NeuroImage 73 (2013) 265–267

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Review

On evidence, biases and confounding factors: Response to commentaries

Cibu Thomas [a,b,c,*], Chris I. Baker [a,c]

[a] Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892, USA
[b] Section on Tissue Biophysics and Biomimetics, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA
[c] Center for Neuroscience and Regenerative Medicine at the Uniformed Services University of the Health Sciences, Bethesda, MD, USA

ARTICLE INFO

Article history:
Accepted 9 November 2012
Available online 14 November 2012

Keywords:
VBM
DTI
Structural plasticity
Training
MRI

ABSTRACT

In a critical review (Thomas and Baker, 2012), we argued for caution in evaluating reports of training-dependent adult structural plasticity measured with MRI. Here, we respond to the commentaries on our review, clarifying our position and addressing some of the specific criticisms raised.

Contents

We thank the authors of the three commentaries (Draganski and Kherif, 2012; Erickson, 2012; Fields, 2012) for their thoughtful responses to our critical review of evidence for training-dependent structural plasticity in the adult brain (Thomas and Baker, 2012). Rather than respond to every issue raised, here, we focus on key agreements and disagreements and at the same time clarify our position.

The main contention of our review was the need to be cautious in evaluating the current MRI evidence for human structural plasticity due to the potential for confounding factors, small effect sizes, and the lack of clear and consistent replication. We focused on 20 published MRI studies, but it is worth noting that since our review was accepted for publication, at least nine more have been published (Bueti et al., 2012; Colom et al., 2012; Gryga et al., 2012; Mackey et al., 2012; Mårtensson et al., 2012; Sagi et al., 2012; Schlegel et al., 2012; Tang et al., 2012; Voss et al., 2012). There is certainly strong agreement on the potential importance of the reported findings, but opinions clearly differ on the assessment of the current literature.

Critically, we were not condemning or rejecting studies but providing an objective assessment of the strength of the evidence, highlighting the need for more rigorous design and analyses to confirm the findings of earlier studies and move the field forward. As Erickson (2012) notes, it is difficult to have an in-depth discussion of structural plasticity and assess the importance of factors such as duration of training, age and training paradigm, without having demonstrably robust and reliable results.

We are not using an "appeal to the majority" (Fields, 2012), but appealing to the strength of the prevailing evidence. Nor are we using a "negative study" (Thomas et al., 2009) to call "into question all the others reporting positive findings" (Fields, 2012). While we completely acknowledge the hazards of interpreting negative findings, the significance of the study by Thomas and colleagues is not so much in its null result, but in demonstrating how subtle changes in the analysis pipeline can produce dramatic changes in the results observed. Finally, we are not advocating throwing out the existing data, and suggest that the strength of the evidence could be improved by simply reanalyzing existing data using more stringent statistical criteria.

As we noted in the review, a major challenge for MRI-based imaging of structural plasticity is to explicate the underlying biological substrate. Conventional structural MRI methods like T_1-weighted

* Corresponding author at: Section on Tissue Biophysics and Biomimetics, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA. Fax: +1 301 443 7111.
E-mail address: cibu.thomas@nih.gov (C. Thomas).

1053-8119/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2012.11.009

imaging do not offer any insight on the nature of the structural change induced by training. In this respect, we fully agree with Draganski and Kherif (2012) that quantitative techniques like diffusion MRI (Jones, 2010), T_1–T_2 relaxometry (Deoni et al., 2008), and voxel-based quantification (Draganski et al., 2011) have the potential to greatly improve our understanding of structural plasticity. However, it is important to embrace these techniques with caution. Sources of measurement error need to be considered, and the reproducibility of the various quantitative measures should be firmly established (Deoni, 2010; Marenco et al., 2006). Importantly, regardless of the specific measure, the concerns we raised about design and analysis of studies still apply.

