Quantifying the time course of visual object processing using ERPs: it's time to up the game

Guillaume A. Rousselet[1]* and Cyril R. Pernet[2]

[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

Hundreds of studies have investigated the early ERPs to faces and objects using scalp and intracranial recordings. The vast majority of these studies have used uncontrolled stimuli, inappropriate designs, peak measurements, poor figures, and poor inferential and descriptive group statistics. These problems, together with a tendency to discuss any effect $p < 0.05$ rather than to report effect sizes, have led to a research field very much qualitative in nature, despite its quantitative inspirations, and in which predictions do not go beyond condition A > condition B. Here we describe the main limitations of face and object ERP research and suggest alternative strategies to move forward. The problems plague intracranial and surface ERP studies, but also studies using more advanced techniques – e.g., source space analyses and measurements of network dynamics, as well as many behavioral, fMRI, TMS, and LFP studies. In essence, it is time to stop amassing binary results and start using single-trial analyses to build models of visual perception.

Keywords: faces, N170, robust statistics, single-trial analyses, mechanisms

Visual cognition depends on fast and progressive transformations of retinal inputs into higher-order representations useful for decision-making (Rousselet et al., 2004a; DiCarlo and Cox, 2007; Schyns et al., 2009a). Hence, a theory of visual cognition must specify the information content of brain activity from retinal input to decision-making, and the operations performed on this information – the mechanisms. This theory must also specify how information content and mechanisms develop during childhood and deteriorate with age.

Because of the temporal resolution of EEG and MEG, ERP research is well suited to identify the cascade of processes that lead to decision-making (Schyns, 2010). ERP research has matured its techniques and theories since the first reports of larger ERPs to faces compared to objects. Progress has, however, been inhomogeneous: most recent ERP studies use outdated experimental designs and statistical techniques, and poor interpretation frameworks. The field shows its immaturity by its incapacity to make precise predictions about the timing and magnitude of expected effects: the fault of using group statistics and categorical designs, reporting effects as significant or not with no consideration for effect sizes, and a reluctance to model the results for future hypothesis testing. In sum, most ERP studies of visual cognition are plagued by problems that need to be addressed urgently.

USE CONTROLLED STIMULI

The use of uncontrolled stimuli makes the interpretation of ERP differences among image categories difficult to interpret because it is unclear whether the effects are due to low-level, physical, differences or high-level, semantic, differences (VanRullen and Thorpe, 2001). For instance, there have been speculations about the sensitivity of the P1 component to object categorical information,

mostly from studies using uncontrolled images (Itier and Taylor, 2004). Low-level image properties, such as luminance, contrast, and amplitude spectrum can be manipulated using published toolboxes (Knebel et al., 2008; Willenbockel et al., 2010). Low-level image properties should be equated across image categories in studies seeking to reveal high-level differences (Rousselet et al., 2005, 2008a; Honey et al., 2008; Gaspar and Rousselet, 2009). Alternatively, low-level properties can be manipulated parametrically to reveal the influence of low-level factors, such as contrast, on high-level cognition (Macé et al., 2005). Low-level and high-level properties can also be explicitly modeled together (Rousselet et al., 2008b). For other research questions, it might be essential to control local information as well, for instance in studies of facial expression processing (Schyns et al., 2007). Although images cannot be completely equated without loosing meaning, and there is no optimal procedure to control stimuli, the problem can no longer be ignored.

USE A CONSISTENT FRAMEWORK TO INTERPRET TASK EFFECTS

Instead of controlling physical differences among images, an alternative strategy consists in measuring ERP modulations due to task differences while keeping stimuli constant (VanRullen and Thorpe, 2001; Rousselet et al., 2007). More generally, task manipulations are essential to understand the nature of ERP differences, one of the most enduring debates in the field (Carmel and Bentin, 2002; Rossion et al., 2002). However, the interpretation of task effects and their comparison across studies is complicated by the use of inconsistent terms: for instance, the N170 and the M170 to faces have been described as sensitive, selective, or specific responses to faces (Carmel and Bentin, 2002; Liu et al., 2002; Itier and Taylor, 2004; Joyce and Rossion, 2005); the intracranial N200 has been

