Visual object categorization in the brain: what can we really learn from ERP peaks?

Guillaume A. Rousselet*, Cyril R. Pernet², Roberto Caldara² and Philippe G. Schyns¹

1 Centre for Cognitive Neuroimaging, Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK
2 Brain Research Imaging Centre, SINAPSE Collaboration, University of Edinburgh, Edinburgh, UK
3 Department of Psychology, University of Fribourg, Fribourg, Switzerland
*Correspondence: guillaume.rousselet@glasgow.ac.uk

A commentary on

Face-sensitive processes one hundred milliseconds after picture onset

In three experiments, Dering et al. (2011) measured the amplitude and latency of the mean P1 and N170 in response to faces, cars, and butterflies, cropped or morphed. The N170 was sensitive to cropping but did not differentiate frontal views of faces and cars. Throughout, P1 amplitude was larger for faces than objects. The authors concluded that P1, not N170, is a reliable face-sensitive event.

Although Dering et al. (2011) and Thierry et al. (2007) correctly identified problems with P1 and N170 ERP interpretations, these same problems are evident in their own research, and, consequently, their results are equally inconclusive. We identify four shortcomings in Dering et al.'s approach that also apply to the literature they contest (e.g., Eimer, 1998; Allison et al., 1999; Carmel and Bentin, 2002; Itier and Taylor, 2004; Rousselet et al., 2004; Rossion and Jacques, 2008): categorical design, uncontrolled task demands, group statistics, and peak measurements. ERPs are worthwhile measurements of visual processing, but only when the shortcomings reviewed below are addressed.

BEYOND CATEGORICAL DESIGNS

Categorical designs compare few categories of input stimuli against brain measurements. When dealing with complex visual stimuli, categorical designs make it difficult to attribute a brain response to specific object categories (e.g., a face), features (e.g., the eye) or functions, because the specificity of the response (e.g., to the face) is determined by contrast with responses from other categories (e.g., cars) and informal hypotheses tested. Unfortunately, category members share a dense correlational structure of low-level visual properties (e.g., luminance energy, main directions of orientation, spatial frequency composition), which cannot all be controlled with a finite number of contrast categories. Consequently, the brain response's specificity might be due to differences in input statistics, not to the category itself (Schyns et al., 2003).

Whereas Dering et al. (2011) controlled total luminance, contrast, and size, other physical dimensions affecting P1 amplitudes and categorical judgments were not controlled (e.g., spectral profiles of contrasted categories: VanRullen and Thorpe, 2001; Sowden and Schyns, 2006; VanRullen, 2006; Honey et al., 2008; Rousselet et al., 2008a). Consequently, a significant ERP peak difference between two input categories does not ensure that the difference relates to the category per se, as opposed to the uncontrolled statistics of low-level image properties (Schyns et al., 2003; Pernet et al., 2007). Categorical designs typically use few categories, but must still control for a much larger set of low-level image dimensions. By neglecting these experimental controls, it is almost inevitable that such designs will deliver the type of inconclusive results that animate unnecessary debates in the P1 and N170 face literature. Thierry et al.'s (2007) own results elegantly demonstrate this point. In their Experiment 3, they overlaid faces with cars and instructed observers to attend to either category. Here, when the two categories were better controlled for their low-level properties (though still not perfectly), the categorical modulations on the P1 component were abolished. Thus, varying levels of quality of stimulus control in Thierry et al.'s work fuels their own debate: if Dering et al. had correctly identified the P1 face-sensitiveness, Thierry et al.'s (2007) third experiment would demonstrate an increased P1 amplitude for faces overlaid with cars. It did not. This contradiction illustrates the pitfalls of poor stimulus control.

CONTROL OF TASK DEMANDS

Dering et al.'s (2011) participants discriminated between faces and cars or butterflies. Assuming that the stimuli were properly controlled, observers in Dering et al.'s task could still discriminate a face from any car or butterfly using only the presence or absence of one of many face features (global appearance, texture, left or right eye, eyelashes, nose) or their combinations. This weakness highlights the broader (and often neglected) role of task demands in cognitive neuroimaging studies of recognition. A face can be categorized as “John,” a Western Caucasian, who is 40, is handsome, in good health and currently has a happy face. Suggestion that “the P1 (or the N170) is categorically sensitive to faces as opposed to cars or butterflies” leaves unresolved the issue of whether the N170 is sensitive to the face, as an object category, or to the categorization task that is used to test the face with. If the P1 effects of Dering et al. arose from processing categorization-specific facial information, changing task demands (by changing the categorization performed on the same face) should differentially influence the brain response (Schyns, 1998; Pernet et al., 2007). The task demands of Dering et al. are not sufficiently controlled to inform the underlying information processing function of the P1 or N170.

PROBLEMS WITH GROUP STATISTICS

One of the main results of Dering et al. (2011) is that the N170 peak amplitudes do not differ between different stimulus categories. This result is only valid if the data conform to the assumptions of normality.
and heteroscedasticity that underlie the analyses. Standard linear statistics are extremely sensitive to departure from these assumptions leading to low statistical power (Wilcox, 2005). A convincing demonstration that the N170 is not a marker of face processing would require detailed investigation of the data distribution using robust statistics and actually showing the data. Moreover, effects present in individual observers might be washed out by grand average group statistics (Rousselet and Pernet, 2011; Rousselet et al., 2011).

Effects might also be distributed across electrodes, as multivariate multi-electrode analyses can reveal (Philiastides and Sajda, 2006; Philiastides et al., 2006; Ratcliff et al., 2009). Without systematic univariate or multivariate analyses in single-subjects, null effects as reported for the N170 by Dering et al. are inconclusive.

Compared to group analyses, single-trial parametric analyses also provide more powerful tools to infer function because they exploit the richness of single-trial data to examine how parametrically manipulated input information co-varies with brain response variability (Hubel and Wiesel, 1979; Freedman et al., 2002, 2003; Ratcliff et al., 2009; Schyns et al., 2009a; Schyns, 2010; Pernet et al., 2011b; Rousselet and Pernet, 2011). In contradiction to Dering et al. (2011), such analyses have revealed sensitivity to category differences from about 100 ms to about 300 ms, encompassing the N170 (Philiastides and Sajda, 2006; Philiastides et al., 2006; Rousselet et al., 2008b, 2011; Schyns et al., 2009b; van Rijssbergen and Schyns, 2009). Critically, parametric studies demonstrate that mean ERPs cannot elucidate the information content of single-trial fluctuations around the central tendency of the distribution (see Smith et al., 2004, Figure 1, for a graphical demonstration).

EXTENDING FROM PEAK MEASUREMENTS

Finally, by restricting analyses to peak ERP differences between groups, Dering et al.’s (2011) results fall short of providing a complete picture of the relevant brain responses. Important effects occur between peaks, which are quantifiable by applying analyses at all time points and electrodes (Delorme et al., 2007; Schyns et al., 2007, 2009b, 2011; Rousselet et al., 2008b; Vizzioli et al., 2010a,b; Pernet et al., 2011a). Peaks can index the end of a process rather than the specific information processing mechanism leading to this endpoint (Schyns et al., 2007).

In sum, Dering et al. (2011) applied a pervasive but now arguably out-dated approach to an otherwise useful brain measurement. The limitations of their approach apply as equally to their own results as to those they criticize. The issue of “face specificity” of the N170 should be addressed with more powerful methods such as using parametric manipulations of well-controlled stimuli, addressing task demands, and systematic single-subject statistics. Otherwise, numerous inconclusive data will continue to fuel unnecessary and opinionated controversies.

REFERENCES


Frontiers in Human Neuroscience www.frontiersin.org


