Do gridlines on a line graph correct for compatibility effects in quantitative estimates of change?

Misha Ash

5/3/2018

introduction

This experiment aims to elicit a cognitive bias that derives from a mental metaphor that maps the dimension of emotional valence onto a vertical spatial dimension (e.g., see Dolscheid & Casasanto, 2014 and 2015). A mental metaphor like this may be deployed by representing the values along a non-spatial dimension ranging from negative emotions like sadness to positive emotions like happiness in terms of positions along a vertical spatial dimension ranging from low to high. It is expected that quantitative estimates made from line graphs will be affected by this bias such that incompatible graphs—that is, graphs representing a decrease in a positively valenced phenomenon like happiness or an increase in a negatively valenced phenomenon like sadness—will inflate error rates relative to compatible graphs.

- *null hypothesis 1*: The mean error will not differ between compatible and incompatible conditions.
- *alternative hypothesis 1*: The mean error will be lower for compatible than incompatible conditions.

As reported by Heer and Bostock (2010), gridlines can improve accuracy when not spaced too tightly. To the extent that accuracy is affected by compatibility effects, then, this bias may be mitigated by gridlines.

- *null hypothesis 2*: The mean error will not differ between conditions with and without an additional gridline.
- *alternative hypothesis 2*: The mean error will be lower in the condition with additional gridlines.

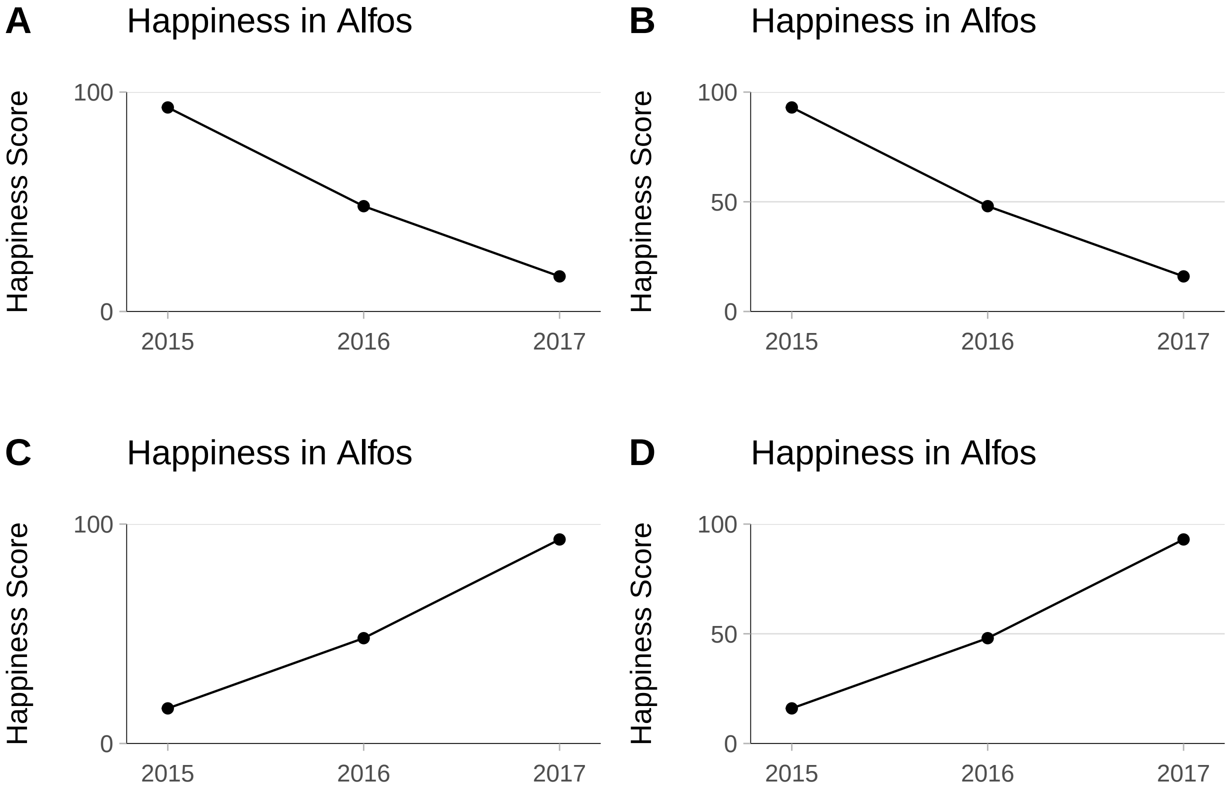
design & methods

The study is organized around a 2x2x2x3 design:

