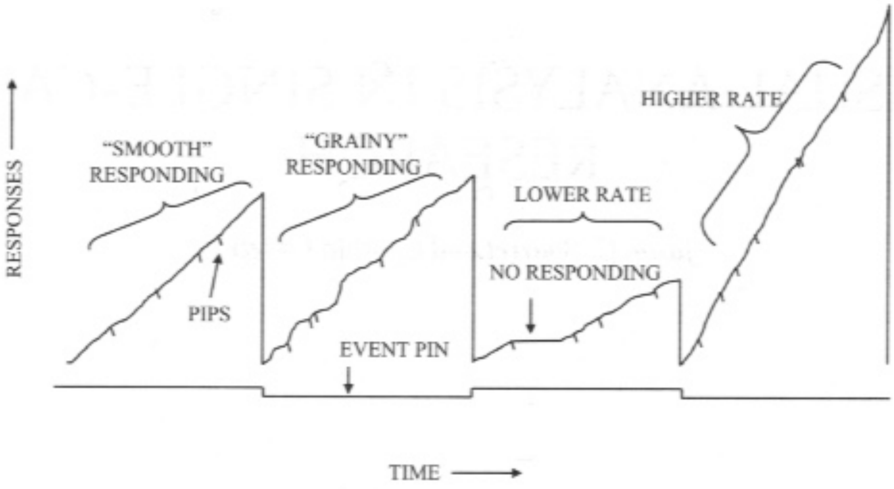

FIGURE 9.1. **Example of patterns that may be observed on a cumulative record. Cumulative responses are shown on the vertical axis, and time is shown on the horizontal axis. Each response moves the response pen a constant distance in the vertical direction, and the paper moves vertically at a constant speed. Shown (right to left) are smooth curves indicating constant rates of responding, grainy curves indicating irregular patterns of responding, shallow curves indicating low rates of responding, and steep curves indicating high rates of responding. Flat portions indicate no responding. Pips (downward deflections of the response pen) usually indicate reinforcer deliveries. Movements of the event pen are used to signal changes in experimental contingencies or stimulus conditions. Data are hypothetical.**

useful in the experimental analysis of behavior not as a primary means of data analysis, but as a means of monitoring within-session performance. Another notable exception to the decline of the cumulative record is research published by Gallistel and his colleagues (e.g., Gallistel et al., 2007). In this research line, cumulative records (some of them quite creatively constructed with more than simple responses on the vertical axis) are used extensively to better understand the process by which organisms allocate their behavior between concurrently available sources of food.

Bar Graphs

Bar graphs show discrete categories (i.e., a nominal scale; Stevens, 1946) along the horizontal (x) axis and values of the dependent variable on the vertical (y) axis (Shah, Freedman, & Vekiri, 2005). In behavior analysis, bar graphs are often used to show percentages (e.g., of correct responses) or average performance across stable sessions or conditions. As shown in Figure 9.2, bar graphs facilitate comparisons of performances (i.e., the height of each bar) across conditions. Typically, bars are separated from each other, but related bars may be grouped together. One variation of the standard (vertical) bar graph is a horizontal bar graph, in which the categorical variable is plotted on the y-axis. On these graphs, the length of the bar along the x-axis shows the value of the dependent variable. Bars graphs may also be drawn so that bars can deviate in either direction from a center line. Such graphs may be used, for example, to show increases or decreases from baseline values that are represented by a center horizontal line at zero. Bar graphs are similar to histograms, but histograms (in which the bars touch each other) have interval x-axis values and typically show frequency distributions (see Figure 9.3). In bar graphs, the y-axis scale usually begins at the lowest possible value (e.g., zero) but may begin at a higher value if the low range would be devoid of data. Constraining the lower range of y-axis values will make differences across conditions appear bigger than they would have been had the range extended to zero, a factor to consider when evaluating the clinical (or other) importance of the visually

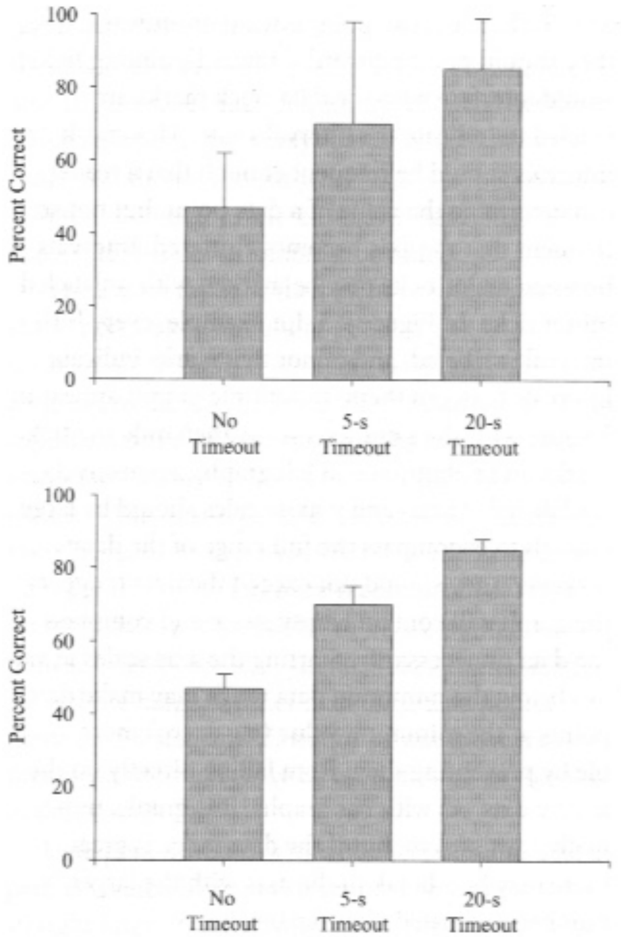

FIGURE 9.2. Examples of bar graphs. Data are hypothetical.

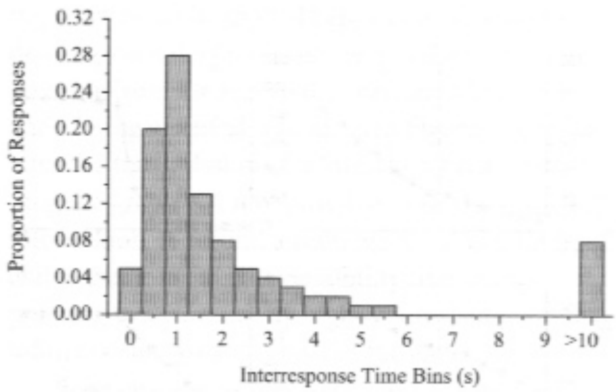

FIGURE 9.3. Example of a histogram. Data are hypothetical.

apparent difference. When bar graphs show measures of central tendency (e.g., means), error bars (the vertical lines in Figure 9.2) should be included to depict the variance in the data (e.g., between-session differences in percentage correct).

When group statistical designs are used, bar graphs frequently summarize the differences in group mean performances. Reliance on statistical analyses of grouped data may lead to the omission of error bars from such graphs, a practice that obscures the size of individual differences within groups. Figure 9.4 shows a variation on the between-groups bar graph that displays the performance of individual participants (individual data points) while maintaining the ease of comparing measures of central tendency (height of the bars). Graphs constructed in this way make it possible to evaluate to what extent the height of the bar describes the behavior of individual participants. A second advantage of this type of bar graph is that it allows readers to evaluate whether the within-group variance is normally distributed, an important factor when evaluating the appropriateness of the statistical analyses used. Finally, the graphing conventions of Figure 9.4 encourage the researchers to consider those individuals in the treatment group for whom the treatment produced no positive effect. As often noted by

FIGURE 9.4. A bar graph that shows mean performance and the performance of individuals making up the mean. The height of the bar shows the mean percentage of urine samples negative for cocaine and opiates, and the closed circles show the percentage of negative samples for each individual undergoing employment-based abstinence reinforcement treatment for cocaine dependence. From "Employment-Based Abstinence Reinforcement as a Maintenance Intervention for the Treatment of Cocaine Dependence: A Randomized Controlled Trial," by A. DeFulio, W. D. Donlin, C. J. Wong, and K. Silverman, 2009, *Addiction, 104*, p. 1535. Copyright 2009 by John Wiley & Sons, Ltd. Used with permission.

Sidman (1960), the data from these individuals serve to illustrate that the behavior is not fully understood and, by investigating the factors affecting these individuals' behavior further, more effective interventions will follow.

Line Graphs (Time Series and Relational)

Line graphs are used to depict behavior across time (time-series line graphs) or relations between two variables (relational line graphs; see Tufte, 1983). With time-series line graphs (e.g., see Figure 9.5), the horizontal axis (referred to as the *abscissa* or the *x*-axis) illustrates the time point at which each data point was collected, and behavior is plotted at each of these time points on the vertical *y*-axis (ordinate). On relational line graphs, the *x*-axis shows values of the independent variable, and the *y*-axis shows a central tendency measure of the dependent variable. Data points in both types of line graphs are connected with straight lines.

