A   Exponential growth rates and curvature

Let us consider a simple equation which we could use to simulate data which grows exponentially in x, with optional error term ϵ:

$$y = \exp\{\beta_1 x + \epsilon\} \quad \text{where} \quad \epsilon \sim N(0, \sigma). \tag{A1}$$
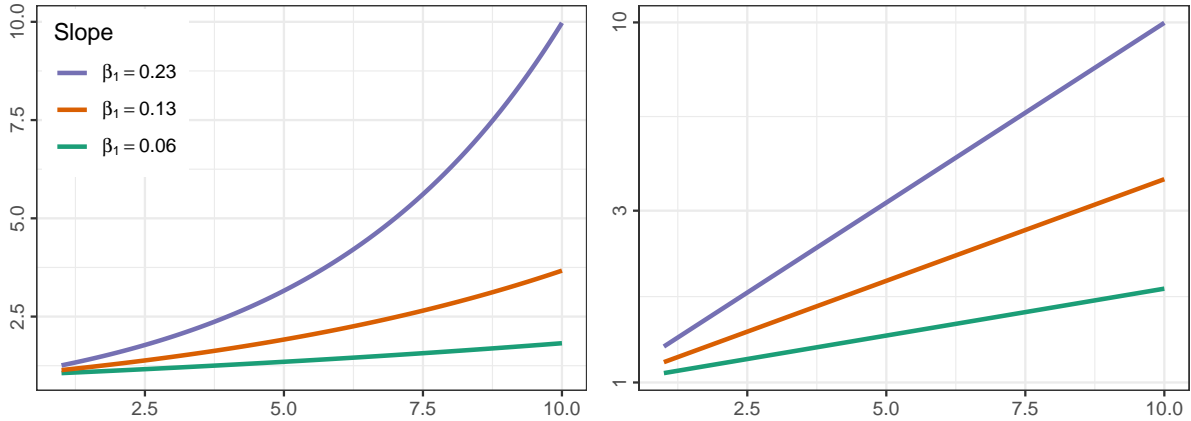

Figure A1: Linear and log scale simple exponential equations as specified in Eq. (A1), with error term excluded. The y-axis endpoints are a primary visual signal when differentiating between the lines in both plots.

It is important in lineup studies to control for extraneous visual signals, and when other visual signals creep in, results of the study can be difficult to interpret [1].

In this particular case, the extraneous visual signals are the endpoints of the lines (the endpoint at $x = 10$ on a linear scale, and both endpoints on the log scale), as shown in Fig. A1. Preattentive features guide attention [2]; one of the first things we do when scanning a graph is to notice the extent of the data within the axes. We must control the endpoints to ensure that participants are using active attention to assess the lineup and draw conclusions, so that the whole visual signal is processed as part of the lineup evalua-tion; if participants are making decisions off of the endpoints, then we cannot interpret the
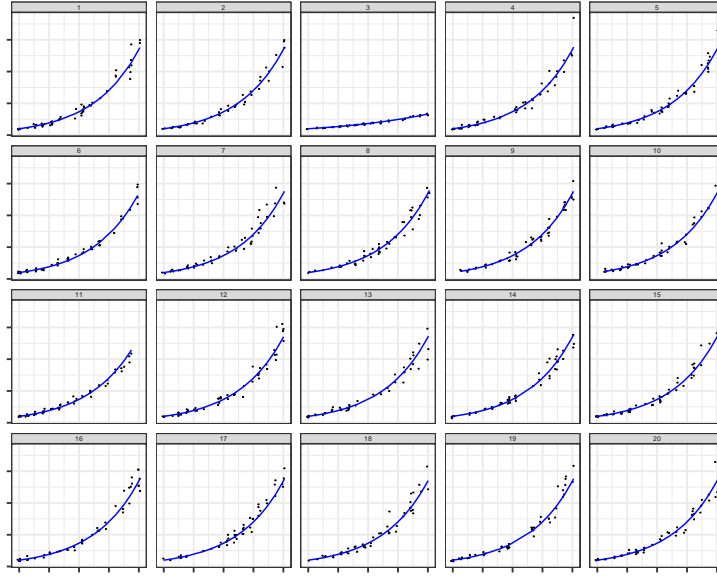
Figure A2: When axis limits are fixed to the most expansive extent in the generated data, the primary signal becomes the extent of the domain which has data, rather than the growth rate itself.

results to say that they are assessing the rate of exponential growth rather than the end result.

