
An appropriate use of *iMap* produces correct statistical results: a reply to McManus (2013) “*iMAP* and *iMAP2* produce erroneous statistical maps of eye-movement differences”

Sébastien Miellet, Junpeng Lao, Roberto Caldara§

Department of Psychology, University of Fribourg, Switzerland; e-mail: roberto.caldara@unifr.ch
Received 23 December 2013, in revised form 11 May 2014

Abstract. McManus (2013, *Perception*, **42**, 1075–1084) contends the validity of the statistical approach adopted in previous versions of *iMap* (namely, *iMap* and *iMap2*; Caldara & Miellet, 2011, *Behavior Research Methods*, **43**, 864–878), casts doubts on earlier results obtained with the toolbox, and offers an altered version of the code. Here we dispute these claims and argue that while some of the arguments put forward are valid, McManus’s conclusions are misleading, since they are based on a *partial* use of the toolbox. Moreover, we compared *iMap* with the alternative code offered by McManus and objectively demonstrate that McManus’s approach is underpowered and flawed. *iMap* offers an appropriate and effective alternative to the commonly used regions of interest approach for statistical analyses of eye-movement data.

Keywords: eye movements, mapping, *iMAP*

1 Introduction

In 2011 we released *iMap*, a freely available open-source Matlab toolbox dedicated to data-driven, robust statistical mapping of eye-movement data (Caldara & Miellet, 2011). Our approach was strongly inspired by the development of open-source toolboxes in neuroimaging, such as SPM (Friston, Worsley, Frackowiak, Mazziotta, & Evans, 1994), EEGLab (Delorme & Makeig, 2004), and Fieldtrip (Oostenveld, Fries, Maris, & Schoffelen, 2011). Open-source toolboxes are constantly updated and improved thanks to user feedback, comments, and contributions. This approach and its scientific philosophy permit constructive and reactive updates to the software made necessary by continuous developments in statistics, novel theoretical and practical interests, types of data, methodology, and equipment.

The initial main goal of *iMap* was to offer solutions for data-driven analysis of eye movements, inspired by methods in functional magnetic resonance imaging. We were mainly aiming to avoid the use of subjective a priori defined regions of interest (ROIs), as discussed in the original paper (Caldara & Miellet, 2011). Facing the analyses of our own eye-movement dataset, the best strategy we found at the time (*iMap1*) was to compute image statistics on the fixation map averaged across observers in order to isolate data-driven fixation clusters; and then, in a separate analysis, to assess interobserver variability and to compute the statistical significance of group differences on individual data extracted by *iMap* from the fixation clusters (Blais, Jack, Scheepers, Fiset, & Caldara, 2008). In his paper McManus presents two valid limitations related to this approach (used in *iMap* from version 1 to version 2.1; hereafter *iMap 1–2.1*). Firstly, graphical outputs were not accounting for interobserver variability or number of participants. Secondly, because we normalized the fixation maps by the variability of their observer-averaged pixels, instead of normalizing by their per pixel variability across observers, spurious data-driven fixation clusters could show up on data with high observer variability, such as the randomized datasets of McManus. In short, *iMap 1–2.1* graphical outputs were revealing data-driven fixation clusters, isolated as significant across pixels in the

observer-averaged maps according to the random-field theory and the pixel test (Adler, 1981; Chauvin, Worsley, Schyns, Arguin, & Gosselin, 2005). Hence McManus is right when saying that, with particular data distributions, contrast maps might have revealed fixation clusters that were not statistically significant across observers. However, using the full possibilities of *iMap* 1–2.1 (ie statistical analyses on the full eye-movement data, but selected from the data-driven fixation clusters only) allowed users to also assess pixel-wise statistical significance with regard to the interobserver variability. Therefore, here we dispute the idea according to which using *iMap* 1–2.1 ineluctably leads to erroneous conclusions.

In addition, McManus attempted to address those limitations by offering a modified version of our code. Problematically, as demonstrated below, the approach put forward by McManus does not produce valid statistical fixation maps. In contrast, our new version of *iMap* (*iMap* 3), which adopts an original statistical approach, provides a self-contained statistical graphical analysis: the data-driven fixation clusters represent the correct rejection of the null hypothesis across observers. Consequently, the new version of *iMap* (version 3) simplifies the interpretation of the statistical maps and, overall, properly addresses the aforementioned limitations. In the following sections we will in turn contend McManus's criticisms and present limitations of his approach.

