**OSF Pre-registration**

**Title:** The effect of conceptual metaphors on the interpretation of graphs

**Authors:** Greg Woodin, Lace Padilla

**Research questions:** Will (1) axis inversion, (2) graph orientation, (3) emotional valence, or (4) graph literacy affect participants’ understanding of line graphs?

**Study rationale:**

Line graphs are usually structured according to spatial metaphors commonly reported in the literature: quantity normally increases upwards along the y-axis, in line with the quantity metaphor more is up (e.g., Hartmann et al., 2012), and time increases right along the x-axis (e.g., Boroditsky, 2000). Some graphs (perhaps more frequently bar charts) also depict quantity as increasing rightwards along the x-axis, in line with the quantity metaphor more is right (e.g., Dehaene et al., 1993). If these graphical conventions are subverted, will people find the resultant graphs more difficult to interpret? This is what Pandey et al. (2015) showed; the current research is, in part, a conceptual replication of this study.

Another question is whether graphs are easier to interpret when quantity is mapped onto the y-axis rather than the x-axis, or the other way around. Previous research suggests that numbers may be more readily conceptualised using the horizontal axis (e.g., Woodin & Winter, 2018), but Fischer et al. (2005) report that participants found vertically-oriented bar charts easier to interpret than horizontally-oriented ones. Because line graphs normally depict quantity using the vertical axis, we predict that the y-axis will be more useful to participants than the x-axis.

In addition to quantity, spatial metaphors have also been reported for emotional valence. Along the vertical axis, research suggests that English speakers associate higher vertical space with more positive valence, and lower vertical space with more negative valence (e.g., Meier & Robinson, 2004). For the horizontal axis, the body-specificity hypothesis predicts that English speakers will associate leftwards space with negative valence, and rightwards space with positive valence (e.g., Casasanto, 2009). Based on this, we are interested to see whether participants will find it easier to understand graphs that are congruent with these valence metaphors (e.g., where the magnitude of a positively valenced quantity increases upwards along the vertical axis) compared to ones that are incongruent. We hedge this prediction with the observation that the infamous Stand Your Ground visualisation relating to gun deaths in Florida was notoriously difficult to understand chiefly *because* it reversed the y-axis, despite this axis reversal being congruent with the valence metaphor bad is down.

We are also interested to see if people with higher graph literacy will be less affected by our experimental manipulations than people with lower graph literacy, due to their familiarity with how to read and interpret graphs.

All in all, we hope that our results will demonstrate the complex interplay between metaphor and graphical conventions in determining how easy a graph is to interpret.

**References for work cited in this document:**

Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial

metaphors. *Cognition,* 75, 1-28.

Casasanto, D. (2009). Embodiment of abstract concepts: Good and bad in right- and

left-handers. *Journal of Experimental Psychology: General,* 138(3), 351-367.

Cleveland, W. S. (1984). Graphical methods for data presentation: Full scale breaks,

dot charts, and multibased logging. *The American Statistician*, 38(4), 270-280.

Dehaene, S., Bossini, S. & Giraux, P. (1993). The mental representation of parity and

number magnitude. *Journal of Experimental Psychology: General*. 122(3): 371-

396.

Fischer, M. H., Dewulf, N. & Hill, R. L. (2005). Designing bar graphs: Orientation

matters. *Applied Cognitive Psychology,* 19, 953-962.

Hartmann, M., Grabherr, L. & Mast, F.W. (2012). Moving along the mental number

line: Interactions between whole-body motion and numerical cognition. *J. Exp.\*

*Psychol. Hum. Percept. Perform*, 38, 1416–1427.

Meier, B. P. & Robinson, M. D. (2004). Why the sunny side is up: Association between

affect and vertical position. *Psychol. Sci.,* 15, 243–247.

Pandey, A. V., Rall, K., Satterthwaite, M. L., Nov, O. & Bertini, E. (2015). How

deceptive are deceptive visualisations?: An empirical analysis of common

distortion techniques. *Proceedings of the 33rd Annual ACM Conference on*

*Human Factors in Computing Systems,* 1469-1478.

Smith, N. J. & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3), 302-319.

Woodin, G. & Winter, B. (2018). Placing abstract concepts in space: Quantity, time

and emotional valence. *Frontiers in Psychology*, 9.

**Hypotheses:**

* Directional 1: normal line graphs (congruent with quantity metaphors) will be easier to understand than line graphs with either the x-axis or y-axis reversed in direction (incongruent with quantity metaphors)
* Directional 2: line graphs with quantity on the y-axis will be easier to understand than line graphs with quantity on the x-axis
* Directional 3: line graphs congruent with valence metaphors will be easier to understand than line graphs incongruent with valence metaphors
* Directional 4: participants with higher graph literacy will be less affected by experimental manipulations than those with lower graph literacy

**Existing** **data:** Pre-registering prior to collecting new data

**Explanation of existing data:** No existing data

**Data** **collection procedures:** Study will be advertised on Amazon Mechanical Turk and users of the platform will volunteer to participate. Participation will be restricted to US users of the platform. Payment will be $1 per participant based on US minimum wage. Participants have to be 18 years or older to use MTurk. Data from participants who quit partway through study will be excluded, as will data from participants who self-report not being a native speaker of English, or speaking a language with a different reading/writing directionality (e.g., right to left) to English, even if not a native language.

**Sample size:** 300 participants

**Sample size rationale:** Pandey et al. (2015) collected data from 240 participants and achieved a significant effect. Because of our more complicated statistical models, we will collect more to pre-emptively account for model convergence issues.