There are several different options for controlling the visual signal in a lineup:

0. Do nothing and set each subplot's limits to the overall maximum limits.

1. allow each subplot to have different axis limits

2. truncate the displayed range of each subplot to the minimum range generated in the lineup

3. add extra parameters to the exponential equation to adjust the range of the data so that it fits within a pre-specified domain, as in Algorithm 2.

Typically, lineups do not show axis labels or scaling, because the goal is to assess the signal from the data without additional context. In addition, the traditional 20-panel lineup would be quite visually confusing in this circumstance:
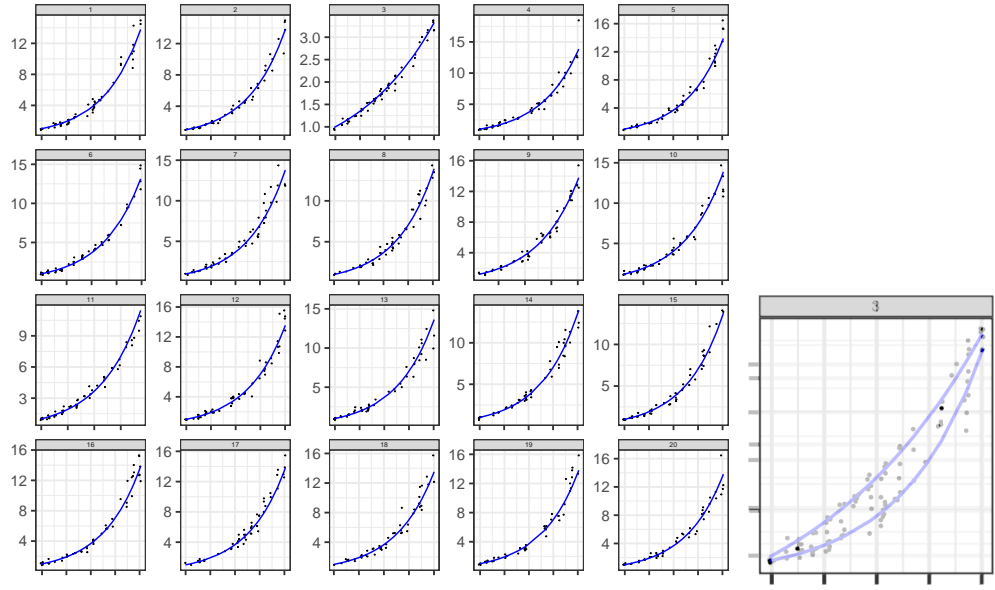
(a) Lineup

(b) Overlay of null and signal panels.

Figure A3: When y-axis limits vary by panel, the y-axis labels are the primary visual signal; the secondary signal (when the panels are overlaid) is the curvature of the line.

Fig. A3 demonstrates that when we ignore the y-axis labels and overlay the signal panel with one of the null panels, the remaining difference is the curvature of the line. If we want to keep the standard convention of not including y-axis labels in lineups for simplicity and to reduce plot clutter, this alternative signal seems promising (and as we will show, is approximately equivalent to option 3).

Let us next consider option 2: Crop each panel to the minimum limits of all generated data. The result is shown in Fig. A4. This operation shifts which axis becomes the visually important factor, but doesn't change the problem: previously the issue was the y-axis extent, now it is the x-axis extent. Both of these parameters are implicitly affected by how we generate exponential data. In order to truly assess whether people can discriminate between comparable exponential growth rates graphically, it is more useful to approach the problem from a curvature perspective rather than to artificially limit the data shown.

This issue is a fundamental problem when testing graphics: the test must meet the "goldilocks" standard - not too hard, not too easy, but just right. Both option 0 and option 2 fail this standard.

Visually, both the x and y axes matter equally, even if it might on paper make sense to truncate x in order to control the y axis range.

An additional benefit of controlling endpoints is that it also provides some realism: our goal was to assess the ability to examine exponential growth **rates**, motivated by the COVID-19 pandemic. As each geographic reporting region started with different numbers of cases and had different control policies, the α and θ parameters are reasonable to represent some of these differences while still examining the underlying exponential behavior.