2 *iMap* 1–2.1

On the basis of the limitations presented above, McManus distrusts previous results obtained with *iMap* 1–2.1. However, MacManus did not exploit all of the possibilities offered by the toolbox. *iMap* not only generates heat maps and isolates data-driven fixation clusters. It also gives access to various eye-tracking measures within these data-driven clusters—measures that are not averaged across observers. McManus ignores these numerical outputs and focuses exclusively on the graphical outputs, thus making inconclusive any assumptions about *iMap* 1–2.1 validity. In contrast, we used those numerical reports, as they allowed the extraction of the raw or normalized data inside the detected clusters for each participant to subsequently test statistically the robustness of an effect across observers. This is the approach we applied from our first study (Blais et al., 2008) up to the more recent work (eg Mielllet, Caldara, & Schyns, 2011). Altogether, this statistical approach ensured firm conclusions could be established, which would have been similar if we had used more conventional *arbitrary* ROIs. Moreover, the cultural fixation bias we observed in face processing has been replicated in multiple studies from our lab (Blais et al., 2008; Caldara, Zhou, & Mielllet, 2010; Kelly, Mielllet, & Caldara, 2010; Kelly et al., 2011; Mielllet, He, Zhou, Lao, & Caldara, 2012; Mielllet, Vizioli, He, Zhou, & Caldara, 2013). It is then unlikely that the same false positive was systematically found on identical locations in these studies. More importantly and objectively, the same bias has also been observed in several independent studies carried out by other labs using different methods (eg figure 3 in Kita et al., 2010 or figure 3 in Watanabe, Matsuda, Nishioka, & Namatame, 2011), as well as in infants (Fu, Hu, Wang, Quinn, & Lee, 2012; Liu et al., 2011).⁽¹⁾ In several of our studies the graphical contrast maps outputted by *iMap* 1–2.1 revealed fixation clusters that we did not interpret as representing genuine effects because analysis of fixation durations across participants revealed an absence of statistical significance (see figure 4 in Mielllet, Zhou, He, Rodger, & Caldara, 2010 or figure 2 in Mielllet et al., 2012). To conclude on this point, in general, if one distrusts the group statistics performed on individual data extracted from data-driven clusters or ROIs (which is the conventional approach in the eye-movement literature), then not only results obtained with *iMap* 1–2.1 should be questioned, but practically all previous findings in the field.

⁽¹⁾There are several other studies reporting a central fixation during face processing in Eastern observers with a variety of experimental designs. However, we felt that for the current purpose it was not appropriate to review them all.

Recently, we became aware of the risk of double-dipping in using our approach (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009) and addressed this issue in Miellet et al. (2013), where we introduced a bootstrapped split-half verification method in order to rule out this problem. Note that *iMap3* does not present this risk of double-dipping, as the maps are representing self-contained statistical analyses.

Obviously, as for any toolbox or software, our position is that users should be fully responsible for their use (or misuse) of the tool. Thus, while being as transparent and explicit as possible in the manual and open-source code, being responsive, and offering useful support and advice to users, we cannot bear responsibility for potential misinterpretation of the results. Nonetheless, we must admit that our original paper about *iMap* could have been more explicit on this matter. We indeed *implicitly* assumed that users would make full use of the toolbox potentialities (ie computing group statistics on the basis of individual data extracted from the data-driven clusters), conforming to the analysis pipeline we adopted in our empirical work.

It is worth noting that interpreting graphical outputs is now more straightforward in the current version of *iMap* (from *iMap3*), as it provides a self-contained statistical graphical analysis. *iMap3* has been formally validated and already used in a recent study (Miellet, Caldara, Raju, Gillberg, & Minnis, 2014). It would nonetheless be naive to think that there is no room for improvement in the latest version of the toolbox. As a matter of fact, we are currently testing and validating several crucial developments that will be released in the next version of *iMap*.

3 McManus's approach

McManus offers an altered version of the *iMap* 1–2.1 code—a version that was not validated despite the author's claim of a necessity to validate any released method. McManus performed a comparison between his approach and *iMap* 1–2.1. Firstly, McManus's approach is by far too conservative from a statistical point of view, requiring an unrealistic number of data points in eye-movement research (at least 100 000 simulated fixations) to capture statistical differences. More critically, it also reveals an erroneous spatial extent of the effects and generates false positives.