Figure 9.5 shows the parts of a time-series line graph. Axis labels indicate the variables plotted. If multiple graphs appear in a figure with the same axes, then, to reduce the amount of text in the figure, only the *x*- and *y*-axes on the bottom and left-most graphs, respectively, are labeled (for an example, see Figure 9.6). The scales of both the *x*- and *y*-axes are divided into intervals. Successive intervals on an axis are equally sized and are marked with short lines, called tick marks, that intersect the axis. Tick marks can point inward or outward, but they should point outward if inward-pointing ticks would interfere with the data. Tick marks are labeled to indicate the interval value. Tick-mark intervals should be frequent enough that a reader can determine the value of a data point, but not so frequent that the axis becomes cluttered. Intervals between major ticks may be marked with unlabeled minor ticks. In Figure 9.5, for example, every fifth interval is labeled, and minor tick marks indicate intervals between them. If multiple graphs appear in a figure with the same axis scale, then only the tick marks on the bottom and left graphs, respectively, are labeled. The *x*- and *y*-axis scales should be large enough to encompass the full range of the data; however, they should not exceed the data range, or the graph will contain empty space and compress the data unnecessarily. Starting the axis scales at values below the minimum data range may make data points at the minimum value (e.g., zero) more visible by preventing them from falling directly on the *x*- or *y*-axis. As with bar graphs, line graphs normally start at zero, but if the data range is great, there may be a break in the axis with the larger numbers indicated after the break.

When plotting data for individual participants in separate graphs within a single figure, it is sometimes not realistic to represent data for each

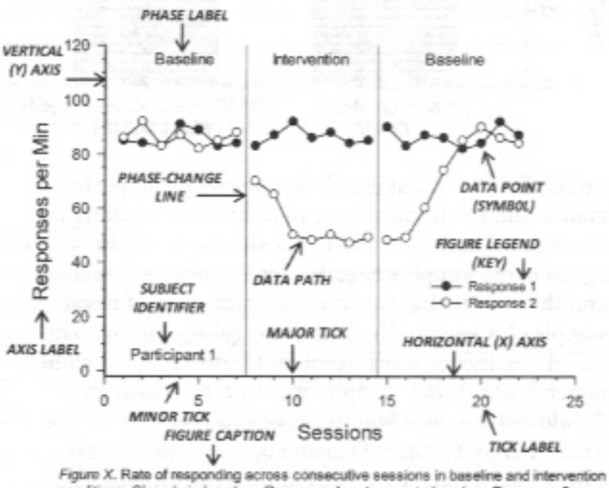

FIGURE 9.5. Diagram of parts of a time-series line graph. Data are hypothetical.

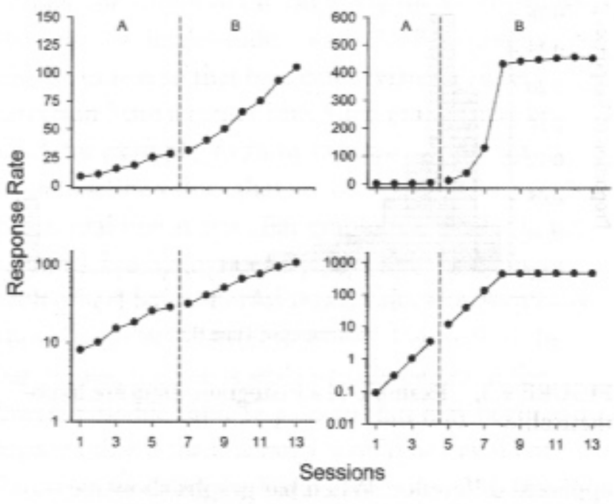

FIGURE 9.6. Examples of linear and logarithmic (log) scales. The upper two graphs show data plotted using linear *y*-axis scales. The lower two graphs show the same data, but plotted using log (base 10) *y*-axis scales. Data are hypothetical.

participant within a single range on the *y*-axis (e.g., the range for one participant may be between one and 10 responses and for another, between 400 and 450 responses). It is always better to use the same ranges on the *y*-axis, but when this is not possible, different ranges for different participants may be used, and this deviation must be noted.

The aspect ratio, or the height-to-width ratio of the *y*- and *x*-axes, should not distort the data. Too great an aspect ratio (i.e., a very tall graph) may magnify variability, or small effects, and too small an aspect ratio (i.e., a very long graph) may obscure important changes in behavior or variability in a data set (Parsonson & Baer, 1986). A 2:3 *y:x* aspect ratio (Parsonson & Baer, 1978) or a 1.0:1.618 aspect ratio (Tufte, 1983) has been recommended. Breaks on the *y*-axis may be used if there are outlier data points. *Outliers* are idiosyncratic data points that far exceed the range of other data (e.g., data points more than 3 standard deviations from the mean). Breaks on the *x*-axis may be used if there are breaks in data collection.

Data points are marked with symbols, and a data path is created by connecting the data points with straight lines. When multiple data paths are shown on a graph, each data type is represented by a distinct symbol, and a central figure legend provides a concise description of each path. A common graphing convention in applied behavior analysis is to describe each data path with text and an arrow pointing from the text to the corresponding data path. Using a central legend, as in Figure 9.5, facilitates the transmission of information because scanning the graph is not required to find figure legend information. A second advantage of the central legend is that it avoids the possibility that arrows pointing to a particularly high or low data point may influence the visual analysis of the data.

Phase changes in time-series line graphs are denoted by vertical lines extending the length of the *y*-axes. Phase-change lines are placed between the last data point of a condition and the first data point of the new condition. Data paths are broken across phase-change lines to avoid the appearance that behavior change produced by an independent variable occurred before the change in condition. Descriptive phase labels are centered at the top of the space allocated to each phase. Figure legends, and phase labels, are usually placed within the rectangular space created by the *x*- and *y*-axes. Figure captions are placed below graphs and describe what is plotted, the axes, any abbreviations or symbols that appear in the graph, and any axis breaks.

Linear interval scales are the most common scales used in time-series line graphs, but logarithmic (log) interval scales and semi-log interval scales (in which the *x*- or *y*-axis is log scaled and the other is linearly scaled) are also used. Log scales are helpful in more normally distributing data sets that are skewed toward large values (Cleveland, 1994), transforming curvilinear data into linear data (Shull, 1991) and showing proportional changes in behavior (Cooper, Heron, & Heward, 1987). Because the logarithm of zero is undefined, log scales have no zero. Log base 10, base 2, and base *e* (natural logs) are the most common log scales (see Cleveland, 1994, for some recommendations for the use of various log bases).

Illustrations of data plotted on both a linear scale and a semi-log (base 10) scale are shown in Figure 9.6. In the upper left graph, response rates in Phase B appear to increase more quickly between sessions than in Phase A. This difference, however, may be attributed to the greater absolute value of the response rate. Plotted on a log scale (lower left graph), it is visually apparent that the rate of change is similar in both phases. In the upper right graph, there appears to be a large shift in performance from Phase A to Phase B. The arithmetic scale of the *y*-axis, however, compresses the low rates in Phase A. When data are plotted on a log scale, the low rates are more visible and the increase in responding in Phase B can be seen to be part of an increasing trend that began in Phase A.

Scatterplots

Scatterplots present a dependent variable in relation to either an independent variable (in which case the graph may be described as a *relational graph*; see Tufte, 1983) or another dependent variable. When both measures are dependent variables, either measure can be plotted on the horizontal axis, although if one measure is conceptualized as a predictor variable, it is plotted on the *x*-axis, and the other, the

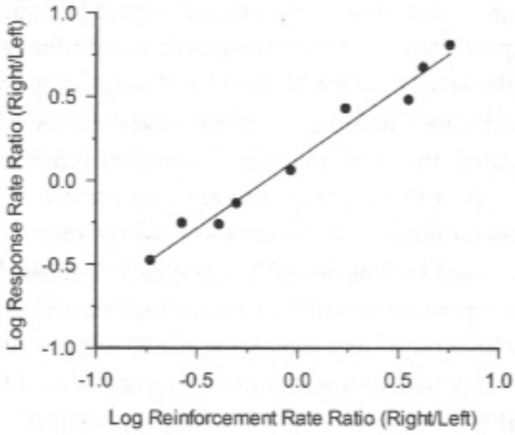

FIGURE 9.7. Example of a scatterplot. Data are hypothetical.

criterion variable, is plotted on the *y*-axis. An example of a scatterplot is shown in Figure 9.7. In this figure, which shows data from a hypothetical experiment investigating choice between two concurrently available reinforcement schedules, the log of the ratio of response rates on the right and left alternatives is plotted on the *y*-axis and the log of the ratio of reinforcement rates is plotted on the *x*-axis. In scatterplots, data points are not connected with lines, usually because measures are independent of each other (e.g., they are data points from different conditions or participants) or because they are not sequentially related. Sometimes, however, lines or curves are fit to data on scatterplots to indicate the form of the relation between the two variables (see Interpreting Relational Graphs section). The line in Figure 9.7 shows the best-fitting linear regression line. That data points fall near this line indicates a linear relation (matching) between response rates and reinforcement rates.

