Preregistration: Can a small stimulus set reliably estimate individual differences in semantic salience?

Benjamin de Haas

Department of Psychology, Justus Liebig Universität Giessen, Germany

Corresponding Author: Benjamin de Haas
benjamindehaas@gmail.com
Department of Psychology, Justus Liebig Universität Giessen, Germany
Otto-Behaghel-Str. 10F
35394 Giessen, Germany
Fon: +49-641-9926125

AIMS

The aims of this study are two-fold

- To replicate the recent finding of consistent individual differences in the tendency to fixate objects along six semantic dimensions (de Haas et al., 2019)

- To test whether such individual differences can be reliably estimated using a smaller stimulus set than that of the original study (de Haas et al., 2019)

DESIGN AND ANALYSIS PLAN

The design and analysis will stay as close as possible to that of (de Haas et al., 2019), as detailed below.

Subjects. We will aim to test 100 healthy adults with normal or corrected-to normal vision and between 18 and 60 years old, to be recruited by the Leibniz Institute of Psychology Information (ZPID). Testing will take place at the ZPID in Trier, Germany utilizing their service *PsychLab offline*. If 100 participants cannot be recruited until 1 November 2019, we will conclude the study with a smaller sample, unless sample size at that point is still below 50 (in which case the study

will be concluded when a sample size of 50 is reached). These sample sizes were based on power calculations determining uncertainty about observed correlations indicating acceptable predictive validity ($r > .7$; see below, Analysis). The 95% confidence interval for an observed correlation falling just into the acceptable range ($r = .7$) extends from .52-.82 for n = 50 and from .58-.79 for n = 100 (Fig. 1; note the diminishing reduction of uncertainty with increasing n).

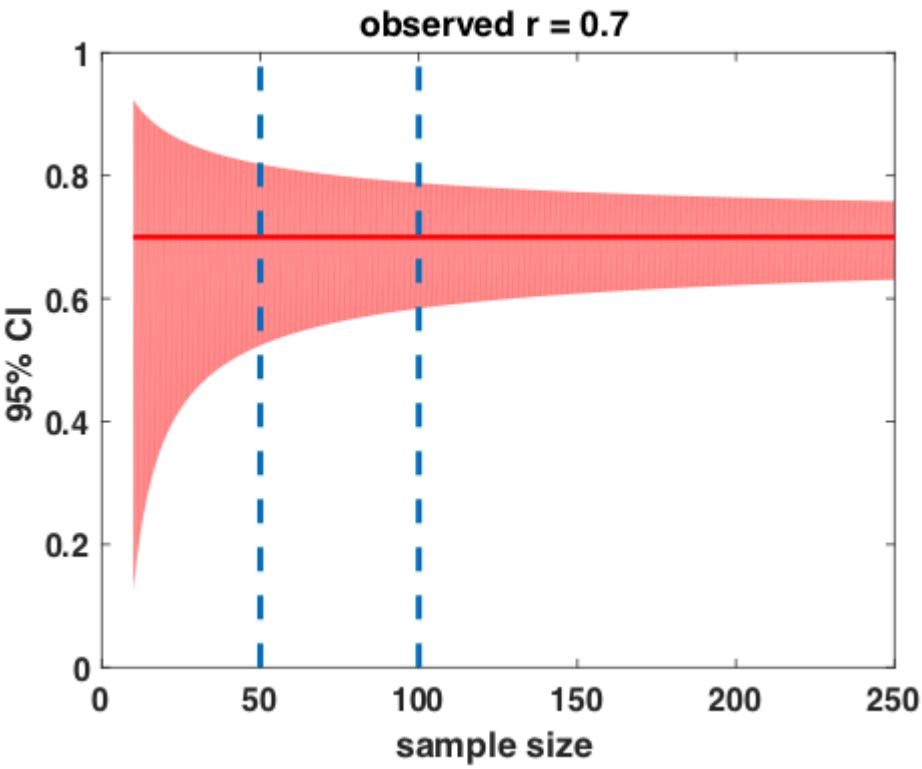

Figure 1. Uncertainty about an observed correlation as a function of sample size. The solid red line indicates the observed correlation of .7, corresponding to just acceptable predictive validity. The error shade shows the extent of the 95% confidence interval as an indicator of uncertainty about the effect size. Dashed blue lines mark the planned minimum and desired sample sizes of n = 50 and n = 100, respectively.

The local institutional review board (Lokale Ethikkomission FB06, JLU Giessen) approved the study and participants will have to provide written informed consent before participating. Every participant will be invited for two appointments, 3 to 14 days apart and reimbursement is 10 € per hour.