These limitations are clearly apparent in McManus's simulation comparing his approach with *iMap2*, in the *absence* and the *presence* of an effect. Firstly, according to this simulation, *iMap2* is oversensitive and shows spurious noisy clusters in the absence of an effect [ie no difference (A vs B) comparison in figure 7, top row, in McManus's paper]. In contrast, McManus's approach does not reveal significant areas in the absence of an effect. However, we expect that statistics performed on individual fixation durations extracted from *iMap2* data-driven fixation clusters would not reveal any significant effect. McManus does not report this information in his paper.⁽²⁾ Secondly, figure 7 in McManus's paper also clearly shows that *iMap2* *accurately* and *consistently* estimates the spatial extent of the true simulated effect [ie difference comparison (A vs B) in figure 7, bottom rows]. Hence, *iMap2* data-driven fixation clusters could be exploited to extract individual data and perform statistical analyses across observers. In contrast, even with a simulated effect, McManus's approach requires no less than 100 participants (1000 fixations each, 100 000 data points) to reveal significant 'genuine' effects, but also more problematically false positives. Moreover, it is clear from the figure that the spatial extent of the 'genuine' effect is largely overestimated (ie figure 7: the 50+50 and 100+100 simulations).

This observation suggests that—with noisier, but more realistic, experimental data—an even larger number of participants would be required to reveal such spatially overestimated effects. In fact, we tried McManus's code on several of our datasets, and the results are

⁽²⁾We do not have access to McManus simulated data and could not verify this aspect formally.