described as a specific response from a face module (Puce et al., 1999). Clear operational definitions of specific, selective, and preferential responses have been described, providing a useful framework to interpret task effects and compare them across studies (Pernet et al., 2007). A *specific response* is a brain response for which activity is exclusively observed in the context of an interaction between a category (information) and a task (process) (Fodor, 2001). Concretely, if the N170 was face specific, one should observe the N170 only for face stimuli in a given task, and no evoked activity (no difference from baseline) for other stimuli and tasks. A *selective response* is defined as a category by task interaction, in which the target condition is higher than the control conditions, which themselves are higher than baseline. For ERPs, that means that a stronger component should be observed for one category (e.g., faces) relative to others (e.g., cars) but only for a given task (e.g., categorization vs. discrimination). Finally, a *preferential response* is task-independent, such that brain responses are stronger for a given category compared to all the others. Preferential activity reflects some specialization for the considered category; however, it does not capture the interaction with the task. The point is that the criterion for category selectivity utilized in most publications is not sufficient to ascribe functional specialization. Based on these definitions, most categorical ERP effects reported so far seem to be preferential responses, including the N170, the M170, and the N200.

USE ROBUST STATISTICS

ERP researchers, similarly to most psychologists and neuroscientists, tend to have misguided understanding of basic statistical procedures. The most important problem is that mean, variance, *t*-tests, ANOVAs, correlations, and linear regressions are not robust to deviations from the optimal distribution parameters they assume, which can lead to substantial errors in descriptive and inferential statistics (Wilcox, 2005). Although there is no one-size-fits-all procedure, alternative techniques have been available for more than a decade and should no longer be ignored.

Using mean and variance can lead to distorted data description and poor statistical power. When data are skewed, or contain outliers, or both, the mean is a poor measure of central tendency and the variance a poor measure of dispersion. As a consequence, confidence intervals relying on mean and variance tend be too large, *t*-tests and ANOVAs tend to lack power, which means that null results from these tests are not convincing evidence of a lack of effect. Many robust alternatives to mean and variance exist, such as trimmed means and winsorized variance. Such robust measures of central tendency and dispersion behave appropriately under normality and when normality assumptions are violated. In particular, Wilcox (2005) has shown that the 20% trimmed mean performs well in many situations. Robust estimators have been used to derive robust *t*-tests and ANOVAs, some of them relying on bootstrap procedures. These modern statistical techniques are available in the R environment (Wilcox, 2005) and several Matlab toolboxes (Maris and Oostenveld, 2007; Litvak et al., 2011; Pernet et al., 2011).

Contrary to *t*-tests and ANOVAs, correlations and linear regressions tend to be biased toward false positives, which means that when a significant effect is found, its effect size might be artificially inflated and it remains unclear whether there is a true effect or whether the data suffer for instance from heteroscedasticity, i.e.,

variance inhomogeneity. Robust correlation and linear regression techniques are available in the R environment, for instance bootstrap tests under heteroscedasticity, skipped estimators and the percentage bend correlation (Wilcox, 2005).

Another important problem in ERP research is the use of ineffective multiple comparison corrections (MCCs). In ERP studies focusing on peaks, it is important to control for the number of linear contrasts to maintain the false positive error rate at the nominal level. Bonferroni correction tends to be too conservative but many other options exist, depending on the experimental design and the estimators tested (Wilcox, 2005). However, most of these MCCs, developed to deal with psychology data, are not appropriate for ERP studies in which tests are performed at many time points, electrodes or temporal frequencies. Indeed, ERP effects have temporal, spatial, and frequential correlations that need to be taken into account to provide efficient statistical tests. To take into account the temporal structure of ERP effects, a popular MCC consists in dismissing all effects that are significant for less than a certain number of time points, e.g., 15 consecutive significant *t*-tests (Rousselet et al., 2004b). This MCC and other *ad hoc* techniques should be abandoned because of poor control of false positive and false negative errors. Data driven approaches provide a better control of the false positive error rate, without sacrificing power, by taking into account the correlations inherent to ERP data. These MCCs rely on permutation and bootstrap techniques and are available in Matlab toolboxes (Maris and Oostenveld, 2007; Litvak et al., 2011; Pernet et al., 2011).