**Stopping rule:** 300 participants

**Manipulated variables and the different levels of each variable:**

1. Axis Inversion
   * No inversion
   * Either x- or y-axis inverted
2. Orientation
   * Quantity on y-axis, time on x-axis
   * Quantity on x-axis, time on y-axis
3. Valence
   * Positive
   * Negative
4. Trend
   * Rising
   * Falling

Trend included to control for this possible confound; no specific hypotheses are formulated regarding this factor, but we will look at it exploratorily afterwards to see if there is an unpredicted effect.

**Indices:** N/A

**Study type:** Experiment

**Blinding:** All dependent variables will be measured automatically and fed into statistical models. Because no coding is necessary for this study, no blinding is necessary, either.

**Study design:** Axis inversion and graph orientation varied between subjects, because varying these two factors from task to task would be misleading and thus confusing to participants. Valence and trend varied within subjects to assess effect of valence on responses, to control for the possible confounding influence of trend, and to ensure that participants cannot use the same strategy to interpret each graph.

**Randomisation:** Participants will be randomly allocated into one of the following between-participants conditions:

G1. Normal graph with quantity on y-axis and time on x-axis

G2. Normal graph with quantity on x-axis and time on y-axis

G3. Y-axis inverted with quantity on y-axis and time on x-axis

G4. Y-axis inverted with quantity on x-axis and time on y-axis

G5. X-axis inverted with quantity on y-axis and time on x-axis

G6. X-axis inverted with quantity on x-axis and time on y-axis

Participants will then complete the following four trial types in a randomised order:

V1. Positive valence, rising trend

V2. Positive valence, falling trend

V3. Negative valence, rising trend

V4. Negative valence, falling trend

We will do this for 250 participants and then see how much incomplete data there is for the conditions G1-G6. We will then selectively run trials for specific conditions (e.g., G1) that have more incomplete data than others to achieve a balance across groups. We will do this until we have exactly 300 participants.

**Statistical models:** Linear and logistic regression with random slopes and intercepts for between-participant variation

Phase 1:

Logistic regression (categorical dependent variable):

Accuracy ~ AxisInversion + Orientation + AxisInversion:Valence

(1 + Valence|Subject)

Linear regression (continuous dependent variable):

Confidence ~ AxisInversion + Orientation + AxisInversion:Valence

(1 + Valence|Subject)

Linear regression (continuous dependent variable):

RT ~ AxisInversion + Orientation + AxisInversion:Valence

(1 + Valence|Subject)

Phase 2 (looking at how effects observed in phrase 1 are mediated by graph literacy):

Logistic regression (categorical dependent variable):

Accuracy ~ (AxisInversion + Orientation + AxisInversion:Valence) \* GraphLiteracy +

(1 + Valence|Subject)

Linear regression (continuous dependent variable):

Confidence ~ AxisInversion + Orientation + AxisInversion:Valence \* GraphLiteracy +

(1 + Valence|Subject)

Linear regression (continuous dependent variable):

RT ~ AxisInversion + Orientation + AxisInversion:Valence \* GraphLiteracy +

(1 + Valence|Subject)

If these models do not converge, in the following order we will (1) sum-code AxisInversion, Orientation and Valence (2) use all\_fit from the R package ‘afex’ to find a suitable optimiser, (3) remove random slopes, and (4) remove random intercepts.

**Transformations:**

Confidence and RT will be centred. RT will be log-transformed to the base *e* (2.718282) (see Cleveland, 1984) due to research suggesting that reaction times are scaled logarithmically (e.g., Smith & Levy, 2013).

**Follow-up analyses:** We will perform likelihood ratio tests on the above models to test the likelihood of the full models (given the data) against models with each predictor left out individually. In the examples below, the term Accuracy can be swapped out for Confidence and RT.

Full model:

xmdl\_def <- Accuracy ~ AxisInversion + Orientation + AxisInversion:Valence

(1 + Valence|Subject)

AxisInversion left out:

xmdl\_noaxis <- Accuracy ~ Orientation + AxisInversion:Valence +

(1 + Valence|Subject)

Orientation left out:

xmdl\_noorient <- Accuracy ~ AxisInversion + AxisInversion:Valence +

(1 + Valence|Subject)

AxisInversion:Valence left out:

xmdl\_nointer <- Accuracy ~ AxisInversion + Orientation +

(1 + Valence|Subject)

Tests:

anova(xmdl\_noaxis, xmdl, test = 'Chisq')

anova(xmdl\_noorient, xmdl, test = 'Chisq')

anova(xmdl\_nointer, xmdl, test = 'Chisq')

**Inference criteria:** In line with mainstream Null Hypothesis Significance Testing, we will use *p* = 0.05 as a hard cut-off for statistical significance. We will use this criterion specifically to assess the statistical significance of the likelihood ratio tests.

**Data exclusion:** Outliers will not be excluded. Participants will be told that they should answer as quickly and accurately as possible, and responses that are greater than 3 standard deviations above the mean response will be excluded. Respondents with incomplete data will be excluded.

**Missing data**: Data partially missing because participant quit study halfway through, or because they failed an attention check question, will be excluded. If applicable, data partially missing due to unforeseen software errors will also be excluded.

**Exploratory analyses**: We will look to see if trend had an effect on responses. We will also look at whether x-axis inversion had a greater effect on task responses than y-axis inversion, or vice versa.