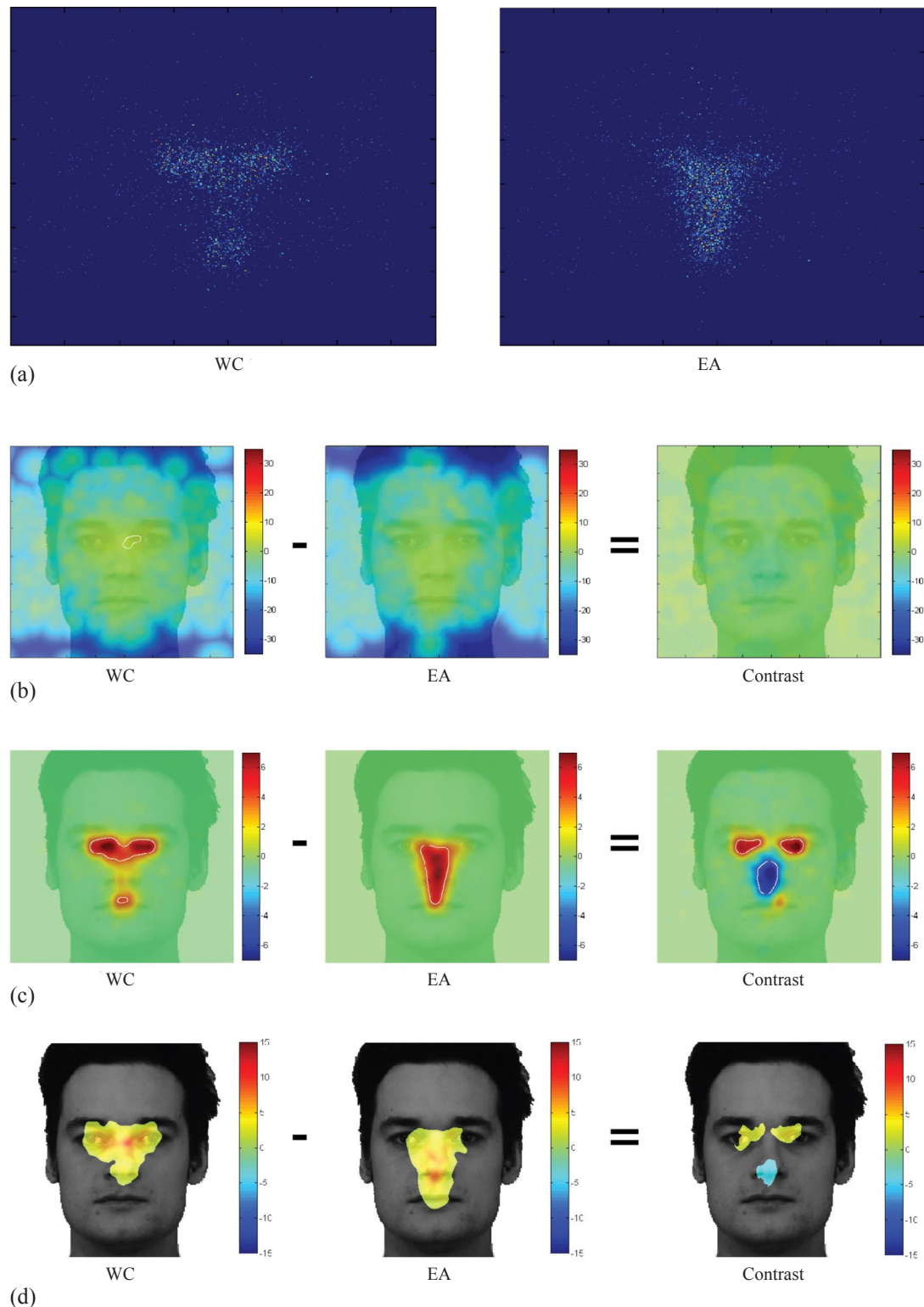


Figure 1. [In colour online, see <http://dx.doi.org/10.1068/p7682>] (a) Raw data for Western Caucasian (WC, on the left) and East Asian (EA, on the right) observers followed by (b) statistical fixations (first and second columns for WC and EA, respectively) and contrast maps (third column) generated by McManus's approach, (c) *iMap2.1*, and (d) the latest version, *iMap3*. For each toolbox the colour scale is centered on 0 and the range is set according to the largest absolute value computed by the toolbox in any of the three maps.

far from convincing. For an illustrative purpose, we ran McManus's code on one of the datasets (face example) presented in the *iMap3* manual (downloadable on our website: <http://www.unifr.ch/psycho/ibmlab>). Figure 1 shows a representation of the raw data (a) for the two groups of observers (Westerners and Easterners). Each dot represents a fixation location (without smoothing or normalization), and the cumulated fixation duration is colour coded (warm colours for longer fixations). The figure also shows both group fixation biases and the difference map for McManus's approach (b), *iMap2.1* (c), and *iMap3* (d).

In the group maps McManus's approach generates the largest absolute statistical values (-30), in locations where the participants do not, or very rarely, look. For instance, the absence of significant effects in the East Asian observers' map should be interpreted as an absence of preferentially fixated areas for this group of observers (ie a homogeneous fixation distribution across the stimulus space). This observation is completely unfounded given the raw fixation data distribution represented above (figure 1a). The contrast map is also disconcerting. The absence of significant contrast is not only inconsistent with the contrast map and individual fixation durations reports generated by *iMap2.1*; it is also inconsistent with the self-contained statistical maps generated by *iMap3*, which adopt a completely different statistical approach. In the example of the cultural bias during free viewing of face stimuli, *iMap3* reveals a similar results pattern as *iMap2.1* and leads to the same conclusions.

It is important to mention that in his paper McManus claims that we implemented in *iMap3* "a method which essentially is that described here". We find this statement misleading, as it suggests that both approaches are similar. This is absolutely not the case; the only similarity between both methods is that they use pixel-wise statistics across participants, as is performed for comparing voxel activations in most functional magnetic resonance imaging approaches. The critical point in statistical mapping is to find a way to determine significance and effect sizes. On this matter, McManus's and *iMap3*'s approaches are fundamentally different and lead to radically divergent results and conclusions, as shown in figure 1. Contrary to *imap2icmBeta* (McManus's approach), *iMap3* does not rely on the random field theory to correct for multiple comparisons. Instead, it makes use of a bootstrapping procedure to determine the significance threshold after applying threshold-free cluster enhancement (Smith & Nichols, 2009) on *t*-values across participants for signal enhancement, as recently adopted in LIMO EEG (Pernet, Chauveau, Gaspar, & Rousselet, 2011).

4 Conclusions

We demonstrated that, despite the limitations put forward by McManus, a proper use of the *iMap 1–2.1* toolbox (ie group statistics based on individual data extracted from data-driven fixation clusters; split-half procedure to control for double-dipping) allowed researchers to draw solid conclusions. Therefore, we do not concur with McManus's doubts on the validity of previous research using *iMap 1–2.1*. We also demonstrated that McManus's current approach is statistically flawed, even with simulated data. Moreover, McManus's approach is not suited to satisfy the realistic constraints of conventional eye-movement research and to characterize and statistically test empirical findings. We challenge the author to replicate any of the effects shown in the eye-tracking literature in the last 40 years with his approach, using conventional sample sizes (typically around 30 observers) and number of fixations.

On a more positive note, *iMap3* has been released. This new version keeps the general data-driven philosophy of *iMap* (by avoiding the use of arbitrary ROIs) and implements a radically different statistical approach. Crucially, *iMap3* has been formally validated with our data distributions⁽³⁾ and has already been used in a published study (Miellel et al., 2014). We will not

⁽³⁾ Obviously, we cannot consider all the potential experimental situations and cannot anticipate all the possible data characteristics and distributions. We fully rely on users' critical feedback to inform us about unexpected problems that might appear with specific equipment, tasks, stimuli, or data structures.

describe the current version here, but the open-source code and a detailed manual are available on our website (<http://www.unifr.ch/psycho/ibmlab>). Contrary to McManus's approach, our data and simulations show that *iMap3* is an appropriate and effective method for statistical fixation mapping of eye-movement data.