USE OPTIMIZED AVERAGING

In addition to non-robust statistics, low statistical power can result from the choice of electrodes entered into group analyses. In group analyses, ERPs are typically measured at the same electrodes in all subjects. However, these electrodes will not necessarily pick up functionally equivalent signals because even minor differences in brain fissuration or skull and scalp inhomogeneity can lead to different scalp projections. A potentially more fruitful way to do group statistics is to optimize electrodes independently in each subject, for instance by selecting the electrodes most sensitive to image and task parameters (Foxe and Simpson, 2002). Hence, this kind of optimized averaging tends to average signals that reflect common processing across subjects, whereas using the same spatial electrodes may lead to averaging signals reflecting different processes. Statistical circularity can be avoided by selecting the electrodes using an independent dataset (Liu et al., 2002), or an orthogonal condition (Kriegeskorte et al., 2009). Moreover, there is no or minimal circularity when the selected electrodes correspond to electrodes extensively reported in the literature, and when they reveal large and reliable effects in highly expected time windows (Rousselet et al., 2010). Group averaging can also be optimized by using independent components (Delorme et al., 2007) or by projecting data in a common source space (Gross et al., 2007). In source space, different locations can be studied to reveal their information content over time (Smith et al., 2009). Equivalent independent components are more difficult to cluster, although progress has been made in this direction (Onton et al., 2005; Gramann et al., 2010). These techniques have the potential to help make more meaningful comparisons across subjects and to increase statistical power.

ANALYZE ALL DATA POINTS, NOT JUST PEAKS

ERP component peaks are the main independent variables of ERP research on visual cognition. It is not clear why ERP peaks have such a special status, except their ease of measurement and a history of cumulated data. Indeed, an ERP component is not equivalent to a functional brain component (Luck, 2005), and there is very little evidence supporting the implicit belief that peak amplitudes and latencies convey two independent sources of information. A peak latency difference implies an amplitude difference that starts before the peak; latency and amplitude effects are therefore confounded. One study has nevertheless suggested a link between peak latency and information accumulation speed: the N170 to faces peaks when diagnostic information has been integrated (Schyns et al., 2007). Therefore, peaks might indicate the outputs of brain mechanisms rather than mechanisms themselves. It nevertheless remains to determine if this finding applies to other tasks and object categories. There is also evidence supporting the view that information integration starts at the transition between peaks (Schyns et al., 2007; Rousselet et al., 2008b) and that ERP sensitivity to visual information changes with age, following a temporal continuum that ignores ERP peaks (Rousselet et al., 2009, 2010); that is, in some subjects, maximum information sensitivity occurs between peaks. Finally, measuring peaks cannot be justified solely by the need to compare with the existing literature, because of the poor descriptive and inferential statistics in the field, making comparisons across studies difficult. Overall, there is no justification for limiting analyses to peaks or time windows of interest and throwing away the rest of the data. A systematic approach is thus necessary: analyzing all time points to reveal the complete time course of the effects. This systematic approach requires an adequate group averaging (Use Optimized Averaging) and a proper control for multiple comparisons (Use Robust Statistics).

USE DESCRIPTIVE STATISTICS AND MEANINGFUL FIGURES

Most ERP studies report only F, T, and p values, figures limited to data averaged across subjects, and no descriptive statistics – this poor standard makes it impossible for readers to evaluate the effects and to compare them across studies. Result descriptions must be improved. In addition to F, T, and p values, report effect sizes or confidence intervals around the effects, or both. Do not round p values and do not use a star system to mark, e.g., $*p < 0.05$, $**p < 0.01$ – p values are not error rates or effect sizes. Instead, report the exact p values to let readers make their own mind about the results. Provide the full time course of the important effects, whether they are significant or not. For instance, if two conditions are compared, plot the time course of the difference and a confidence interval around the difference. Further, plot the y-axes correctly: negativity down and positivity up – inverting the axes makes ERP research look silly to researchers from other fields. Show individual subjects' results to complement figures of group data. Illustrate effect sizes using boxplots and scatterplots, so that readers can appreciate how many subjects show an effect, and the shape of the data distribution. In essence, show the data in details, at least for the main results of an article.