Other Types of Graphs

The types of graphs we have discussed do not represent an exhaustive list of the types of graphs used by behavior analysts and subjected to visual analysis. For example, sometimes examining data across time and within individual sessions is useful, in which case a three-dimensional graph would be appropriate, with the dependent variable on the *y*-axis, within-session time on the *x*-axis, and successive sessions on the third (*z*) axis (e.g., Cançado & Lattal, 2011). Three-dimensional graphs may also be used to show other types of interactions, such as changes in interresponse time distributions on a reinforcement schedule across sessions (Gentry, Weiss, & Laties, 1983) or effects of different drug doses on response run length distributions (Galbicka, Fowler, & Ritch, 1991).

Other graphing techniques have been used to depict specific kinds of relations. Staddon and Simmelhag (1971), for example, used detailed flow charts to graphically show the conditional probabilities of different responses given an initial response. Davison and Baum (2006) depicted the number of responses to different alternatives in a choice experiment as different-sized circles (*bubbles*). This technique could also be useful in showing, for example, time allocated to playing with multiple toys by a child across successive time periods.

These examples are but a few of specialized graphs that may be useful in enhancing the visual depiction of specific data sets or aspects of data sets. For a more complete description of graph types, see Harris (1996). Cleveland and McGill (1984, 1985) offered some useful advice on how to choose graph types to show data with maximum clarity. In undertaking a graphical analysis of data, there are no immutable rules concerning which graphs to use for depicting what. Use is based on precedence, but investigators also need to think outside the axes (so to speak) in using graphs to tell the story of their data.

General Recommendations for Graph Construction

Many features of a graph influence a reader's reaction to the data. Even small details such as tick marks, axis scaling, data symbols, aspect (*y*:*x*) ratio, and so forth can affect a graph's impact, and poor graphing methods can lead to misinterpretations of results (Cleveland, 1994). Creating graphs that are accurate, meaningful, rich in information, yet readily interpretable, therefore, requires planning, experimenting, reviewing, and close attention to detail (Cleveland, 1994; Parsonson & Baer, 1992). For some additional recommendations on producing useful graphs, see Baron and Perone (1998), Cleveland (1994), Johnston and Pennypacker (1993), Parsonson and Baer (1978, 1986), and Tufte (1983).

When preparing graphs for publication, the *Publication Manual of the American Psychological Association* (American Psychological Association, 2010) also offers valuable advice.

INTERPRETING GRAPHICAL DATA

In the sections that follow, we describe some strategies for visually analyzing graphical data presented in cumulative records, bar graphs, time-series line graphs, and scatterplots. We also discuss the visual analysis of graphical data generated by some commonly used single-subject experimental designs.

If the graph is published, the first step in visual analysis is to determine what is plotted by reading all of the text describing the graph, including the axis labels, condition labels, figure legend, and figure caption. The next step is the analysis of patterns in the graphical data.

Interpreting Cumulative Records

In cumulative records, changes in rate of responding and variability in responding are analyzed by examining changes in the slope of the records (Johnston & Pennypacker, 1993). Several patterns that may be distinguished in cumulative records are shown in Figure 9.1, a hypothetical cumulative record. The first smooth curve shows responding occurring at a steady, constant rate, whereas the second shows grainy responding, or responding occurring in unsystematic bouts of high and low rates separated by varying periods of not responding. The flat portion of the third curve indicates no responding. The greater slope of the fourth curve compared with the third curve indicates a higher rate of responding. Cumulative records also allow an analysis of responding at a more local level. For example, in Figure 9.1, the second curve from the left, between the third and fourth pip, shows that responding occurred first at a low rate, then rapidly increased, then gradually decreased again before the reinforcer delivery. Such a fine-grained analysis is not possible with other graph types.

Interpreting Bar Graphs

When visually analyzing bar graphs, the relative heights of bars are compared across conditions (see Figure 9.2). When making this comparison, attention should be given to the y-axis scale to determine whether the range has been truncated. In bar graphs depicting average performance within a phase, a critical element to evaluate is the length of the error bars. Very long error bars suggest that the performance may not have been stable, and so it will be important to evaluate the stability criterion used. If the range of the data is used, long error bars may also occur if an outlying data point was included in the data set depicted in the graph. If so, then the average value depicted by the height of the bar may not represent most of the data; in such cases, the median would be a better measure of central tendency. Error bars also indicate the overlap of data points across conditions. For example, Figure 9.2 shows results from a hypothetical experiment that evaluated the effects of three time-out durations after incorrect responses on match-to-sample performance. The height of the bar is the mean, and the error bar shows the standard deviation. In the top graph, the error bars are long, and the mean of the 20-second condition overlaps with the variance in the 5-second condition. Thus, differences between the 5- and 20-second time-out duration conditions are less convincing than the difference between the no time-out and the 20-second conditions. Error bars provide no information about trends in the data, however, and a reader must look to the text of the article or to other graphs for evidence that the performances plotted in a bar graph represent stable responding.

Care should also be taken to consider which measure of variability is represented by the error bars. The standard deviations plotted in Figure 9.2 quantify the average deviation of each data point from the condition mean and, therefore, are an appropriate measure of variability when mean values are reported (interquartile ranges usually accompany medians). Some errors bars will depict the standard error of the mean, and readers should interpret these with caution. The standard error of the mean is used to estimate the standard deviation among many different means sampled from a normally distributed population of values. As such, it tells one less about the variability in the data than does the standard deviation. Moreover, the standard error of the mean is calculated by dividing the sample standard deviation by the square root of n (i.e., the number of values used to calculate the mean); thus, error bars depicting the standard error of the mean will be increasingly more narrow than the standard deviation as the number of data points included in the data set increases. If the standard error of the mean had been plotted in Figure 9.2 instead of the standard deviation, the visually apparent difference between all three conditions would seem greater even with no change in the data set plotted.

The general strategies we have outlined (i.e., consider the difference in the measure of central tendency in light of the variability in the data to evaluate how convincing the difference is) are formalized by common inferential statistical tests. Behavior analysts wanting to reach a broader audience of scientists (including extramural grant reviews), professionals, and public policymakers may wish to use these tests in addition to conducting a careful visual analysis of the data.

Analyzing Time-Series Data in Line Graphs

The analysis of time-series data is the most prevalent visual inspection practice in behavior analysis. Basic and applied behavior analysts use visual analysis techniques to determine when behavior has stabilized within a phase, to judge whether behavior has changed between phases, and to evaluate the evidence that the experimental variable has affected individual behavior.

Evaluating stability. Once an experiment is underway, one of the first decisions a researcher must make is, When should a phase change be made? In most cases, this question is answered by evaluating the stability (i.e., consistency) of responding over time (see Chapter 5, this volume). If behavior is not stable before the condition change (e.g., there is a trend in the direction of the anticipated treatment effect), then attributing subsequent shifts in responding to the experimental manipulation will be unconvincing (Johnston & Pennypacker, 1993). Moreover, an unstable baseline (i.e., one containing a great deal of between-session variability) will inhibit one's ability to detect small but clinically important effects of an experimental manipulation (Sidman, 1960). Thus, whenever possible, conditions should remain unchanged until stability is achieved.

Stability of time-series data may be assessed by visual inspection or quantitative criteria (see Perone, 1991; Sidman, 1960; Chapter 5, this volume). Both will evaluate bounce and trend, with the latter catching patterns that may be missed by the quantitative criterion. *Bounce* refers to unsystematic changes in behavior across successive data points, whereas *trend* refers to a systematic (directional) change (Perone, 1991).