To control these endpoints on both scales then we have to move to an exponential model
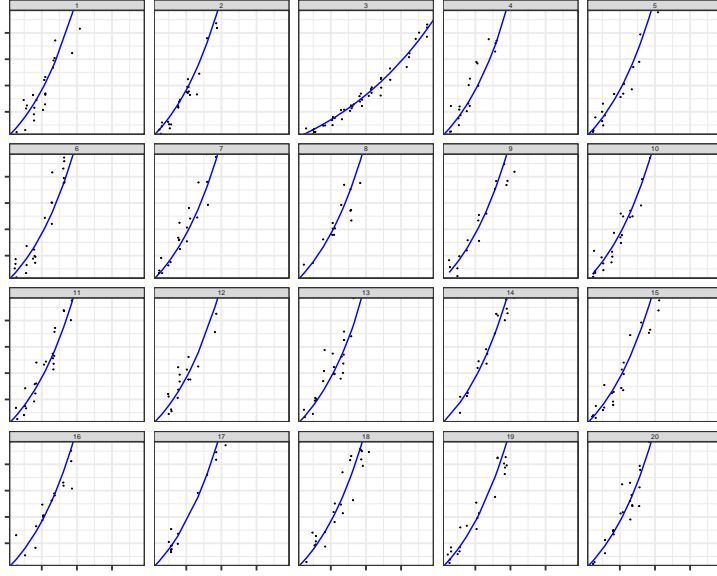
Figure A4: When axis limits are fixed to the most restrictive in all generated data, the primary signal becomes the extent of the x axis which is covered with data, rather than the growth rate itself.

that is a bit more complicated:

$$y = \alpha \cdot \exp\{\beta_1 x + \epsilon\} + \theta \quad \text{where} \quad \epsilon \sim N(0, \sigma). \tag{A2}$$

α and θ were used to make endpoints consistent among the growth rates. As a result, the lineups examine the degree of inflection of the trend rather than the explicit exponential growth rate. An explanation for how the ultimate values for α and θ were determined is provided in Algorithm 2. Fig. A5 uses the parameters in Table 1 and the data generating method described in Algorithm 2. The result is a series of panels which have obvious variation due to the points (a desirable feature) but where the underlying relationship shown in blue has consistent endpoints. As a result, the question becomes whether participants can identify that plot 3 has a different growth rate (as measured by the curvature of the line).
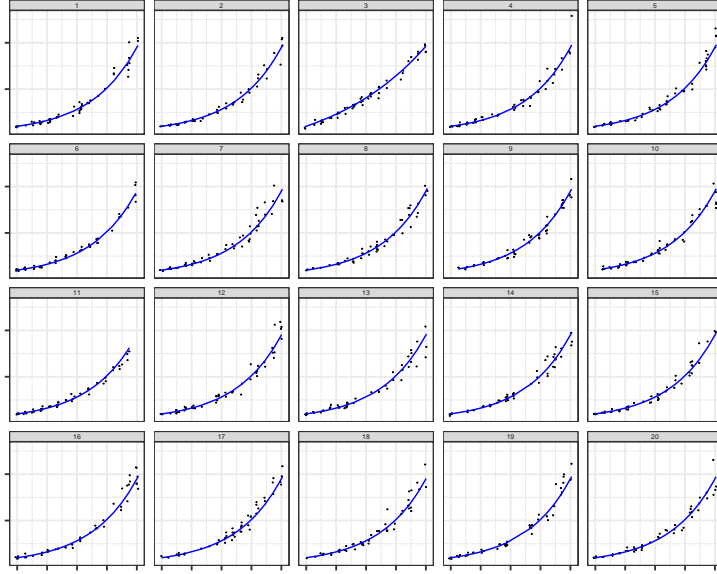
Figure A5: When the data are scaled such that endpoints are similar across all conditions, the primary feature becomes the curvature of the line, while individual plots show random variation due to the scatter around the line.

B Rorschach Lineup Evaluation

In addition to the experimental lineups, we generated three sets of homogeneous "Rorschach" lineups, each featuring panels simulated from the same distribution. Importantly, participants were unaware that there was no designated target panel for identification. Fig. B6 illustrates the selection proportions for each panel on the "Rorschach" lineups corresponding to different curvature combinations and replicated data sets, when presented on both log and linear scales. Notably, the panel number is irrelevant, and the darker shade (bottom) denotes panels chosen more frequently than others.