PROVIDE DATA DRIVEN INTERPRETATIONS

Showing the data in sufficient details for readers to assess them will help progress in ERP research of object processing. Progress will be even stronger once another recurrent problem is tackled:

the large gap between the result and the discussion sections of publications. With too much focus on story telling, rather than on the data, one runs the risk of writing lengthy discussions about small effects, non-existent effects, or even things that were not actually quantified, such as mechanisms. Indeed, most researchers, including the authors, have been guilty of over-interpreting significant ERP effects. Over-interpretations stem mostly from an obsession for p values and the tendency to discuss any effect $p < 0.05$. Better statistics and better illustrations are a first step to go beyond the description of binary group effects. Balanced data interpretations must take effect sizes into account and compare them across studies. Individual differences should also be highlighted, to avoid unwarranted general conclusions – for instance a significant group effect tells nothing about the number of subjects showing that effect, which could be surprisingly low [Rousselet et al., 2010; Modeling single-trial ERP reveals modulation of bottom-up face visual processing by top-down task constraints (in some subjects), in revision; Rousselet et al., 2008a].

To increase the quality of data interpretations, ERP researchers need to learn about the limitations of null hypothesis significance testing (Goodman, 1999; Wagenmakers, 2007). Among many problems, p values cannot be used to weigh the importance of an effect because they are calculated under the null hypothesis, H0. In addition, "marginally significant effects" do not exist: the false positive error rate must be decided before the experiment is run and cannot be re-adjusted after looking at the data; because p values are not accurate, even if you use robust statistics, in practice it might be impossible to dissociate, e.g., $p = 0.04$ from $p = 0.06$. Describing marginally significant effects also puts you in the awkward position of describing a threshold for marginal significance – will it be 0.06, 0.07, 0.08? Given the large number of tests performed in a typical study, it is safe to ignore effects with p values close to but larger than 0.05. Especially, readers should treat significant effects with weak effect sizes with caution, because significant effects can be guaranteed under H0 using dishonest subject sampling strategies, for instance by performing a new test after each subject is tested (Wagenmakers, 2007). Hence, weak and unexpected effects should be interpreted more often as false positive errors, instead of engaging into long discussions about discrepancies among studies. However, if an effect is expected, or it is important to demonstrate that an effect is absent, p values are of no use: robust parametric statistics or Bayesian statistics and detailed figures are needed. Moreover, p values give no information about the probability of replicating an effect (Miller, 2009). The reliability of an effect can be shown only by replicating the experiment (Reliability of ERP and single-trial analyses, in revision; Rousselet et al., 2010). Finally, p values are undefined in many laboratory situations, for instance if no clear rule was established to stop subject recruitment (Wagenmakers, 2007).

In addition to understanding p values and taking effect sizes into account, it is essential to limit interpretations to what was actually studied, and how. In most experiments, the mean of a component peak is studied using an ANOVA. Other measures of central tendency and other tests might reveal different effects; other effects might be located somewhere else in the distribution (e.g., lower or upper quartiles); changes might occur in the shape of the

distribution (e.g., skewness and kurtosis). Therefore conclusions should be limited to what was measured and not generalized to entire data distributions.