Figure 9.8 shows baseline data for two participants. The baseline depicted in the top panel has considerably more between-session variability than

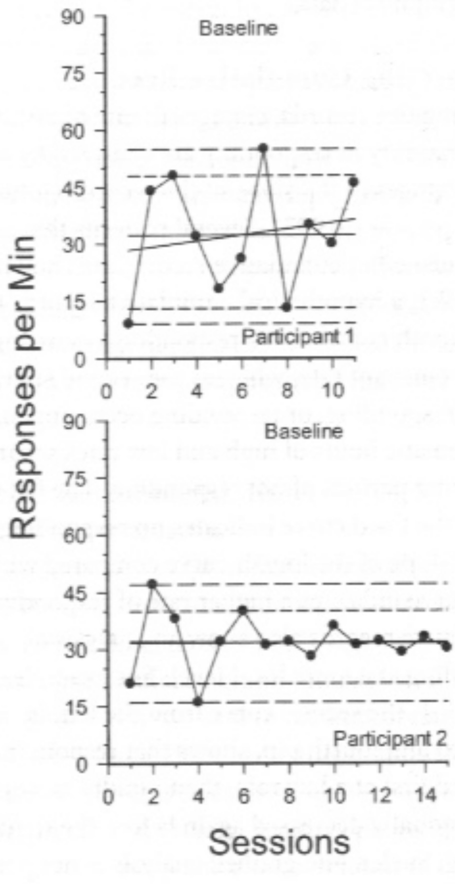

FIGURE 9.8. Hypothetical baseline data with added mean lines (solid horizontal lines), range lines (long dashed horizontal lines), trimmed range lines (short dashed horizontal lines), and regression lines (solid trend lines).

that depicted in the lower panel. The extent to which the researcher will be concerned with this bounce in the data will depend on how large the treatment effect is likely to be. If a very large effect is expected, then the intervention data should fall well outside of the baseline range, and therefore the relatively weak experimental control established in the baseline phase would be acceptable. If, however, a smaller effect is anticipated, then the intervention data are unlikely to completely separate from the range of data in the baseline, making detection of an intervention effect impossible. Under these conditions, the researcher would be well served to further identify the source of variability in the baseline. Indeed, if the researcher succeeds in this endeavor, a potent behavior-change variable may be identified.

Visually analyzing bounce may be facilitated by the use of the horizontal lines shown in Figure 9.8. The solid horizontal line shows the mean of the entire phase (i.e., the mean level of the data path) and allows one to see graphically how much each data point deviates from an ostensibly appropriate measure of central tendency.[1] The dashed lines furthest from the mean line illustrate the range of the data (i.e., they are drawn through the single data point furthest from the mean), whereas the dashed lines within these dashed lines show a trimmed range in which the furthest data point from the mean is ignored (see Morley & Adams, 1991). Drawing range and trimmed range lines may be useful when considering how large the intervention effect will have to be to discriminate a difference between the baseline and intervention data. Clearly, to produce a visually apparent difference, the intervention implemented in the top panel will have to produce a much larger effect than that implemented in the bottom panel. Neither range lines nor trimmed range lines will depict changes in variability within a condition, however. To visualize changes in variability within conditions, Morley and Adams (1991) suggested plotting trended range lines. To construct these, the data in a condition are divided in half along the *x*-axis, and the middle *x*-axis value of each half is located. For each half, the minimum and maximum *y*-axis data points are located, and those values are marked at the corresponding *x*-axis midpoint. Finally, two lines on the graph are drawn connecting the two minimum data points from each half and the two maximum data points from each half. Converging lines suggest decreasing variability (bounce) across the phase, diverging lines suggest increasing variability, and parallel lines suggest that the variability is constant.

The next characteristic of the baseline data to be considered, when deciding when to change phases, is the extent to which there is a trend in the data. A first step can be to plot a line of best fit (i.e., linear regression) through the baseline data. Any graphing software package will suffice. Researchers should be aware, however, that a line of best fit can be unduly affected by outliers (Parsonson & Baer, 1992). One alternative to linear regression that was recommended by Cleveland (1994) is the curve-smoothing loess technique. The loess technique is less sensitive to outliers and does not assume that data will conform to any particular function (e.g., a straight line). This technique smoothes data and makes patterns more visible by plotting, for each *x*-axis value, an estimate of the center of the distribution of *y* values falling near that *x* value (akin to a moving average; for descriptions, see Cleveland, 1994; Cleveland & McGill, 1985). Linear regression, however, has the advantage of being a more widely used technique, and it quantifies the linear relation between the two variables (i.e., estimates the slope and *y*-intercept).

In the upper panel of Figure 9.8, the line of best fit indicates an upward trend in the baseline data, suggesting that if no intervention is implemented, the rate of response will continue to increase over time. This is problematic if one expects the intervention to increase response rates. In the lower panel of Figure 9.8, the trend line is horizontal and overlaps with the mean line. Thus, in the absence of an experimental manipulation, the best prediction about future behavior is that it will remain stable with little between-session variability. Baseline data need not be completely free of trends before a phase

[1]Plotting a mean line is appropriate only if the data in the phase are free of extreme values that will pull the mean line away from the center of the distribution. Under such cases, a median line would be a better visual analysis tool.

is ended and the intervention is begun. If the baseline data are trending down (up), and the intervention is anticipated to increase (decrease) responding, then the baseline trend is of little concern. A modest trend in the direction of the anticipated intervention effect is also acceptable as long as the intervention proves to produce a large change in trend and mean level. Finally, continuing a baseline until it is free of trends and the bounce is minimal is sometimes impractical. In applied settings, it may be unethical to continue a baseline until stability has been achieved because to do so delays the onset of treatment. These practical and ethical concerns, however, must be balanced with the goal of constraining the effects of extraneous variables so that orderly effects of independent variable manipulations can be observed in subsequent conditions. It may be of little use (and may also be considered unethical) to conduct an intervention if the data are so variable that it is impossible to interpret the effects of treatment. Thus, researchers should be especially concerned with achieving stability when treatment effects are unknown or are expected to be small (i.e., when one is conducting research rather than practice).

Visual inspection of trends within a data set sometimes reveals nonlinear repetitive patterns, or cycles. Some cycles result from feedback loops created by self-regulating behavior–environment interactions (Baum, 1973; Sidman, 1960), whereas others result from extraneous variables. Identifying the source of cyclical patterns is sometimes necessary to produce behavior change. In Figure 9.9, for example, every other data point is higher than the preceding one. Such a pattern could be the result of different experimenters conducting sessions, changes in levels of food deprivation, or perhaps practice effects if two sessions are conducted each day. Cycles may be difficult to detect if there is a good deal of between-session variability, but plotting data in various formats may help reveal cyclical patterns. For example, plotting each data point as a deviation from the mean using a vertical bar graph can make patterns in the variability more apparent (see Morley & Adams, 1991).

The same strategies for evaluating the stability of baseline data are used to evaluate the stability of data in an intervention phase. Figure 9.10 repeats the data from Figure 9.8, but adds data from an intervention phase. In the upper panel, the line of best fit reveals an upward trend in the intervention phase, although the final four data points suggest that the behavior may have asymptoted. The researcher collecting these data should continue the intervention phase to determine whether the performance has reached an asymptote or will increase further given continued exposure to the intervention.

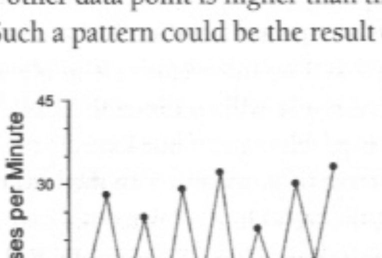

FIGURE 9.9. Example of a figure showing a cyclical pattern. Data are hypothetical.

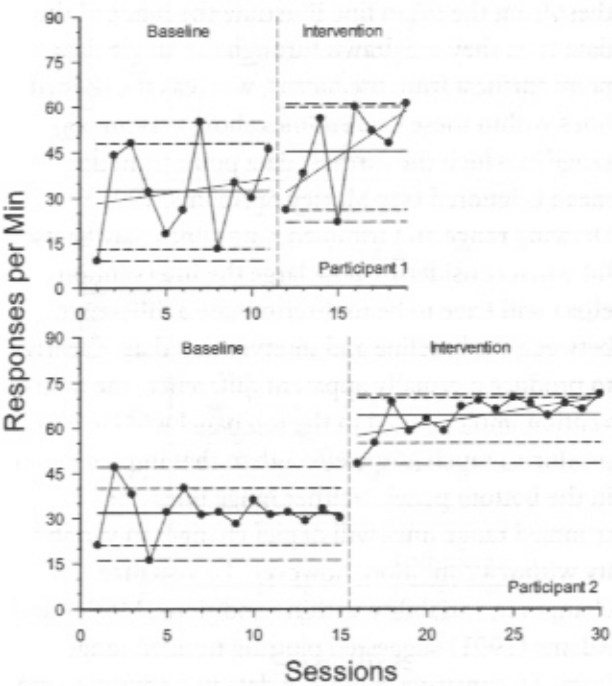

FIGURE 9.10. Hypothetical baseline and intervention data with added mean lines (solid horizontal lines), range lines (long dashed horizontal lines), trimmed range lines (short dashed horizontal lines), and regression lines (solid trend lines). Baseline data are the same as in Figure 9.8.

In the lower panel, a similar upward trend may be observed in the intervention phase, but over the final 10 sessions of the phase, the performance has stabilized because there is little bounce around the mean line and no visually apparent trend.

Evaluating differences across phases. The second use of visual analysis of time-series data involves comparing the baseline and intervention data to determine whether the difference makes a compelling case that behavior has changed between phases. Determining whether behavior change was an effect of the intervention (assuming a compelling difference is observed) is a different matter, and one that we consider in more detail next.