This approach allows us to explore whether any null panels were significantly more unique, providing insights into the null plot sampling method. The plots exhibit a relatively even distribution, offering multiple candidates for the most interesting panel. It is worth noting that in the low curvature "Rorschach" lineups simulated in data set replication one,
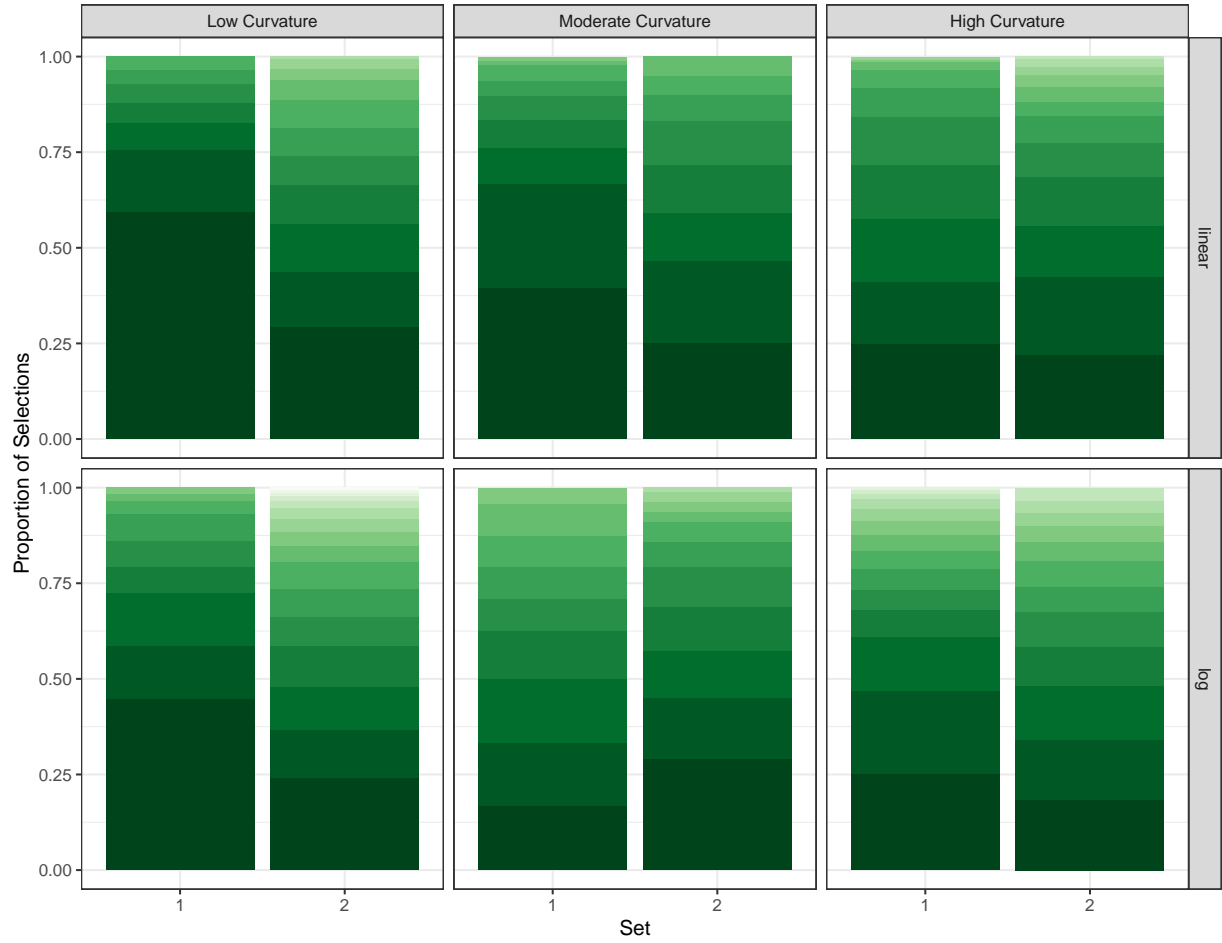
Figure B6: Observed proportions of panel selections for the 'Rorschach' lineups. The darker shade (bottom) denotes panels chosen more frequently than others.

a single panel stands out, with over 50% of participants selecting the same null panel on the linear scale and just under 50% on the log scale. Through a visual examination of "Rorschach" evaluations, we determined that our null sampling method is appropriate.