References

- Adler, R. J. (1981). *The geometry of random fields*. New York: Wiley.
- Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PLoS ONE*, *3*:e3022. doi:10.1371/journal.pone.0009708
- Caldara, R., & Miellet, S. (2011). *iMap*: a novel method for statistical fixation mapping of eye movement data. *Behavior Research Methods*, *43*, 864–878.
- Caldara, R., Zhou, X., & Miellet, S. (2010). Putting culture under the 'spotlight' reveals universal information use for face recognition. *PLoS ONE* *5*(3):e9708. doi:10.1371/journal.pone.0009708
- Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., & Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *Journal of Vision*, *5*(9):1, 659–667.
- Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21.
- Friston, K. J., Worsley, K. J., Frackowiak, R. S. J., Mazziotta, J. C., & Evans, A. C. (1994). Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, *1*, 214–220.
- Fu, G., Hu, C. S., Wang, Q., Quinn, P. C., & Lee, K. (2012). Adults scan own- and other-race faces differently. *PLoS ONE*, *7*:e37688. doi:10.1371/journal.pone.0037688
- Kelly, D. J., Liu, S., Rodger, H., Miellet, S., Ge, L., & Caldara, R. (2011). Developing cultural differences in face processing. *Developmental Science*, *14*, 1176–1184.
- Kelly, D. J., Miellet, S., & Caldara, R. (2010). Culture shapes eye movements for visually homogeneous objects. *Frontiers in Psychology*, *1*, doi:10.3389/fpsyg.2010.00006
- Kita, Y., Gunji, A., Sakihara, K., Inagaki, M., Kaga, M., Nakagawa, E., & Hosokawa, T. (2010). Scanning strategies do not modulate face identification: Eye-tracking and near-infrared spectroscopy study. *PLoS ONE*, *5*:e11050. doi:10.1371/journal.pone.0011050
- Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience*, *12*, 535–540.
- Liu, S., Quinn, P. C., Wheeler, A., Xiao, N., Ge, L., & Lee, K. (2011). Similarity and difference in the processing of same- and other-race faces as revealed by eye tracking in 4- to 9-month-olds. *Journal of Experimental Child Psychology*, *108*, 180–189. doi:10.1016/j.jecp.2010.06.008
- McManus, I. C. (2013). *iMAP* and *iMAP2* produce erroneous statistical maps of eye-movement differences. *Perception*, *42*, 1075–1084. doi:10.1068/p7520.
- Miellet, S., Caldara, R., Raju, M., Gillberg, C., & Minnis, H. (2014). Disinhibited reactive attachment disorder symptoms impair social judgements from faces. *Psychiatry Research*, *215*, 747–752. doi:org/10.1016/j.psychres.2014.01.004
- Miellet, S., Caldara, R., & Schyns P. G. (2011). Local Jekyll and global Hyde: the dual identity of face identification. *Psychological Science*, *22*, 1518–1526. doi:10.1177/0956797611424290
- Miellet, S., He, L., Zhou, X., Lao, J., & Caldara, R. (2012). When East meets West: Gaze-contingent blindspots abolish cultural diversity in eye movements for faces. *Journal of Eye Movement Research*, *5*(2):4, 1–12.
- Miellet, S., Vizioli, L., He, L., Zhou, X., & Caldara, R. (2013). Mapping face recognition information use across cultures. *Frontiers in Psychology* *4*:34. doi:10.3389/fpsyg.2013.00034
- Miellet, S., Zhou, X., He, L., Rodger, H., & Caldara, R. (2010). Investigating cultural diversity for extrafoveal information use in scenes. *Journal of Vision* *10*(6):21, 1–18. doi:10.1167/10.6.21
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, article 156869. doi:10.1155/2011/156869
- Pernet, C. R., Chauveau, N., Gaspar, C. M., & Rousselet, G. G. (2011). LIMO EEG: A toolbox for hierarchical linear modeling of electroencephalographic data. *Computational Intelligence and Neuroscience*, article 831409. doi:10.1155/2011/831409

-
- Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, **44**, 83–98. doi:10.1016/j.neuroimage.2008.03.061
- Watanabe, K., Matsuda, T., Nishioka, T., & Namatame, M. (2011). Eye gaze during observation of static faces in deaf people. *PLoS ONE* **6**(2):e16919. doi:10.1371/journal.pone.0016919