- valence: happiness (positive), sadness (negative)
- direction: increasing, decreasing
- grid lines: 0 and 1 (mid-line added at 50)
- city name: Oswa, Escor, Aflos

A set of 24 stimuli graphs is generated by crossing these factors (8 x 3 for each city):

Valence and direction are crossed for two compatible conditions (increasing happiness and decreasing sadness) and two incompatible conditions (decreasing happiness and increasing sadness). All stimulus graphs have labels for the minimum and maximum of the scale (0 to 100), and the half in the gridline present condition have an additional label at 50 with a horizontal grid line.



method

Each visualization displays the same three data points in ascending or descending order. The data value triplet used across the graphs is randomly sampled from its respective third of the 0 - 100 range ±5 (to keep the value from being directly on a grid line and force estimation). The method used by the base R sampling function is analogous to Cleveland and McGill's use of a "uniform random number generator" (1984, p. 539). Resulting stimulus graphs were saved in a .png format via ggsave with uniform dimensions (width = 7, height = 5) and imported to Qualtrics with a fixed aspect ratio and 750 px width.

The framing of each graph was used to elicit compatibility effects by activating a valenced interpretation of what the quantities represent. Prior to being presented with the target stimulus, participants read the following instructions:

> "The next screen will display a graph of annual [happiness/sadness] scores for the city of [Oswa/Escor/Aflos] over the last three years. The graph will be shown for 20 seconds. Study it carefully during this time. You will then be asked a question about this graph. Please answer the question as quickly but accurately as possible. When you are ready, proceed to view the graph."

To increase the possibility of seeing an effect of valence, fictional city names are used as a control variable. People vary in the valence and strength of their associations with concepts represented by words. For example, the name "Ludwig" might trigger strong positive feelings for someone who has fond memories of positive experiences with someone named Ludwig, which may affect how the happiness or sadness of Ludwig is interpreted. Alternatively, someone who has never known a Ludwig will likely be much more neutral in their affective response to this name, or their affective response may be colored by associations activated by resemblence with similar sounding words, such as "wig," or by their memories of the music of Ludwig van Beethoven or the arguments of Ludwig Wittgenstein. To mitigate and control for world knowledge effects on valence, the graphs will be presented as being about change in an annual happiness or sadness score in a fictional city (participants are not informed about the fictive status of the city and the data). Three city names were generated on the Fantasy Name Generators website (each two syllables; aforementioned).