Five characteristics of the data should control the evaluation of behavior change across phases. The first is the change in level. *Level* refers to the immediate change in responding from the end of one phase to the beginning of the next (Kazdin, 1982). Level is assessed by comparing the last data point from a condition to the first data point of the subsequent condition. In the top graph of Figure 9.10, the change in level was a decrease from about 46 responses per minute to about 26 per minute. In the lower panel, the level increased from about 30 to 48 responses per minute. Level may be used to evaluate the magnitude of treatment effect. Large changes in level suggest a potent independent variable, but only when the data collected in the remainder of the intervention phase continue at the new level, as in the lower panel of Figure 9.10. The level change in the upper panel of Figure 9.10 is inconsistent with most of the remaining intervention data and, therefore, appears to be another instance of uncontrolled between-session variability. As this example illustrates, a level change is neither necessary nor sufficient to conclude that behavior changed in the intervention phase.

The second, related characteristic that will affect judgments of treatment effects is latency to change. *Latency to change* is the time required for change in responding to be detected after the onset of a new experimental condition (Kazdin, 1982). To evaluate latency to change, a researcher must examine multiple data points after the condition change to determine whether a consistent change in level or a change in trend occurs (at least three data points are required to detect a trend). A short latency to change indicates that the experimental manipulation produced an immediate effect on behavior, whereas a long latency to change indicates either that an extended exposure to the change in the independent variable is required before behavior changes (such as during extinction) or that the change is caused by an extraneous variable. Again, we consider the question of the causal relation between the behavior change and the intervention later in the chapter.

In the top panel of Figure 9.10, approximately six sessions were required before the trend and mean level in the intervention phase appear distinguishable from baseline. In the lower graph, changes in trend and mean level were observed in the first three sessions after the phase change, showing more clearly that the data in the two phases are distinct. Although short latencies to change suggest that behavior has changed across phases, this change may be temporary and, therefore, additional observations should be made until one is convinced that the change is enduring. How many additional observations are necessary will be affected by factors such as baseline variability (as in the top panel of Figure 9.10) and how novel the finding is (skeptical scientists prefer to have many observations when the intervention is novel). Under most conditions, an intervention that produces a large but temporary behavior change is of limited utility.

The third characteristic of time-series data that is used when visually evaluating differences across phases is the mean shift (Parsonson & Baer, 1992). *Mean shift* refers to the amount by which the means differ across phases. In both panels of Figure 9.10, there is an upward mean shift from baseline to intervention. The bottom graph, however, illustrates a shift that is visually more compelling. The reason for this takes us to the fourth characteristic controlling visual analysis activities: between-phase overlap. In the upper panel of Figure 9.10, as the range lines illustrate, five of eight data points in the intervention condition fall within the range of the preceding baseline data, and, therefore, the difference is not convincing. Perhaps, in the top graph, if additional data were collected during the intervention phase, and assuming responding remained at the upper plateau characterizing the final intervention sessions, the difference might be compelling. In the lower graph of Figure 9.10, the level change, mean shift, and limited between-phase overlap in range make the difference visually apparent.

The fifth characteristic of the data that will affect visual evaluation of between-phase differences is trend. As noted earlier, if the baseline data are trending up (or down) and responding increased (or decreased) during the intervention phase (upper graph of Figure 9.10), then the mean shift will not be convincing unless the trend line is much steeper in the intervention phase than at baseline. In the upper graph of Figure 9.10, the baseline data show a slight upward trend. Data in the subsequent intervention phase show a steeper trend. The greater the difference in trend is, the clearer it is that the mean shift in the intervention is not simply a continuation of the baseline trend.

When evaluating mean shifts, floor or ceiling effects must be considered. These effects occur when performance has reached a minimum or maximum, respectively, beyond which it cannot change further. For example, if baseline response rates are low and an intervention is expected to decrease responding, mean shifts may be small because response rates have little room to further decrease.

Readers skeptical of visual analysis practices may be unsettled by the use of terms and phrases such as *judgment, visually apparent,* and *much steeper.* How much steeper is "much steeper?" Although any interpretation of data requires the researcher make a variety of judgment calls (e.g., which statistic to use, how to handle missing data), Fisher, Kelley, and Lomas (2003) sought to reduce the number of judgments by developing the conservative dual-criterion (CDC) technique to aid the visual analysis of single-case data. This method, illustrated in Figure 9.11, involves extending the baseline mean and trend lines into the intervention phase and raising both of these lines by 0.25 standard deviation (or lowering the lines by 0.25 standard deviation, if the intervention is anticipated to decrease responding). A difference across conditions is judged as meaningful when the number of intervention-phase data points falling above both lines (or below both lines in the case of an intervention designed to decrease a behavior)

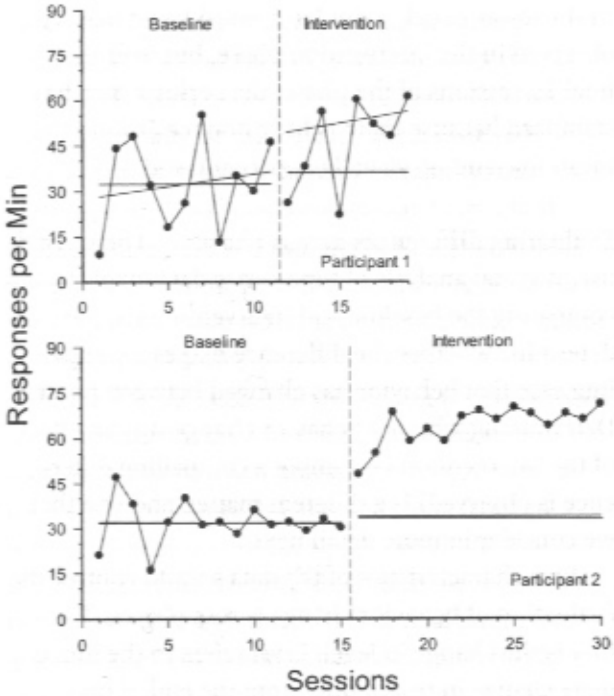

FIGURE 9.11. **Example of the conservative dual-criterion technique applied to hypothetical intervention data. In the baseline phase, solid horizontal lines are mean lines, and solid trend lines are regression lines. To analyze intervention effects, these lines are superimposed onto the intervention phase and raised by 0.25 standard deviation from the baseline means. See text for details.**

exceeds a criterion based on the binomial equation (i.e., exceeds the number that would be expected by chance). Fisher et al. found that the use of CDC procedures improved agreement on visual inspection (data were hypothetical, and intervention effects were computer generated).

Figure 9.11 shows the CDC applied to the data shown in Figure 9.10. In the top panel, three data points in the intervention phase fall above the two criterion lines. Following the table presented in Fisher et al. (2003) for treatment conditions with eight data points, seven data points should be above both lines to conclude that a compelling difference exists between phases. In the lower panel, the CDC requires that 12 of the 15 data points in the intervention condition appear above both lines, a criterion easily met, so the researcher may conclude that behavior changed across phases.

Although the CDC method appears to improve the accuracy of judgments of behavior change, only a few studies have yet investigated this technique (Fisher et al., 2003; Stewart, Carr, Brandt, & McHenry, 2007). The Fisher et al. (2003) procedure is, of course, but one technique for making visual assessment of data more objective. It is ultimately incumbent on the investigator or therapist to provide convincing evidence of an effect, whether through some formal set of rules as illustrated by Fisher et al. or by amplifying the effect to the point at which reasonable people agree on it, through increased control over both independent and extraneous variables.

Assuming that appropriate decisions were made about stability and a visually apparent behavior change was observed from baseline to the intervention phase, the next task is to evaluate the role of the intervention in that behavior change. Evaluating the causal role of the intervention requires that an appropriate experimental design be used, a topic falling under the scope of Chapter 5, this volume. Here, we largely confine our discussion to the visual analysis techniques appropriate to the most commonly used single-case research designs. In these sections of the chapter, the visual analysis focuses on answering the question, "Did the intervention change behavior?"

Comparison designs. The data shown in the lower panel of Figure 9.11 come from a comparison design (or A-B design). There is evidence of a convincing change in behavior across conditions, level and mean level differ, variability and overlap of the data points across conditions are not interfering, and the latency to change is short. Despite stable data, one cannot conclude that the intervention produced the visually apparent behavior change. Although the rapid level change suggests an intervention effect, one cannot rule out extraneous variables that may have changed at the same time that the intervention was introduced (e.g., in addition to the intervention, Participant 2 may have been informed that if his productivity did not improve, his job would be in jeopardy). When visually analyzing data, a difference in behavior between two phases is insufficient evidence that the intervention, and not extraneous variables, produced the change.