Finally, we all talk about mechanisms, a term often found in titles and abstracts of articles using ERP, fMRI, and other techniques – see for instance (VanRullen and Thorpe, 2001; Carmel and Bentin, 2002; Rossion et al., 2002; Heekeren et al., 2004; Peelen et al., 2009). However, it is not clear what is gained by using this term because it can be substituted in almost all occurrences with "brain activity" or "processes." More importantly, it gives the misleading impression that researchers are studying brain mechanisms; as disappointing as it might be, most people working on visual cognition, including the authors, have not yet actually studied brain mechanisms directly. To study brain mechanisms, we need, similarly to specific, selective, and preferential brain responses, to provide clear definitions of mechanisms. At minima, we can define a mechanism as a process by which information is transformed. Therefore, to describe a mechanism, one needs to quantify the information content of brain activity at two stages of visual processing and specify how one goes from one state to the next, i.e., describe the transition states (Smith et al., 2007; Schyns et al., 2009a). Based on this simple definition, it is easy to see how animal electrophysiology has contributed to our understanding of brain mechanisms, by describing how photoreceptors, horizontal, and bipolar cells contribute to the formation of center-surround receptive fields in the retina and how the output of LGN cells is integrated to form the receptive fields of simple cells in the primary visual cortex. At a more integrated level, brain imaging and in particular ERPs can be useful to study mechanisms implemented in large neuronal populations – which one might call cognitive mechanisms. Indeed, many brain-imaging studies have used well-controlled stimuli and tasks, and can therefore describe to some extend the information or cognitive task that modulates brain activity. Nevertheless, such studies do not provide information about transition states; instead they constrain the range of possible mechanisms but do not study them explicitly. In fact, very few brain-imaging studies have provided explicit descriptions of brain mechanisms. Regarding ERP studies, very few of them have described the task-relevant information content of ERPs, a prerequisite to the description of mechanisms (e.g., Philiastides et al., 2006; Ratcliff et al., 2009; Schyns et al., 2009a; van Rijsbergen and Schyns, 2009). Even fewer studies have described explicit mechanisms (Smith et al., 2007; Schyns et al., 2009b). In an important publication, Smith et al. (2007) provided the most detailed description of a mechanism based on ERP data. First, they measured the single-trial ERP sensitivity to local facial information in different temporal frequency bands and at different time points, while subjects were engaged in discrimination tasks. This analysis revealed a succession of information processing states, including early mandatory sensitivity to the eye contralateral to occipital–temporal electrodes, followed by task-dependent sensitivity to diagnostic features – the eyes in a gender task, the mouth in an expression task. Then, they quantified state transitions, which were summarized in tables describing the conditional probabilities of transitions between information states. These tables, or stochastic automata, are the best example of mechanisms from the ERP literature. Because the vast majority of ERP studies do not quantify information content and transitions between information states, they should thus limit their discussions to ERP differences.

MODERN TOOLS TO ADDRESS A MODERN RESEARCH AGENDA

Implementing the changes described above would lead to great improvements in the research output from the literature. These improvements will nevertheless be insufficient to contribute significantly to models of visual cognition. Indeed, a study using well-controlled stimuli, several tasks and robust statistics will still not be able to quantify the information content of brain states, transitions between brain states, and how they lead to decision-making (Schyns et al., 2009a; Schyns, 2010). This can only be accomplished by mapping systematically the relationship between stimulus space, brain activity and behavior, and how task constraints affect this mapping (Pernet et al., 2007). The information content of brain states can be revealed using reverse-correlation techniques and statistical modeling approaches, by determining what global and local image properties modulate single-trial ERPs (Liu et al., 2009; Schyns et al., 2009a; Pernet et al., 2011). Transitions between brain states can be established by quantifying the probabilities of ERP sensitivity to one image feature at time t_{+1} given the sensitivity at time t (Smith et al., 2007). This ambitious research program would require abandoning the main tools of ERP research – group statistics and categorical designs. Averaged ERPs can be informative but their analyses should be focused primarily on single subjects: group averages are an abstraction that does not necessarily reflect the brain dynamics of single subjects. If group results are analyzed, it is essential to use robust statistics and to study the entire time course of brain activity (Analyze All Data Points, Not Just Peaks); it is also important to select adequate electrodes (Use Optimized Averaging). Group and single-subject analyses based on categorical (factorial) designs can be useful to constrain cognitive models. However, to study the information content of brain activity and its transformations, single-trial analyses and parametric designs are mandatory. Indeed, the brain is doing its job on each trial of an experiment, and our ultimate goal should be to understand single-trial brain activity, not activity averaged within or across subjects. More importantly, crucial information is available in the variability across trials. To understand this variability, we need to stimulate the visual system with parametrically manipulated stimuli (or at least identify task-relevant – diagnostic – features from each stimulus), to be able to establish statistical links between image properties and brain activity. Parametric designs are growing in popularity, extending the psychophysics approach to ERP research, to reveal how global and local image properties modulate responses from the visual system (Jemel et al., 2003; Tanskanen et al., 2005; Philiastides and Sajda, 2006; Smith et al., 2007; Rousselet et al., 2008b; Scholte et al., 2009; van Rijsbergen and Schyns, 2009). Single-trial analyses are used to map the relationships between image properties, brain activity and behavior with unprecedented details (Smith et al., 2006, 2007; Philiastides et al., 2006; Schyns et al., 2007; Liu et al., 2009; Ratcliff et al., 2009). For instance, in a series of experiments using single-trial analyses, Philiastides et al. described a cascade of events, revealing, first, task-independent activity related to the perceptual encoding of object categorical information, second, a time-window

sensitive to task instructions, third, a time-window related to post-sensory evidence available for decision-making (Philiastides and Sajda, 2006; Philiastides et al., 2006; Ratcliff et al., 2009). This timing of events cannot be revealed by group analyses, because the information is available from the trial-to-trial variability in brain activity, and how it relates to behavioral performance.

