

Embodiment as preparation: Pupillary responses to words that convey a sense of brightness or darkness (Supplementary Information)

Sebastiaan Mathôt^{1*}, Jonathan Grainger¹, Kristof Strijkers²

¹ Aix-Marseille University, CNRS, LPC UMR 7290, Marseille, France

² Aix-Marseille University, CNRS, LPL UMR 7309, Aix-en-Provence, France

* Corresponding author

Address for correspondence:

Sebastiaan Mathôt, PhD
Aix-Marseille Université / CNRS
Laboratoire de Psychologie Cognitive
3 Place Victor Hugo
Centre St. Charles, Bâtiment 9, Case D
13331 Marseille
France

s.mathot@cogsci.nl

<http://www.cogsci.nl/smathot>

Robustness checks

To verify that our main conclusions do not depend on the specifics of the analyses, we conducted several alternative analyses, or ‘robustness checks.’

Main analysis

In the original analysis (described in the main text), we conducted separate LMEs for each 10 ms, using the following model (in the style of the R package `lme4`):

```
pupil_size ~ word_category
             + (1+word_category|subject_nr)
             + (1+word_category|word)
```

Here, *word_category* was a discrete factor with two levels (bright or dark). For this original analyses, we excluded neutral words, because neutral words were not matched to the darkness- and brightness-conveying words.

As a robustness check, we used the following alternative model:

```
pupil_size ~ rating_brightness
             + (1+rating_brightness|subject_nr)
             + (1+rating_brightness|word)
```

Here, *rating_brightness* was a continuous predictor corresponding to the brightness of the words as rated by the participants on a scale from 1 to 5. We now also included neutral words. Crucially, also when analyzing the effect of semantic brightness in this way, but using the same significance criteria ($p < .05$ for at least 200 consecutive milliseconds) there was a reliable effect of semantic brightness (from 1670 to 1920 ms after word onset).

Individual-participant analysis

In the original analysis, we looked at mean pupil size during the 1.5 - 2 s window for each participant, and tested whether pupil size was larger for darkness- than brightness-conveying words, using both a Bayesian and a classical one-sided paired-samples t test.

As a robustness check, we tested whether this difference also holds when considering mean pupil size during the entire 3 s window of word presentation. When changing the analysis window, but keeping the analysis otherwise identical, there was again evidence for an effect ($B_f = 6.3$; $t(29) = 2.6$, $p = .007$).

Individual-word analysis

In the original analysis, we looked at mean pupil size during the 1.5 - 2 s window for each word, and tested whether pupil size was larger for darkness- than brightness-conveying words, using both a Bayesian and a classical one-sided independent samples t test.

As a robustness check, we tested whether this difference also holds when considering mean pupil size during the entire 3 s window of word presentation. When changing the analysis window, but keeping the analysis otherwise identical, there was again evidence for an effect ($B_f = 4.0$; $t(64) = 2.2$, $p = .014$).

Model comparisons

To test how well the semantic brightness, emotional intensity, and valence of the words (as rated by the participants) predicted pupil size, we conducted separate LMEs for each 10 ms window, using the following models (in the style of the R package `lme4`):

```
pupil_size ~ rating_brightness
              + (1+rating_brightness|subject_nr)
              + (1+rating_brightness|word)
```

```
pupil_size ~ rating_intensity
              + (1+rating_intensity|subject_nr)
              + (1+rating_intensity|word)
```

```
pupil_size ~ rating_valence
              + (1+rating_valence|subject_nr)
              + (1+rating_valence|word)
```

Here, *rating_brightness*, *rating_intensity*, and *rating_valence* are the subjective word ratings, as described in the Methods of the main text. (We used three separate models with a single predictor, rather than a single model with three predictors, because the high correlations between ratings would cause a combined model to produce a misleading impression of what the actual data looks like.) The results are shown in [Figure 1](#), in which the absolute t values (as a proxy for effect size) are plotted for each of the predictors as a function of time from word onset:

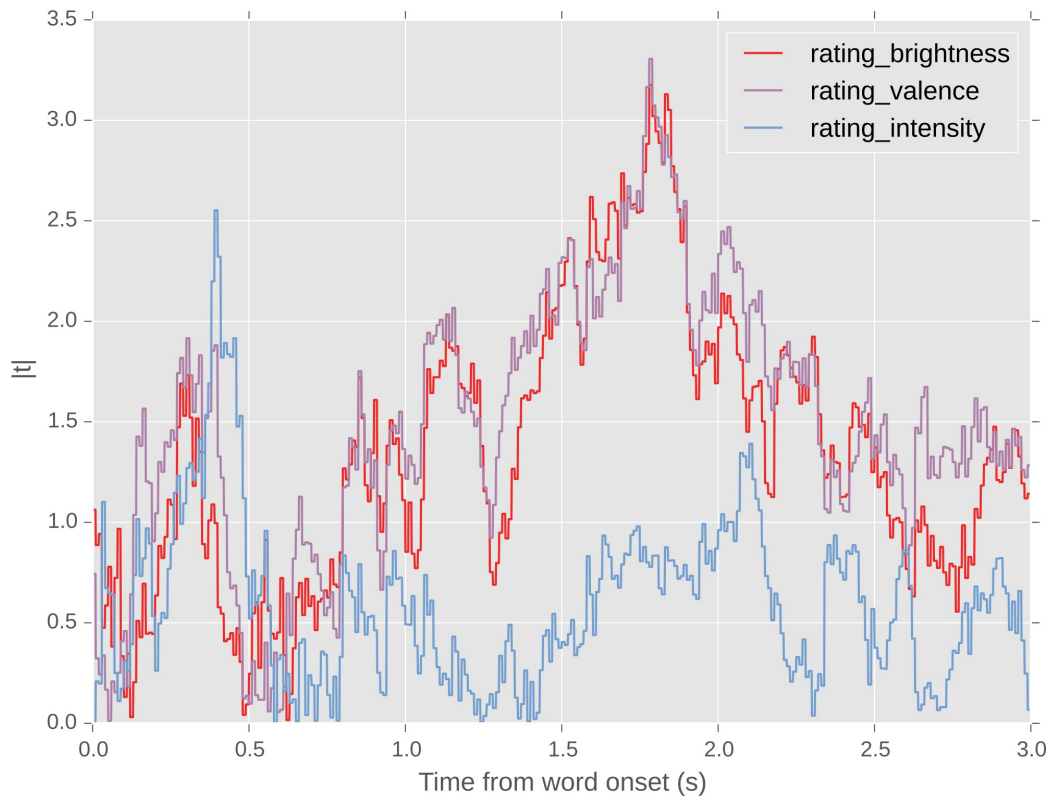


Figure 1. The absolute t values of the fixed effect of rated brightness, valence, and emotional intensity as a function of time from word onset.

A few things are clear from Figure 1.

First, the t values of `rating_brightness` and `rating_valence` are highly correlated. This is to be expected, because brightness and valence ratings were highly correlated ($r = .89$; described in the main text). This correlation is too high to allow a meaningful statistical dissociation (for example through partial effects) between brightness and valence ratings. However, as we have described in the main text, from a theoretical point of view it is far more likely that the effect of valence on pupil size is mediated by semantic brightness than vice versa.

Second, emotional intensity predicts pupil size quite well during the initial 'orienting phase' of the pupillary response, around 500 ms after word onset. This may mean that a word's emotional intensity affects the pupil's orienting response (which would be interesting in itself), but may also result from correlations with other factors, such as word length. More importantly for the present study, during the interval of the semantic pupil effect, emotional intensity does not, or hardly, predict pupil size.