Reversal designs. In a reversal design, the experimental variable is introduced and removed, and systematic behavior changes with each manipulation provide evidence for a causal relation. Figure 9.12 shows the previously considered data set now extended to include a second baseline and a second intervention condition. The bottom graph is easily interpreted. The visually apparent difference between the first baseline phase and the first intervention phase is reversed in the return to baseline. In the second baseline phase, responding was well outside the range in which data should have fallen had no experimental manipulation been implemented. Further evidence for an intervention effect is that responding returned to the level observed in the original baseline. The reintroduction of the intervention (fourth phase) reverses the downward trend in the second baseline, yielding a striking level shift, mean shift, and minimal variability. There is very little overlap in the data across conditions, the latency to change is short, and there are no trends that make interpretation difficult. The mean level is close to the mean level obtained in the first exposure to the intervention, thus replicating the effect. These data thus make a strong case for the intervention as an effective means of influencing behavior.

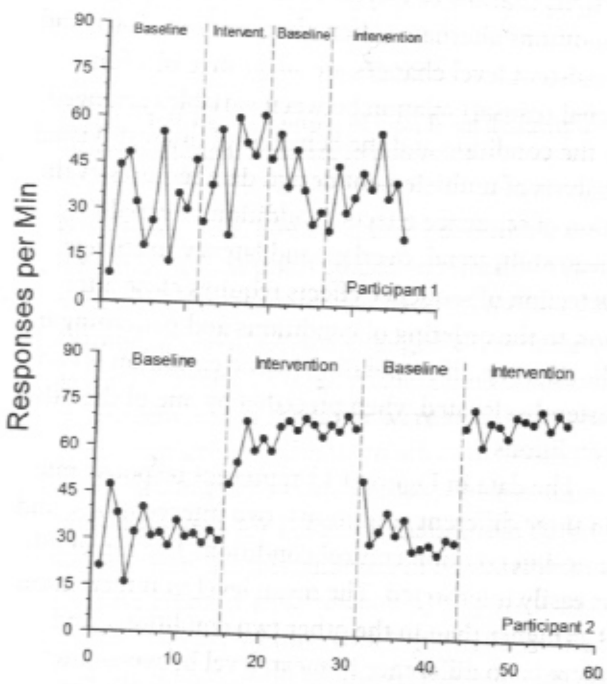

FIGURE 9.12. Example of a reversal design. Data from the first baseline and intervention phases are the same as shown in Figures 9.8 and 9.10. Data are hypothetical.

The upper panel of Figure 9.12 tells a different story. When the baseline conditions are reestablished in the third phase, there is a precipitous downward trend in behavior. Although this behavior change is consistent with the removal of an effective intervention, the between-session variability in the preceding condition renders an unconvincing the argument for a between-phase behavior change. Clearly, the hypothetical researcher who collected these data failed to continue the first intervention phase long enough for a stable pattern of behavior to develop. If responding had stabilized in the upper plateau reached at the end of the first intervention phase, the sharp reduction in responding in the second baseline may have been more compelling. When the intervention is again introduced, the downward trend levels off, but the data points overlap considerably with the data points for the preceding baseline condition. Furthermore, the mean level in the second intervention phase did not closely replicate the mean level of the first intervention phase.

Multielement designs. Figure 9.13 shows data from three hypothetical multielement experimental designs (Barlow & Hayes, 1979). In this design, conditions alternate (often after every session), and consistent level changes are suggestive of a functional (causal) relation between variables arranged in the condition and the behavior of interest. Visual analysis of multielement design data requires evaluation of sequence effects in addition to variability, mean shift, trend, overlap, and latency to change. Detection of sequence effects requires close attention to the ordering of conditions and patterning in the data (i.e., if responding in one condition is consistently elevated when preceded by one of the other conditions).

The data in Figure 9.13 represent response rate in three different conditions, two interventions, and a no-intervention control condition. The top graph is easily interpreted. The mean level in Intervention 1 is higher than in the other two conditions, and there is no difference in mean level between Intervention 2 and the no-intervention condition. The data are relatively stable (i.e., there is little variability), and there are no trends to complicate interpretation. Thus, the difference in behavior between the conditions is obvious. The effects of each experimental manipulation on response rate are reliable (each repetition of a condition allows a test of reliability), and it would be extremely unlikely that some extraneous variable would happen to produce increases in response rate in each Intervention 1 session and none in any other session. Finally, the effect of Intervention 1 does not appear to be dependent on the prior condition, and in at least one case, the effect lasts when two consecutive Intervention 1 sessions are completed. These data provide compelling evidence that Intervention 1 is responsible for producing higher response rates than either Intervention 2 or the no-intervention condition.

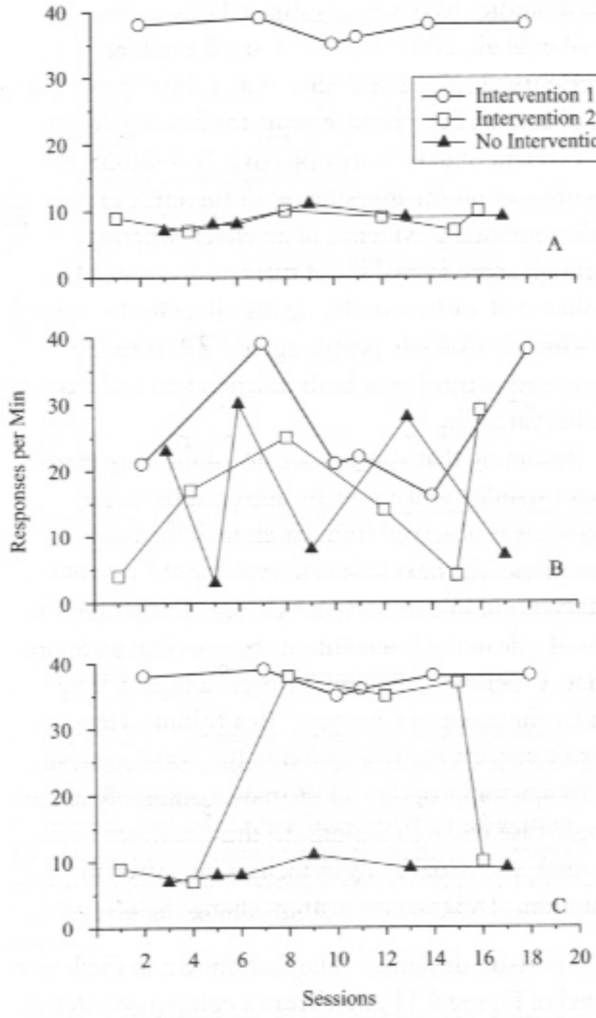

FIGURE 9.13. Examples of data from multielement designs. Data are hypothetical.

The middle graph contains more within-condition variability. The mean level is higher in the

Intervention 1 condition; however, there is considerable overlap in the range of response rates observed in each condition. Because it is not clear that behavior is distinct across conditions, the question of causation is moot. In the third graph, the mean level of the data in Intervention 1 is high, the mean level in the no-intervention condition is low, but the data in Intervention 2 are more difficult to interpret. During some sessions, the response rate is low; during others, it is high. This graph shows a hypothetical sequence effect. Each time an Intervention 2 session follows an Intervention 1 session, the response rate is high; otherwise, the response rate is low. The rate during the Intervention 2 sessions is affected by the preceding condition, which complicates interpretation of the effects of Intervention 2. If the high-rate Intervention 2 sessions were merely a carry-over effect of Intervention 1, then the no-intervention sessions that follow Intervention 1 sessions should show comparable high rates. A researcher who obtains findings of this sort will conduct further experimentation to clarify the processes responsible for the sequence effect.

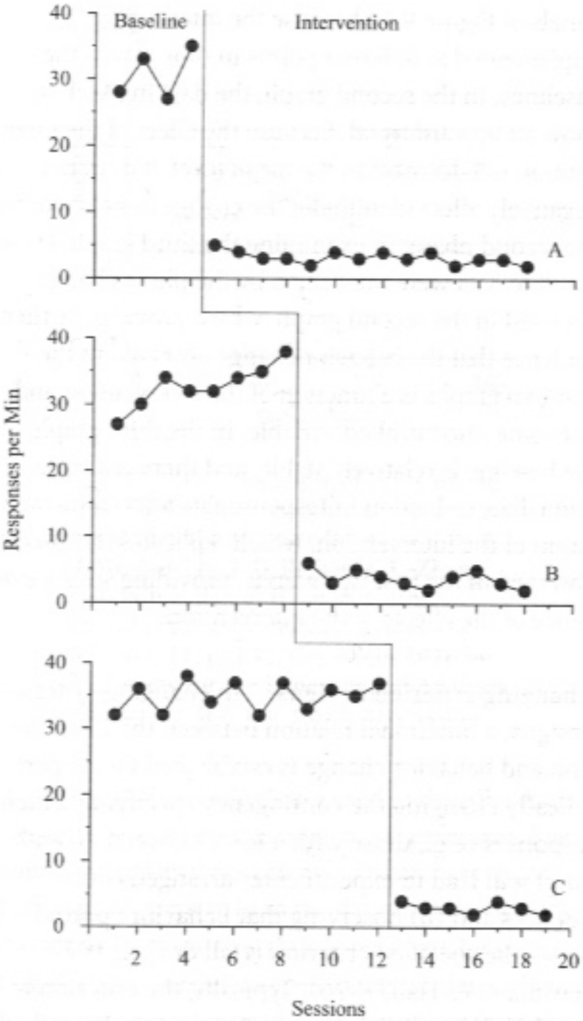

FIGURE 9.14. Example of data from a multiple-baseline design. Data are hypothetical.