This line of research has recently been extended to brain source space, revealing for the first time where and when ERPs reflect task-relevant information (Smith et al., 2009). The potential offered by modern techniques of source localization and intracranial recordings, in conjunction with the present research agenda, remains untapped.

FUTURE DEVELOPMENTS

One of the next challenges will be to produce and test idiosyncratic, single-subject models of visual processing. To illustrate, imagine testing subjects with many well-controlled images, in which certain properties are parametrically manipulated. Using reverse correlations or a Generalized Linear approach, one can then derive formal models linking task constraints and image properties to single-trial ERP amplitude (Pernet et al., 2011). These models could be then tested by bringing subjects back in the lab and showing them new image categories, asking them to perform new behavioral tasks and by manipulating new image properties. Crucially, emphasis on model testing will force

researchers to make quantifiable predictions, making predictions of the sort condition A > condition B a thing of the past. ERP models might also be instrumental in giving more rigorous tests of theories of visual cognition. These developments will help ERP research of visual processing reach the standards of some animal electrophysiologists (Rust and Movshon, 2005). These developments will also give us a better understanding of individual differences in healthy, diseased, young, and old brains. For instance, normative models of the visual system might help us identify idiosyncratic differences in visual processing, and tease apart healthy from non-healthy brain development, by providing rich dissociation tools. Finally, in the more distant future, we will be able to create models of visual processing that integrate information across spatial scales, from local field potentials generated in cortical columns to surface potentials, by combining animal and human data with mathematical approaches (Deco et al., 2008).

ACKNOWLEDGMENTS

REFERENCES

Carmel, D., and Bentin, S. (2002). Domain specificity versus expertise: factors influencing distinct processing of faces. *Cognition* 83, 1–29.

Deco, G., Jirsa, V. K., Robinson, P. A., Breakspear, M., and Friston, K. (2008). The dynamic brain: from spiking neurons to neural masses and cortical fields. *PLoS Comput. Biol.* 4, e1000092. doi: 10.1371/journal.pcbi.1000092

Delorme, A., Westerfield, M., and Makeig, S. (2007). Medial prefrontal theta bursts precede rapid motor responses during visual selective attention. *J. Neurosci.* 27, 11949–11959.

DiCarlo, J. J., and Cox, D. D. (2007). Untangling invariant object recognition. *Trends Cogn. Sci. (Regul. Ed.)* 11, 333–341.

Fodor, J. (2001). *The Mind Doesn't Work that Way: The Scope and Limits of Computational Psychology*. Cambridge, MA: MIT Press.

Foxe, J. J., and Simpson, G. V. (2002). Flow of activation from V1 to frontal cortex in humans. A framework for defining "early" visual processing. *Exp. Brain Res.* 142, 139–150.

Gaspar, C. M., and Rousselet, G. A. (2009). How do amplitude spectra influence rapid animal detection? *Vision Res.* 49, 3001–3012.

Goodman, S. N. (1999). Toward evidence-based medical statistics. 1: the

P value fallacy. *Ann. Intern. Med.* 130, 995–1004.

Gramann, K., Gwin, J. T., Bigdely-Shamlo, N., Ferris, D. P., and Makeig, S. (2010). Visual evoked responses during standing and walking. *Front. Hum. Neurosci.* 4:202. doi: 10.3389/fnhum.2010.00202

Gross, J., Schnitzler, A., Timmermann, L., and Ploner, M. (2007). Gamma oscillations in human primary somatosensory cortex reflect pain perception. *PLoS Biol.* 5, e133. doi: 10.1371/journal.pbio.0050133

Heekeren, H. R., Marrett, S., Bandettini, P. A., and Ungerleider, L. G. (2004). A general mechanism for perceptual decision-making in the human brain. *Nature* 431, 859–862.

