Analyzing 14 years of neuroimaging display practice

University of Zurich[UZH]

UNIVERSITY OF NOTRE DAME

M. Christen (University of Zurich, Switzerland & University of Notre Dame, USA; christen@ethik.uzh.ch), D.A. Vitacco (University Hospital Zurich), L. Huber (Bergische Universität Wuppertal, Germany), P. Brugger (University Hospital Zurich), J. Harboe (Lucerne University of Applied Sciences and Arts, Switzerland)

Problem under study

Neuroimaging is a complex methodology involving the conceptual steps of experimental design, measurement, data analysis, and data presentation. Whereas the first three steps have been intensely discussed with respect to methodological issues, data presentation in neuroimaging has only rarely been investigated. We close this gap by providing data and interpretation of 14 years (1996-2009) of neuroimaging display practice in six major journals. Our focus was on the breadth of display styles, trends in standardization and potential shortcomings in the use of colors.

Methodology

Using a complete sampling approach, we created a dataset of 9179 figures from 3993 contributions published in *Annals of Neurology* and *Brain* (neurological journals), *Human Brain Mapping* and *NeuroImage* (imaging journals), and *Nature* and *Science* (broad interest journals) that contained either fMRI or PET displays of a brain. We collected information on the origins of the contributions, the software used in image analysis, image structure and complexity, and the presence or absence of numerical explanations of neural activation. We also coded all pictures with respect to the use of color scales in brain activation in order to identify different styles of brain images and evaluated their underlying regimes of data presentation. Coder reliability, data accuracy, and data completeness were carefully checked. Below is an outline of the dataset:

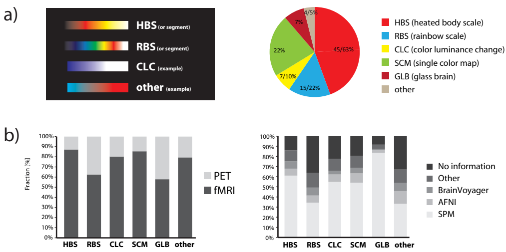
Journals	# publications (% of all public.)	# figures (%PET)
HBM	712 (57.4)	1,659 (13.0)
NeuroIm.	2,352 (42.2)	5,678 (15.3)
Annals	181 (5.9)	301 (60.5)
Brain	448 (14.4)	1,089 (46.1)
Nature	96 (0.3*)	140 (34.3)
Science	202 (0.7*)	312 (18.3)

90.6% of all images used colors, 20.2% of all images were created using PET. The temporal development of the fraction of PET images per year shows a dramatic decrease over time from 78.3% in 1996 to 9.9% in 2009 (*: upper bound).

Results in Detail: Display Style, Standardization, Shortcomings

a)

b)

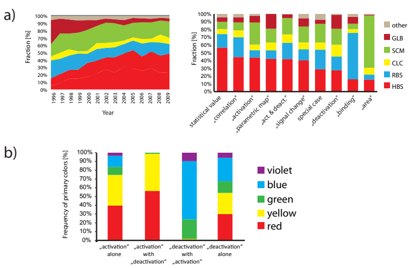


Display styles. a) Distribution of the five main display styles and the style class "others" among the data set (left side) and examples of color scales (right side): Most common is the heated body scale (HBS) in which the luminance increases from black through red, yellow and white. 44.4% of all images used this style (either the full scale or sections) and the transition red-yellow (20.4%) is the single most common color scale used. The second most common style (22.4%) is the use of single color maps (SCM) for denoting, e.g. activated regions. Third is the rainbow scale (RBS, 15.2%), where the hue is varying in the order of the spectrum (violet-blue-green-yellow-red; or sections of this sequence). About equally frequent (6.8% and 7.4%) are the use of single color luminance changes (CLC, e.g. from blue to white) and glass brains (GLB) – a characteristic display style for statistical maps using usually black or grey shades for localizing activation and displaying the brain as a mere (usually) black contour on a (usually) white background. 3.9% of all images used color scales that were different from those of the five main groups (e.g. red-blue transitions).

b) Relationship between display style and method (fMRI or PET, left) and software used (right): The largest fractions of PET images compared to fMRI images are discernible in the GLB (42.1%) and RGB (37.5%) style. Glass brains are preferably, but not exclusively produced by SPM (83.5%), whereas pictures in the RGB display style often lack information about the software used.

a)

b)

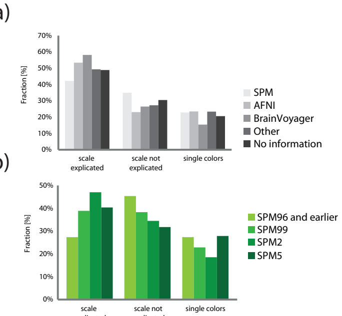


Standardization. a) Temporal development of the fraction of the main display styles (left) and fraction of the main display styles according to the reference of the color. We coded any labeling of the colors presented given either in the figure caption or in the figure itself (scale) into 10 classes: "activation" (explicit wording like "active", "increase"), "activation & deactivation" (when the wording refers both to activation and deactivation), "area" (when the wording refers to a specific area of the brain), "binding" (binding potential etc.), "correlation" (when the color scale codes for the strength of a correlation between parameters), "deactivation" (for "less active", "decrease" and the like), "parametric map" (if the wording refers to a statistical parametric map without further explications), "signal change" (for a wording that only refers to a change in the signal without further indication), special cases (e.g. color-coding of latency times), or "statistical value" (if the color scale is explicitly said to code for F, P, T, or Z values). The largest labeling group "statistical values" (31.1%) displayed a slight increasing trend (slope of linear trend approximation: +0.003), whereas the second largest group "activity" (22.8%) displayed a slight decreasing trend (slope of linear trend approximation: -0.003), which, however, cannot explain the overall trend to the HBS scale.