Multiple-baseline designs. Multiple-baseline designs are frequently used in applied settings, either when it would be unethical to remove the treatment or because the treatment is expected to produce an irreversible effect. The design involves a series of comparison designs in which the researcher implements the treatment variable at different times across participants, behaviors, or contexts. A researcher visually analyzing data from a multiple-baseline design must evaluate whether there are convincing changes in mean level from baseline to treatment conditions, whether the effects are replicated across baselines, and whether changes in behavior occur only when the treatment is implemented for each baseline.

Figure 9.14 illustrates data from a multiple-baseline design. In the top panel, a brief baseline precedes the intervention. The intervention produces level changes and mean shifts easily discriminated as behavior change. There is no latency to change, and there are no trends or overlap between data points across conditions to complicate data interpretation. As noted in the Comparison Designs section, however, these data alone are insufficient to support causal statements. The behavior change could also have been caused by an extraneous variable that changed at the same time as the intervention (e.g., a change in classroom contingencies). If the latter were true, then one might expect this variable to affect behavior in the other baselines. To evaluate this, one examines the other baselines for behavior change that corresponds with the introduction of the intervention in the first graph (i.e., at Session 5). Figure 9.14 shows evidence of this, which strengthens the case that the intervention produced the behavior change observed in the top panel.

Further evidence that the intervention is related to the behavior change must be gathered in the remaining panels of Figure 9.14 because the intervention is implemented at different points in time across the baselines. In the second graph, the data in baseline show an upward trend. Because the effect of the intervention is a decrease in the mean level, it does not negatively affect identifying the change in behavior in the second phase. In examining the third graph, these baseline data were unaffected by the phase change depicted in the second graph, which provides further evidence that the behavior change observed in the first two graphs is a function of the intervention and not some uncontrolled variable. In the third graph, the baseline is relatively stable, and there is a large, immediate reduction in response rate after implementation of the intervention, which replicates the effects observed in the first two graphs, providing strong evidence of the effects of the intervention.

Changing-criterion designs. In changing-criterion designs, a functional relation between the intervention and behavior change is established by (a) periodically changing the contingency specifying which responses (e.g., those with a force between 20 and 30 g) will lead to experimenter-arranged consequences and (b) observing that behavior approximates the specified criterion (Hall & Fox, 1977; Hartmann & Hall, 1976). Typically, the criterion in graphs of changing-criterion designs is indicated by horizontal lines at each phase. Visual analysis of changing-criterion designs, as with that of other designs, requires an assessment of variability, level, mean shift, trend, overlap, and latency to change but also requires an assessment of the relation between behavior and the criterion.

Figure 9.15 shows data from Hartmann and Hall (1976), who used a changing-criterion design to assess the effectiveness of a smoking cessation program. During the intervention, the participant was fined a small amount of money for smoking above a criterion number of cigarettes and earned a small amount of money for smoking fewer cigarettes. In the top graph, the number of cigarettes smoked per day is shown across successive days of the intervention. In baseline, the number of cigarettes smoked per day was stable over the first 6 days but fell precipitously on Day 7. Ideally, the researchers would not have begun the intervention on Day 8, as they did, because if a trend line was drawn through these baseline data, the subsequent decreases in smoking would be predicted to occur in the absence of an intervention. Had the researchers collected more baseline data, they would likely have found that Day 7 was uncharacteristic of this individual's rate of smoking and could have more clearly established the stable rate of baseline smoking.

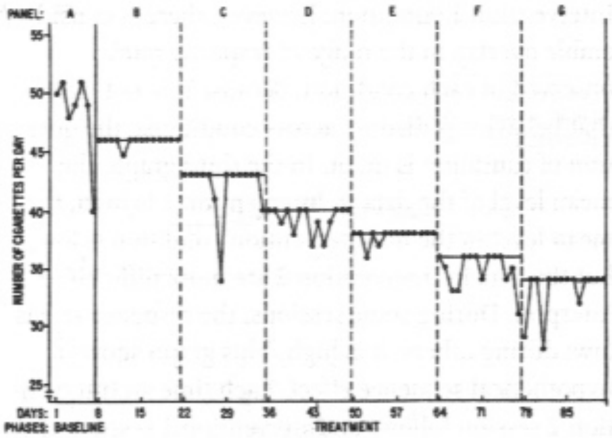

FIGURE 9.15. Example of data from a changing-criterion design. The figure shows the number of cigarettes smoked per day. Solid horizontal lines depict the criterion for each phase. From "The Changing Criterion Design," by D. P. Hartmann and R. V. Hall, 1976, *Journal of Applied Behavior Analysis*, *9*, p. 529. Copyright 1976 by the Society for the Experimental Analysis of Behavior, Inc. Used with permission.

In subsequent phases (B–G), the criterion number of cigarettes was systematically decreased, as indicated by the horizontal line in each phase. Changes in the criterion tended to produce level changes, with many subsequent data points falling exactly on the criterion specified in that phase. Each phase establishes a new mean level approximating the criterion. There are no long latencies to change, and the variability in the data and overlap in data points across conditions are not sufficient to cause concern. Finally, with the exception of Phase F, there is no downward trend in any phase, suggesting that if the criterion remained unchanged, smoking would remain at the current depicted level. Thus, the visual analysis of these data raises concerns about the downward trend in the baseline, but these concerns are largely addressed by the repeated demonstrations of control over smoking rate in each condition. If the study were ongoing and concerns remained, the researchers could set the next criterion (Phase H) above the last one. If smoking rate increased to the new criterion, then additional evidence for intervention control would be established while nullifying concerns about the downward trend in baseline.

Interpreting Relational Graphs

Researchers who conduct time-series research may report their outcomes using relational graphs. In these cases, each data point represents the mean (or another appropriate measure of central tendency) of steady-state responding from a condition. When evaluating these data, measures of variability, such as error bars, are also assessed to help determine whether responding was stable (see Interpreting Bar Graphs section).

Data on relational graphs are evaluated by analyzing the clustering and trend of the data points. Data that appear horizontal across all values of the *x*-axis indicate that the independent or predictor variable has no effect on behavior or that there is no correlation between the two dependent variables. Sometimes behavior changes in a linear fashion across the range of *x*-axis values of the independent variable. When both axes of the graph are scaled linearly, a linear relation indicates that changing the independent variable produces a constant increase or decrease in behavior. Nonlinear relations indicate that the behavioral effect of the independent variable changes across *x*-axis values. Figure 9.16 provides an example. Here, the subjective value of a $10 reward is plotted as a function of the delay to its delivery. Both axes are linear, and the relation between the variables is nonlinear.

Relational graphs, when properly constructed, allow the researcher to quickly evaluate the relation between variables. It is common, however, to evaluate relational data more precisely with quantitative methods, including curve-fitting techniques (e.g., linear and nonlinear regression), Pearson's correlation coefficient, and quantitative models (see Chapters 10 and 12, this volume). Curve-fitting techniques clarify the form of the relation between the independent and dependent variables, and Pearson's correlation coefficient quantifies the relation between two dependent variables. Quantitative models may describe more complex behavior–environment relations and are used to make predictions about behavior. Even

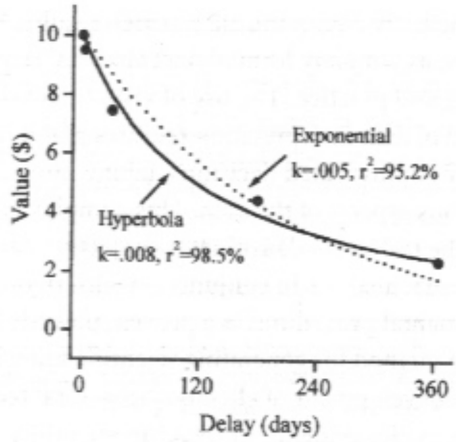

FIGURE 9.16. Example of curve fitting. From "Delay or Probability Discounting in a Model of Impulsive Behavior: Effect of Alcohol," by J. B. Richards, L. Zhang, S. H. Mitchell, and H. de Wit, 1999, *Journal of the Experimental Analysis of Behavior, 71*, p. 132. Copyright 1976 by the Society for the Experimental Analysis of Behavior, Inc. Used with permission.

when quantitative methods are used to describe data, however, visual analysis is used as a supplement. For example, visual analysis can help researchers choose which type of curve to fit to the data, evaluate whether data trends are linear and thus appropriate for calculating Pearson correlation coefficients, or determine whether the data have systematic deviations from the fit of a quantitative model. For example, Figure 9.16 shows the best fits of both exponential and hyperbolic models to the subjective value of delayed money. The figure shows that the exponential model systematically predicts a lower *y*-axis value than that obtained at the highest *x*-axis value.