Honey, C., Kirchner, H., and VanRullen, R. (2008). Faces in the cloud: Fourier power spectrum biases ultrarapid face detection. *J. Vis.* 8, 1–13.

Itier, R. J., and Taylor, M. J. (2004). N170 or N1? Spatiotemporal differences between object and face processing using ERPs. *Cereb. Cortex* 14, 132–142.

Jemel, B., Schuller, A. M., Cheref-Khan, Y., Goffaux, V., Crommelinck, M., and Bruyer, R. (2003). Stepwise emergence of the face-sensitive N170 event-related potential component. *Neuroreport* 14, 2035–2039.

Joyce, C., and Rossion, B. (2005). The face-sensitive N170 and VPP components manifest the same brain processes: the effect of reference

electrode site. *Clin. Neurophysiol.* 116, 2613–2631.

Knebel, J. F., Toepel, U., Hudry, J., le Coutre, J., and Murray, M. M. (2008). Generating controlled image sets in cognitive neuroscience research. *Brain Topogr.* 20, 284–289.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., and Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nat. Neurosci.* 12, 535–540.

Litvak, V., Mattout, J., Kiebel, S., Phillips, C., Henson, R. N., Kilner, J., Barnes, G., Oostenveld, R., Daunizeau, J., Flandin, G., Penny, W., and Friston, K. (2011). EEG and MEG data analysis in SPM8. *Comput. Intell. Neurosci.* doi:10.1155/2011/852961

Liu, H., Agam, Y., Madsen, J. R., and Kreiman, G. (2009). Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron* 62, 281–290.

Liu, J., Harris, A., and Kanwisher, N. (2002). Stages of processing in face perception: an MEG study. *Nat. Neurosci.* 5, 910–916.

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: MIT Press.

Macé, M. J.-M., Thorpe, S. J., and Fabre-Thorpe, M. (2005). Rapid categorization of achromatic natural scenes: how robust at very low contrasts? *Eur. J. Neurosci.* 21, 2007–2018.

Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190.

Miller, J. (2009). What is the probability of replicating a statistically significant effect? *Psychon. Bull. Rev.* 16, 617–640.

Onton, J., Delorme, A., and Makeig, S. (2005). Frontal midline EEG dynamics during working memory. *Neuroimage* 27, 341–356.

Peelen, M. V., Fei-Fei, L., and Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature* 460, 94–97.

Pernet, C., Schyns, P. G., and Demonet, J. F. (2007). Specific, selective or preferential: comments on category specificity in neuroimaging. *Neuroimage* 35, 991–997.

Pernet, C. R., Chauveau, N., Gaspar, C., and Rousselet, G. A. (2011). LIMO EEG: a toolbox for hierarchical linear modeling of eletroencephalographic data. *Comput. Intell. Neurosci.* 2011, doi:10.1155/2011/831409

Philiastides, M. G., Ratcliff, R., and Sajda, P. (2006). Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram. *J. Neurosci.* 26, 8965–8975.

Philiastides, M. G., and Sajda, P. (2006). Temporal characterization of the neural correlates of perceptual

decision making in the human brain. *Cereb. Cortex* 16, 509–518.

Puce, A., Allison, T., and McCarthy, G. (1999). Electrophysiological studies of human face perception. III: Effects of top-down processing on face-specific potentials. *Cereb. Cortex* 9, 445–458.

Ratcliff, R., Philiastides, M. G., and Sajda, P. (2009). Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6539–6544.

Rossion, B., Curran, T., and Gauthier, I. (2002). A defense of the subordinate-level expertise account for the N170 component. *Cognition* 85, 189–196.

Rousselet, G. A., Gaspar, C. M., Pernet, C. R., Husk, J. S., Bennett, P. J., and Sekuler, A. B. (2010). Healthy aging delays scalp EEG sensitivity to noise in a face discrimination task. *Front. Psychol.* 1:19. doi: 10.3389/fpsyg.2010.00019

Rousselet, G. A., Husk, J. S., Bennett, P. J., and Sekuler, A. B. (2005). Spatial scaling factors explain eccentricity effects on face ERPs. *J. Vis.* 5, 755–763.