C   Participant confidence and accuracy

In addition to investigating how scale and curvature rate influence participant accuracy, we asked participants to indicate their level of confidence on a five-point Likert scale by answering, "How certain are you?" Fig. C7 shows the observed confidence rating proportions
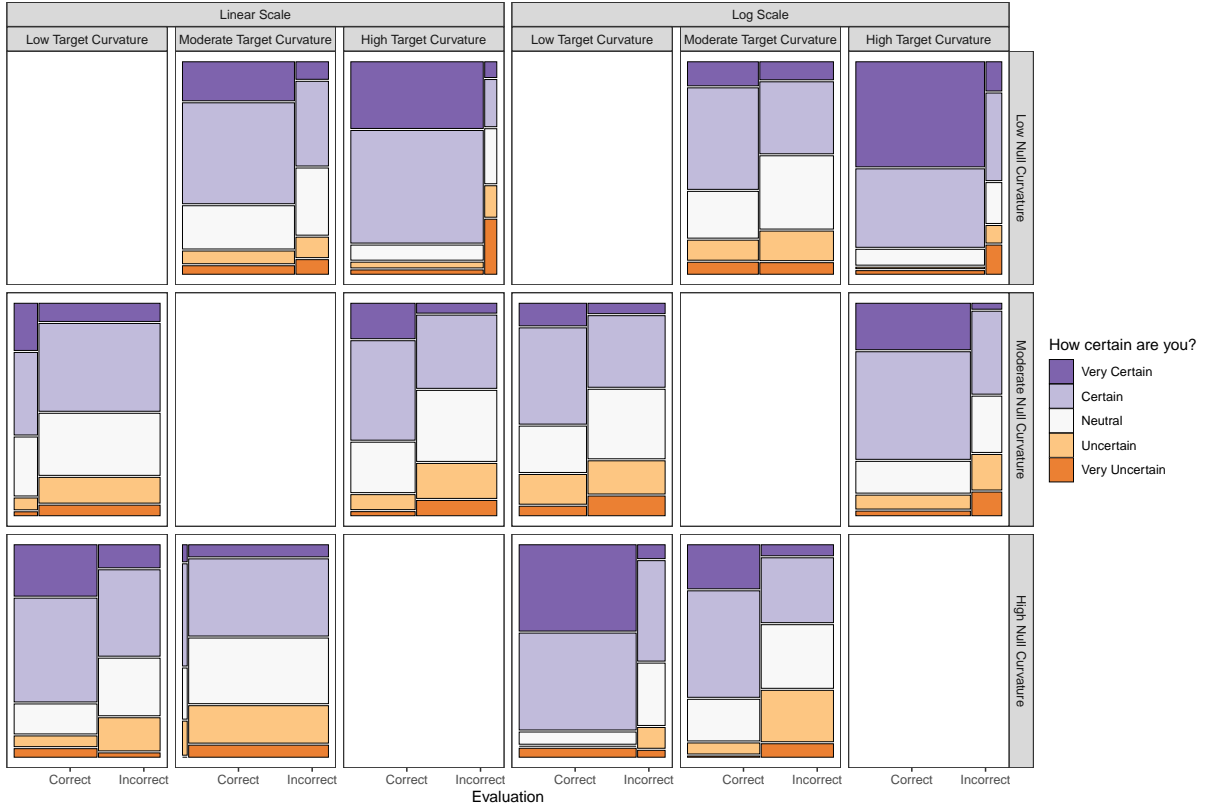
Figure C7: Observed proportion of evaluations for each confidence rating (five-point Likert scale) by accuracy. The width of the bars indicates the number of evaluations for correct and incorrect identifications within each curvature combination.

for correct and incorrect identifications across the different scale and curvature combination scenarios. Overall, participants who correctly identified the target plot were more confident in their plot choice across all conditions with the highest confidence levels and accuracy in scenarios with large differences in curvature between the target and null panels.

References

[1] Susan VanderPlas and Heike Hofmann. Clusters beat trend!? testing feature hierarchy in statistical graphics. *Journal of Computational and Graphical Statistics*, 26(2):231–242, 2017.

[2] Jeremy M Wolfe and Igor S Utochkin. What is a preattentive feature? 29:19–26.