Reference lines may be added to relational graphs (or other graph types) to provide a point of comparison to the data. For instance, in a graph showing discrete-trial performances, such as matching-to-sample, reference lines may be plotted at values expected by chance. In a graph depicting choice data, reference lines might be plotted at values indicative of indifference.

CONCLUSION

Graphs provide clear and detailed summaries of research findings that can guide scientific decisions and efficiently communicate research results. Visual analysis, as with any form of data analysis, requires training and practice. The use of visual analysis as a method of data interpretation requires graph readers to make sophisticated decisions, taking into account numerous aspects of the data. This complexity can make the task seem daunting or subjective; however, visual analysis in conjunction with rigorous experimental procedures is a proven, powerful, and flexible method for generating scientific knowledge.

The development of effective behavioral technologies provides evidence of the ultimate utility of the visual analysis techniques used in behavior-analytic research. Data analyzed by means of visual inspection have contributed to a technology that produces meaningful behavior change in individuals across a wide range of skill domains and populations, including individuals with no diagnoses and those with diagnoses including attention deficit/hyperactivity disorder, autism, an array of developmental disabilities, pediatric feeding disorders, and schizophrenia, to name a few (Didden, Duker, & Korzilius, 1997; Lundervold & Bourland, 1988; Weisz, Weiss, Han, Granger, & Morton, 1995). Because of its history of effective application and advantages for the study of the behavior of individuals, behavior analysts remain committed to visual inspection as a primary method of data analysis.

References

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Barlow, D. H., & Hayes, S. C. (1979). Alternating treatments design: One strategy for comparing the effects of two treatments in a single subject. *Journal of Applied Behavior Analysis, 12*, 199–210. doi:10.1901/jaba.1979.12-199

Baron, A., & Perone, M. (1998). Experimental design and analysis in the laboratory study of human operant behavior. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 45–91). New York, NY: Plenum Press.

Baum, W. M. (1973). The correlation-based law of effect. *Journal of the Experimental Analysis of Behavior, 20*, 137–153. doi:10.1901/jeab.1973.20-137

Cançado, C. R. X., & Lattal, K. A. (2011). Resurgence of temporal patterns of responding. *Journal of the Experimental Analysis of Behavior, 95*, 271–287.

Cleveland, W. S. (1994). *The elements of graphing data* (rev. ed.). Summit, NJ: Hobart Press.

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association, 79*, 531–554. doi:10.2307/2288400

Cleveland, W. S., & McGill, R. (1985). Graphical perception and graphical methods for analyzing scientific data. *Science, 229*, 828–833. doi:10.1126/science.229.4716.828

Cooper, J. O., Heron, T. E., & Heward, W. L. (1987). *Applied behavior analysis*. Columbus, OH: Merrill.

Davison, M., & Baum, W. M. (2006). Do conditional reinforcers count? *Journal of the Experimental Analysis of Behavior, 86*, 269–283. doi:10.1901/jeab.2006.56-05

DeFulio, A., Donlin, W. D., Wong, C. J., & Silverman, K. (2009). Employment-based abstinence reinforcement as a maintenance intervention for the treatment of cocaine dependence: A randomized controlled trial. *Addiction, 104*, 1530–1538. doi:10.1111/j.1360-0443.2009.02657.x

Didden, R., Duker, P. C., & Korzilius, H. (1997). Meta-analytic study on treatment effectiveness for problem behaviors with individuals who have mental retardation. *American Journal on Mental Retardation, 101*, 387–399.

Ferster, C. B., & Skinner, B. F. (1997). *Schedules of reinforcement*. Acton, MA: Copley. (Original work published 1957) doi:10.1037/10627-000

Fisher, W. W., Kelley, M. E., & Lomas, J. E. (2003). Visual aids and structured criteria for improving inspection and interpretation of single-case designs. *Journal of Applied Behavior Analysis, 36*, 387–406. doi:10.1901/jaba.2003.36-387

Galbicka, G., Fowler, K. P., & Ritch, Z. J. (1991). Control over response number by a targeted percentile schedule: Reinforcement loss and the acute effects of d-amphetamine. *Journal of the Experimental Analysis of Behavior, 56*, 205–215. doi:10.1901/jeab.1991.56-205

Gallistel, C. R., King, A. P., Gottlieb, D., Balci, F., Papachristos, E. B., Szalecki, M., & Carnone, K. S. (2007). Is matching innate? *Journal of the Experimental Analysis of Behavior, 87*, 161–199. doi:10.1901/jeab.2007.92-05

Gentry, G. D., Weiss, B., & Laties, V. G. (1983). The microanalysis of fixed-interval responding. *Journal of the Experimental Analysis of Behavior, 39*, 327–343. doi:10.1901/jeab.1983.39-327

Hall, R. V., & Fox, R. G. (1977). Changing-criterion designs: An alternative applied behavior analysis procedure. In B. C. Etzel, J. M. LeBlanc, & D. M. Baer (Eds.), *New developments in behavioral research:*

Theory, method, and application (pp. 151–166). Hillsdale, NJ: Erlbaum.

Harris, R. L. (1996). *Information graphics: A comprehensive illustrated reference.* Atlanta, GA: Management Graphics.

Hartmann, D. P., & Hall, R. V. (1976). The changing criterion design. *Journal of Applied Behavior Analysis, 9*, 527–532. doi:10.1901/jaba.1976.9-527

Johnston, J. M., & Pennypacker, H. S. (1993). *Strategies and tactics of behavioral research* (2nd ed.). Hillsdale, NJ: Erlbaum.

Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings.* New York, NY: Oxford University Press.

Lundervold, D., & Bourland, G. (1988). Quantitative analysis of treatment of aggression, self-injury, and property destruction. *Behavior Modification, 12*, 590–617. doi:10.1177/01454455880124006

Morley, S., & Adams, N. I. (1991). Graphical analysis of single-case time series data. *British Journal of Clinical Psychology, 30*, 97–115. doi:10.1111/j.2044-8260.1991.tb00926.x

Parsonson, B. S., & Baer, D. M. (1978). The analysis and presentation of graphic data. In T. R. Kratochwill (Ed.), *Single-subject research: Strategies for evaluating change* (pp. 101–165). New York, NY: Academic Press.

Parsonson, B. S., & Baer, D. M. (1986). The graphic analysis of data. In A. Poling & R. W. Fuqua (Eds.), *Research methods in applied behavior analysis: Issues and advances* (pp. 157–186). New York, NY: Plenum Press.

Parsonson, B. S., & Baer, D. M. (1992). The visual analysis of data, and current research into the stimuli controlling it. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp. 15–40). Hillsdale, NJ: Erlbaum.

Perone, M. (1991). Experimental design in the analysis of free-operant behavior. In I. H. Iversen & K. A. Lattal (Eds.), *Experimental analysis of behavior, Parts 1 and 2. Techniques in the behavioral and neural sciences* (Vol. 6, pp. 135–171). New York, NY: Elsevier.

Richards, J. B., Zhang, L., Mitchell, S. H., & de Wit, H. (1999). Delay or probability discounting in a model of impulsive behavior: Effect of alcohol. *Journal of the Experimental Analysis of Behavior, 71*, 121–143. doi:10.1901/jeab.1999.71-121

Shah, P., Freedman, E. C., & Vekiri, I. (2005). The comprehension of quantitative information in graphical displays. In P. Shah & A. Miyake (Eds.), *The Cambridge handbook of visuospatial thinking* (pp. 426–476). New York, NY: Cambridge University Press.

Shull, R. L. (1991). Mathematical description of operant behavior: An introduction. In I. H. Iversen & K. A. Lattal (Eds.), *Experimental analysis of behavior, Parts 1 and 2. Techniques in the behavioral and neural sciences* (Vol. 6, pp. 243–282). New York, NY: Elsevier.

Sidman, M. (1960). *Tactics of scientific research.* Oxford, England: Basic Books.

Staddon, J. E. R., & Simmelhag, V. L. (1971). The "superstition" experiment: A reexamination of its implications for the principles of adaptive behavior. *Psychological Review, 78*, 3–43. doi:10.1037/h0030305

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*, 677–680. doi:10.1126/science.103.2684.677

Stewart, K. K., Carr, J. E., Brandt, C. W., & McHenry, M. M. (2007). An evaluation of the conservative dual-criterion method for teaching university students to visually inspect AB-design graphs. *Journal of Applied Behavior Analysis, 40*, 713–718.

Tufte, E. R. (1983). *The visual display of quantitative information.* Cheshire, CT: Graphics Press.

Weisz, J. R., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin, 117*, 450–468. doi:10.1037/0033-2909.117.3.450