Rousselet, G. A., Husk, J. S., Bennett, P. J., and Sekuler, A. B. (2008a). Time course and robustness of ERP object and face differences. *J. Vis.* 8, 1–18.

Rousselet, G. A., Pernet, C. R., Bennett, P. J., and Sekuler, A. B. (2008b). Parametric study of EEG sensitivity to phase noise during face processing. *BMC Neurosci.* 9, 98. doi: 10.1186/1471-2202-9-98.

Rousselet, G. A., Husk, J. S., Pernet, C. R., Gaspar, C. M., Bennett, P. J., and Sekuler, A. B. (2009). Age-related delay in information accrual for faces: evidence from a parametric, single-trial EEG approach. *BMC Neurosci.* 10, 114. doi: 1410.1186/1471-2202-1410-1114

Rousselet, G. A., Macé, M. J.-M., Thorpe, S. J., and Fabre-Thorpe, M. (2007). Limits of ERP differences in tracking object processing speed. *J. Cogn. Neurosci.* 19, 1–18.

Rousselet, G. A., Thorpe, S. J., and Fabre-Thorpe, M. (2004a). How parallel is visual processing in the ventral pathway? *Trends Cogn. Sci. (Regul. Ed.)* 8, 363–370.

Rousselet, G. A., Thorpe, S. J., and Fabre-Thorpe, M. (2004b). Processing of one, two or four natural scenes in humans: the limits of parallelism. *Vision Res.* 44, 877–894.

Rust, N. C., and Movshon, J. A. (2005). In praise of artifice. *Nat. Neurosci.* 8, 1647–1650.

Scholte, H. S., Ghebreab, S., Waldorp, L., Smeulders, A. W., and Lamme, V. A. (2009). Brain responses strongly correlate with Weibull image statistics when processing natural images. *J. Vis.* 9, 21–15.

Schyns, P. G. (2010). Grand challenges in perception science: modeling the future. *Front. Psychol.* 1:10. doi: 10.3389/fpsyg.2010.00010

Schyns, P. G., Gosselin, F., and Smith, M. L. (2009a). Information processing algorithms in the brain. *Trends Cogn. Sci. (Regul. Ed.)* 13, 20–26.

Schyns, P. G., Petro, L. S., and Smith, M. L. (2009b). Transmission of facial expressions of emotion co-evolved with their efficient decoding in the brain: behavioral and brain evidence. *PLoS ONE* 4, e5625. doi: 10.1371/journal.pone.0005625

Schyns, P. G., Petro, L. S., and Smith, M. L. (2007). Dynamics of visual information integration in the brain for categorizing facial expressions. *Curr. Biol.* 17, 1580–1585.

Smith, M. L., Fries, P., Gosselin, F., Goebel, R., and Schyns, P. G. (2009). Inverse mapping the neuronal substrates of face categorizations. *Cereb. Cortex* 19, 2428–2438.

Smith, M. L., Gosselin, F., and Schyns, P. G. (2006). Perceptual moments of conscious visual experience inferred from oscillatory brain activity. *Proc. Natl. Acad. Sci. U.S.A.* 103, 5626–5631.

Smith, M. L., Gosselin, F., and Schyns, P. G. (2007). From a face to its category via a few information processing states in the brain. *Neuroimage* 37, 974–984.

Tanskanen, T., Nasanen, R., Montez, T., Paallysaho, J., and Hari, R. (2005). Face recognition and cortical responses show similar sensitivity to noise spatial frequency. *Cereb. Cortex* 15, 526–534.

van Rijsbergen, N. J., and Schyns, P. G. (2009). Dynamics of trimming the content of face representations for categorization in the brain. *PLoS Comput. Biol.* 5, e1000561. doi: 10.1371/journal.pcbi.1000561

VanRullen, R., and Thorpe, S. J. (2001). The time course of visual processing: from early perception to decision-making. *J. Cogn. Neurosci.* 13, 454–461.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychon. Bull. Rev.* 14, 779–804.

Wilcox, R. R. (2005). *Introduction to Robust Estimation and Hypothesis Testing*, 2nd Edn. New York, NY: Academic Press.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., and Tanaka, J. W. (2010). Controlling low-level image properties: the SHINE toolbox. *Behav. Res. Methods* 42, 671–684.