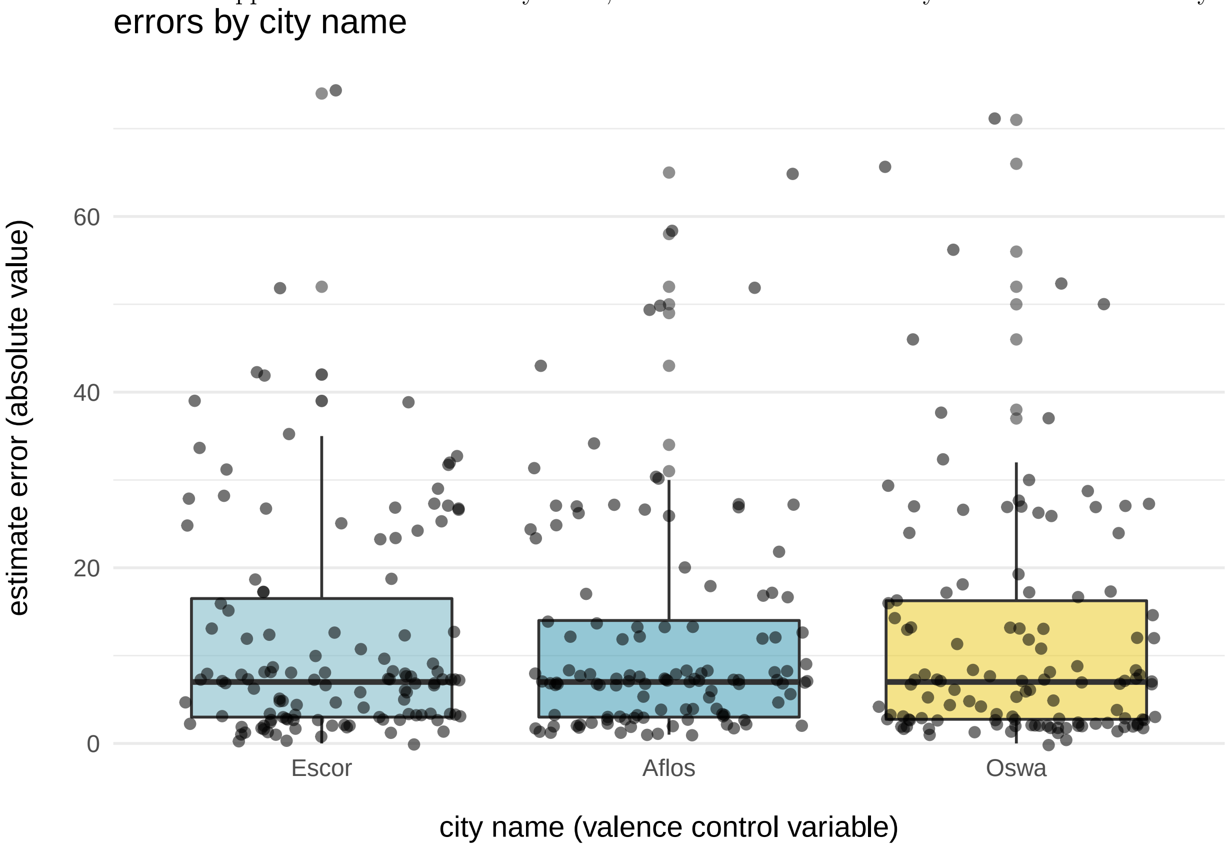
The radomization logic of the stimulus conditions was implemented in Qualtrics (see survey flow in repository materials for more information). After being presented with the stimulus graph for 20 seconds (proceeding prior to timer elapsing was disabled and occurred automatically once timer elapsed), participants made an an estimate of the change between the first and last year plotted by using a slider. The survey began with a multiple-choice qualification task and terminated if the final question was answered incorrectly, which probed participants' understanding of the task by asking about the difference between minimum and maximum values on a graph with 3 gridlines and all points plotted directly on gridlines.

Given the exploratory nature of the current experiment, a one-shot, between-subjects design is used. The study will consist of one HIT with one of 24 possible conditions, randomly assigned and counter-balanced. The study aims to obtain 480 observations, all from distinct participants (ballot box stuffing was disabled in Qualtrics to mitigate multiple responses from the same participant). Although within-subjects designs can increase power for a given sample size, the comparability (and power) gained would require either using the same values for stimuli across conditions, which is prone to order effects and would require extensive counter-balancing of an already complex design or different values across conditions that vary between participants to converge on comparable means. This would require a larger sample, a larger stimulus set, or both.

results

Based on participant ID, observations from 361 Turkers were included in the analysis. (Disqualified or otherwised unfinished submissions were excluded.) All participants were English speakers.