Stimuli. Participants will be presented with 700 natural images depicting a wide variety of complex everyday scenes (http://www-users.cs.umn.edu/~qzhao/predicting.html), or a subset thereof (see below). Semantic metadata (Xu et al., 2014) consist of binary pixel maps for 5551 objects in these images and accompanying labels for 12 semantic dimensions. We modified the provided labels to minimize overlap between them in the following way: The (neutral) *Faces* label was removed from all objects with the *Emotion* label (i.e. emotional faces); The *Smell* label was removed from all objects with the *Taste* label; The *Operable* and *Gazed* labels were removed from all objects with

the *Touched* label; The *Watchable* label was removed from all objects with the *Text* label. This step allowed to test individual differences in fixation attraction for a given dimension largely independently from the others. Without this step some of the attributes would have been perfectly confounded (for instance, all text is watchable; all emotional faces are faces).

Participants will view blocks of 100 images on a screen, with seven blocks in total on the first testing day, two of which will be repeated on the second day. Stimulus presentation and data collection was coded in MATLAB Version R2019a (MathWorks, Natick, MA) using Psychtoolbox Version 3.0.12 (Pelli, 1997; Kleiner et al., 2007). The order of images will be fixed across all participants and each image will be presented for 3s, with a self-paced period of a central fixation dot in between. Participants will simply be instructed to 'look at the images in any way [they] want' and initiate the onset of each image with a press of the space bar.

Participants will sit at a distance of ~64 cm from the screen (BenQ XL2430T) and see the stimuli at a resolution and size of 1920 x 1080 pixels and a size of 29.0 x 22.2 degrees visual angle.

Data Collection. The gaze of participants will be collected from the left eye with an Eyelink 1000 Plus utilizing the desktop mount (SR Research, Ottawa, Canada) at a frequency of 1 kHz.

At the beginning of each block, participants will complete a nine-point calibration and validation procedure, which will be repeated if necessary. Fixation data will be collected online using the 'normal' setting of the Eyelink parser (saccade velocity and acceleration thresholds of 30 d.v.a./s and 9500 d.v.a./s^2, respectively) and the default drift check procedure in each inter-trial interval.

Data Processing and Analyses. Onset fixations, fixations with a duration below 100ms will be disregarded (minimum fixation duration following standard recommendations by SR Research). Fixations that will fall on or within a distance of ~0.5 d.v.a from a labeled object will be assigned the corresponding label. Unlabeled fixations will be disregarded for the calculation of cumulative fixation times and the individual proportion of *first* fixations (see below).
In order to quantify the individual tendency to fixate objects bearing a given attribute label, we will first calculate the cumulative fixation time for all labeled fixations made by a given observer to a given image set. This will allow us to calculate the proportion of this time spent on a given attribute in a second step.

The *first* fixations analysis will consider the proportion of labeled *first* fixations (after image onset) landing on objects with a given attribute for a given observer and image set. Individual proportions of cumulative fixation time and first fixations will be expressed in %.

Consistency and re-test correlations. Data from Day 1: To estimate the consistency of individual differences in % cumulative fixation time or % first

fixations along a given attribute dimension, we will calculate these measures independently for two random halves of the images for each observer and calculated the split-half Pearson correlation. This procedure will be repeated 1000 times. We will inspect the frequency histograms of all correlations and consider the median correlation coefficient across random image splits as an indicator of consistency. The accompanying two-sided *P*-values will be Holm-Bonferroni adjusted for the number of dimensions tested. The dimensions to be tested were determined based on their consistency in (de Haas et al., 2019): Faces, Text, Touched, Taste and Motion for cumulative dwell times and Faces, Text and Touched for the proportion of first fixations (the Faces label will be collapsed across emotional and neutral Faces) .

We will further test the ability to reliably estimate these individual differences with a smaller subset of images, correlating % cumulative and first fixations towards a given feature for the first n images with the corresponding values for the remaining images (starting from n = 30 and up to n = 200). The resulting correlation values will be plotted by n for each dimension to explore the set-size dependence of the estimate consistency. Additionally we will test the statistical significance of the correlations for each dimension for set sizes n = 40 (extra small, XS), n = 100 (very small, VS) and n = 200 (small, S), each time adjusting P-Values with Holm-Bonferroni correction for the six dimension-specific correlations.

Data from Day 2: To estimate the re-test reliability of individual differences in % cumulative fixation time or % first fixations for a given attribute, we will correlate the corresponding values for the first session with those of the second for the XS, VS and S test.

The order of images (and thus the content of the small sets) was determined a priori, based on the dataset by de Haas et al., 2019. An iterative search algorithm aimed to select images in an order that maximised the predictive validity for the full set at each step (i.e. for each number of images < 200). The order of images on the second day will be pseudo-randomised relative to the first, but only within the three subsets (i.e. the order of the first 40 images, will be randomised relative to the first day, then the order of images 41-100, then 101-200).

REFERENCES

de Haas B, Iakovidis AL, Schwarzkopf DS, Gegenfurtner KR (2019) Individual differences in visual salience vary along semantic dimensions. Proceedings of the National Academy of Sciences, in press.

Kleiner M, Brainard D, Pelli D (2007) What's new in Psychtoolbox-3? Perception 36:14

Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 10:437–442

Xu J, Jiang M, Wang S, Kankanhalli MS, Zhao Q (2014) Predicting human gaze beyond pixels. J Vis 14(1): 28, 1-20