b) In neurophysiological brain mapping and EEG it was agreed by convention that red and yellow indicate high activity and positive polarity, whereas green and blue are used for low activity and negative polarity. For testing, whether this convention holds for neuroimaging, we counted the appearance of the basic colors red, yellow, green, blue, and violet in figures of the classes "activation" and "deactivation". A special set of figures are those that included two color-scales that coded for "activation" and "deactivation". For this set, the convention was clearly fulfilled, i.e. activation is coded almost exclusively using red and/or yellow, and deactivation is coded using blue and/or green and/or violet. However, as soon as the figures referred only to "activation" or "deactivation", the standard eroded. In particular for "deactivation", more than half of the colors mentioned were the "warm" colors yellow or red.

a)

b)



Shortcomings. A striking finding is that in 38.2% of the images that displayed neuronal activations using color scales the colors were not associated with numbers by either a scale bar or by outlining the meaning of the colors in the figure caption. If the display style "glass brain" is included in this figure (when grey shadings reflect numerical data in an unspecified way), it rises to 40.9%. Missing scale explication is more common in PET than fMRI papers (46.3% versus 39.5%). Also the software used has an influence on the explication of the scale (Fig. a). In the temporal development, a tendency (although no clear trend) to explain scales is discernible: in 1996, 47.6% of the figures did not have explicated scales; the number rose up to 64.9% in 2000 and then dropped to 35.0% in 2009. Furthermore, images in the CLC style surprisingly often lack explained scales (47.5%), whereas uncommon color scales are usually accompanied by scale explications (missing explications in 19.9%).

Fig. b) Shows the influence of SPM version on explication of the color scale: Newer versions of SPM increase the likelihood that a picture contains a scale, whereas SPM5 also increases the likelihood of using single colors (mostly red) for indicating activation.

Another shortcoming refers to the still considerable popularity of the RBS scale, although there are well-known problems associated with it: First, to some users it might not present an intuitive ordering. Second, yellow is present half way through the color scale, which means that if one is interested in depicting extreme values the middle values might interfere, since yellow has a highlighting effect being perceived as brighter than the other colors. Third, the saturation steps do not equally represent differences between numbers. Yellow has the smallest number of perceived saturation steps and users find it harder to distinguish small saturation variations for yellow than, for example, for blue.

We thank Alvin Chesham, Annina Gläser, Cécile Haldi, Christian Ineichen, Judith Kaemmer and Myriam Nösberger for their support in creating the publication database as well as the participants of the workshop "Visual Communication of Data in Neuroimaging" (Regula V. Burri, Sara I. Fabrikant, Martin H. Fischer) that took place in Zurich on June 30th and July 1st 2011 for their valuable input. This research has been supported by grant No. R-129/09 of the Cogito Foundation, Wollerau, Switzerland

Main Results

Only few image production sites and image creation software systems dominate: three countries (USA, UK, and Germany) produced 65.6% of all images; 72 software systems for brain image analysis have been identified, but 85.8% of all images were produced by three systems (SPM, AFNI, BrainVoyager). The color coding reveals a remarkably diverse use of colors. We identified five main display styles. Despite the rather diverse phenomenology with respect to the use of display styles, the temporal development shows a trend of standardization towards the heated body scale, mainly at the expense of glass brains.

A striking finding is that in 38.2% of the images that displayed neuronal activations using color scales the colors were not associated with numbers by either a scale bar or by outlining the meaning of the colors in the figure caption.

Conclusions

In summary, our results and consideration lead to the following suggestions with respect to the display practice in neuroimaging:

- The process of image production should be discussed in more detail in the methodological section of publications and include also specifications of the image post-processing software.
- If color scales are used in images, they should be clearly explicated by a scale or an appropriate description in the figure caption. If a neuroimage merely displays sites of activation or the like, single colors should be preferred.
- The discerned trend of standardization with respect to using the heated body scale or "cold colors" for increase or decrease of statistical significance should be advanced further, but be decoupled from a mere wording of "activation" and "deactivation". Either denoting the precise statistical meaning of the scale, or a more neutral wording like "signal change" would be more appropriate.
- Non-standard displays of data relations in neuroimages (e.g. correlations, latency times etc.) should be based on color scales other than HBS.
- The use of the rainbow color scale may be further restricted to applications where a quasi-standard has been established, e.g. for displaying binding potentials in PET imaging.
- Producers of imaging analysis tools should support appropriate use of colors with respect to the usability of the programs and instruction manuals.