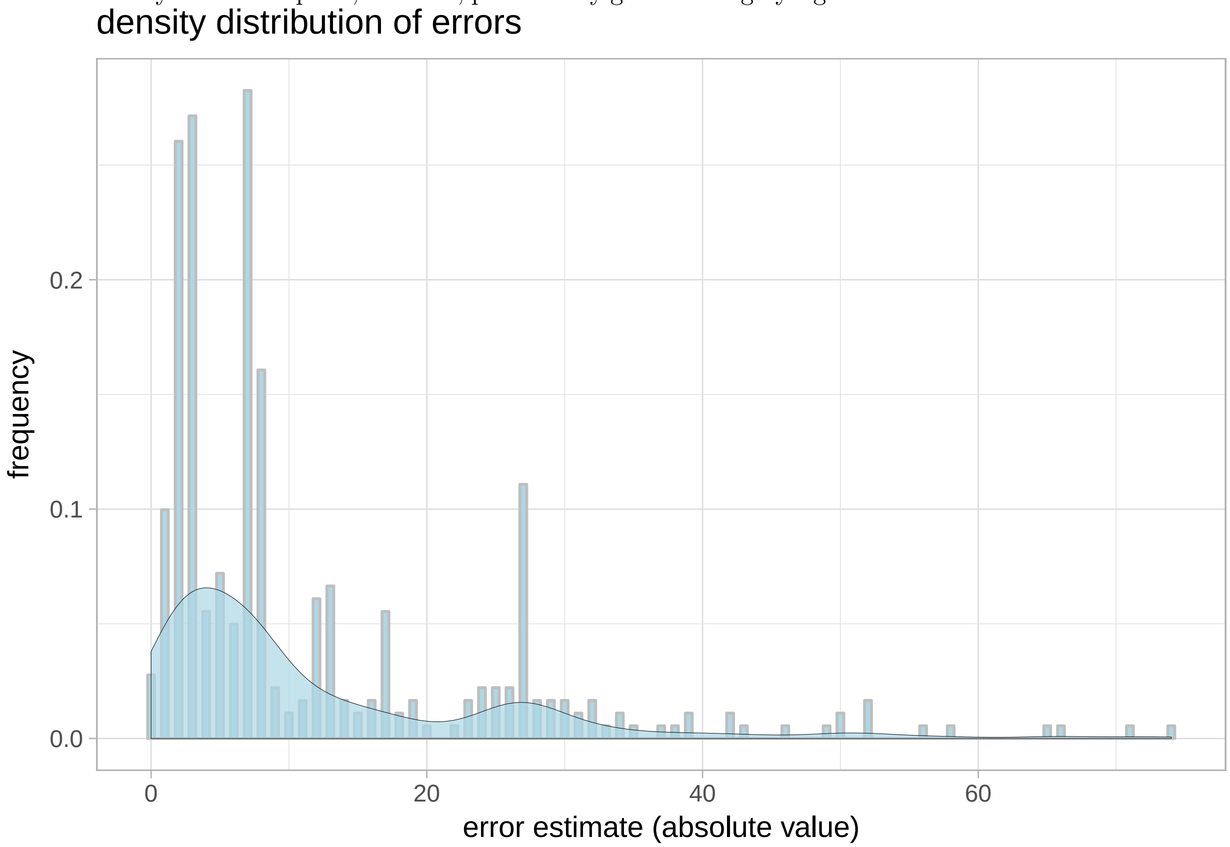
Error rate does not appear to be related to city name, so this control variable may be excluded from analysis.



A first-pass analysis of errors via general linear mixed effects model suggests compatibility effects were not elicited and gridlines did not affect estimates.

```
##
## Call:
## glm(formula = error ~ compatibility * gridline, data = tidy_qualtrics)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -13.659   -8.237   -4.527    2.849   59.473
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 10.2366     1.3539   7.561 3.42e-13 ***
## compatibility                0.9146     1.9533   0.468    0.640
## gridlineyes                  1.2909     1.9253   0.671    0.503
## compatibility:gridlineyes    2.2173     2.7500   0.806    0.421
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 170.4845)
##
##     Null deviance: 61874  on 360  degrees of freedom
## Residual deviance: 60863  on 357  degrees of freedom
## AIC: 2885.5
##
## Number of Fisher Scoring iterations: 2
```

It is possible that this analysis is inadequate, however, particularly given the highly right-skewed distribution of error values:

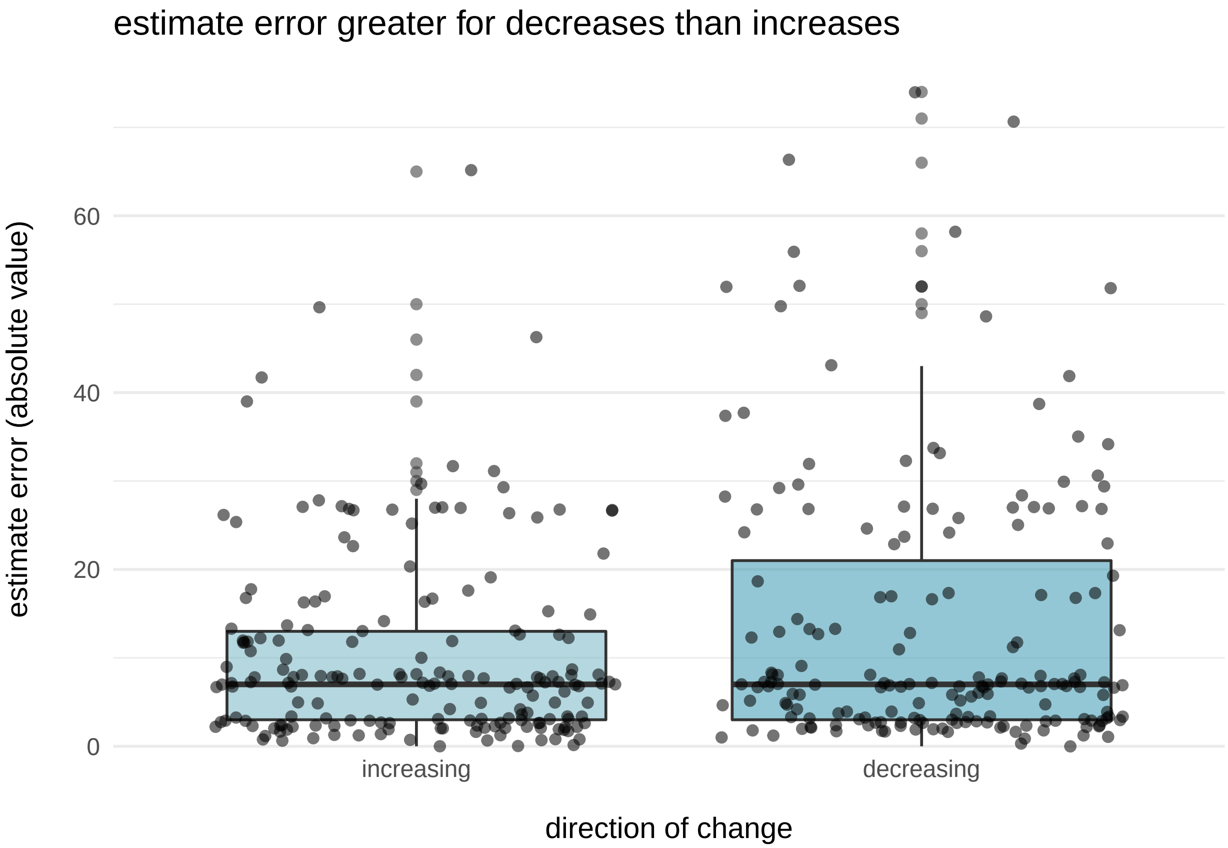


Based on the current analysis, both null hypotheses were not rejected. Further analysis, perhaps with log error, may be more revealing.

Interestingly, a linear mixed effects model (controlling for random effect of time spent on survey) shows a significant effect of direction, which suggests that estimates of decreases are more error-prone when made from a line graph. This may be an interesting avenue for further research.

```
## Type III Analysis of Variance Table with Satterthwaite's method
##                     Sum Sq Mean Sq NumDF  DenDF F value  Pr(>F)  
## direction          1002.61 1002.61     1 355.30  5.9509 0.01520 *
## gridline            552.58  552.58     1 351.95  3.2798 0.07099 .
## direction:gridline  184.72  184.72     1 355.79  1.0964 0.29577  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: error ~ direction * gridline + (1 | duration_sec)
##    Data: tidy_qualtrics
## 
## REML criterion at convergence: 2861.4
## 
## Scaled residuals: 
##     Min      1Q  Median      3Q     Max 
## -1.1193 -0.6641 -0.3559  0.3264  4.5046 
## 
## Random effects:
##  Groups       Name        Variance  Std.Dev.
##  duration_sec (Intercept)   0.01288  0.1135 
##  Residual                 168.48057 12.9800 
## Number of obs: 361, groups:  duration_sec, 116
## 
## Fixed effects:
##                              Estimate Std. Error      df t value
## (Intercept)                    11.622      1.368 279.754   8.494
## directionincreasing            -1.903      1.940 355.064  -0.981
## gridlineyes                     3.906      1.940 356.591   2.013
## directionincreasing:gridlineyes   -2.862      2.733 355.793  -1.047
##                              Pr(>|t|)    
## (Intercept)                   1.2e-15 ***
## directionincreasing            0.3275    
## gridlineyes                    0.0449 *  
## directionincreasing:gridlineyes   0.2958    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Correlation of Fixed Effects:
##             (Intr) drctnn grdlny
## drctnncrsng -0.705              
## gridlineyes -0.705  0.497       
## drctnncrsn:  0.501 -0.710 -0.710
```



estimate error greater for decreases than increases
estimate error (absolute value)
0
20
40
60
increasing
decreasing
direction of change