We argued for the importance of a well-matched control condition in assessing evidence for structural plasticity. Although the need for a control condition is well established in many different domains, including functional MRI, it could be argued that a control group is unnecessary for short-term studies of adult structural plasticity since changes in structure would ordinarily not be expected in healthy individuals. However, this argument makes the assumption that measurements of brain structure are accurate, reliable and noise-free. Unfortunately, this is not the case, and a control condition helps guard against any potential confounding factors such as scanner-related performance (Johansen-Berg, 2011). Specifically, MRI provides indirect measures of brain structure and is susceptible to a number of potential confounds including subject motion (Pierpaoli, 2011), cardiac pulsation (Walker et al., 2011), level of hydration (Streitbürger et al., 2012), and may also be affected by the inhomogeneities in the B1 field which can themselves change over time (Draganski and Kherif, 2012).

Fields (2012) suggests that in repeated-measure designs, each subject is its own control. But this ignores the fact that the measure of interest is the change between time points and the purpose of a control is to assess the variation that can occur in the absence of the specific intervention being tested. Fields (2012) adopts an analogy of blood pressure before and after taking a medication. But it is critical to realize that without a control condition in which blood pressure is monitored without the medication, any changes in blood pressure could reflect measurement error or any one of a number of possible confounding variables such as temperature, time of day, stress, posture, caffeine intake or activity immediately prior to testing. Well-matched control groups provide an estimate of the reliability of repeated measures and take into account any potential uncontrolled factors.

Of course, as Fields (2012) notes, if noise and artifacts are distributed randomly, this should not be a concern. However, with the relatively small sample sizes used in typical MRI studies of structural plasticity and the search across thousands of voxels, an imbalance in any of a number of factors could produce artifactual results. It is important to note that our concern is not specifically about statistical thresholds. Statistical thresholds are important and it is clear that entirely spurious effects can be observed without appropriate correction (Bennett et al., 2009, 2011). Crucially, results arising from systematic biases in uncontrolled factors are not strictly spurious but reflect real effects in the data unrelated to the manipulation of interest and may survive even conservative statistical correction (Streitbürger et al., 2012; Walker et al., 2012). Without taking into account the possibility for such confounding factors in the design and analysis of a study, their potential impact cannot be ruled out.

We argued that in a between-groups design, the critical test is the interaction between group (trained/untrained) and time point (pre/post-training). As Draganski and Kherif (2012) rightly point out, alternative approaches include a simple test between groups on the difference scores between pre- and post-test data or an ANCOVA including pre-training scores as a covariate. That said, in a group by time point ANOVA, while the main effect and interaction effect are orthogonal with equal samples per condition, prior selection of the data can dramatically distort the results. In particular, if voxels are first selected for a main effect, especially within one group only (Scholz et al., 2009; Tang et al., 2010; Woollett and Maguire, 2011), before testing the interaction, the results will be biased toward the effect in that group and the likelihood of finding an interaction will be artifactually elevated (Kriegeskorte et al., 2009). Importantly, re-analysis of existing data avoiding selection biases could address this concern and potentially demonstrate the robustness of the results.

Both Draganski and Kherif (2012) and Fields (2012) question our concern about the small effect sizes typically reported, suggesting that we are conflating “spatial resolution and spatial precision”. In both cases, the commentaries are highlighting the necessity of averaging in detecting any structural changes. We are not, as Fields (2012) implies, suggesting that values below the unit of measure are uninterpretable. We are simply pointing out that differences below the resolution of the technique gleaned from single observations in each participant are wholly dependent on averaging across participants. This dependence implies that a significant effect can be observed even if a clear majority of the participants show no change, which, combined with the sheer number of comparisons being run, suggests that caution is needed and converging evidence and replications are paramount. Given the lack of replication in the literature using even similar paradigms (May and Gaser, 2012; Thomas and Baker, 2012), this caution appears to be warranted.

In light of the concerns we raised, it is worth highlighting that some of the recently published studies adopt design and analyses along the lines we advocate. For example, in a study of possible structural changes during training for the Law School Admission Test (LSAT), difference scores between time points were tested between training and control groups, and possible confounding factors related to stress and amount of sleep were evaluated (Mackey et al., 2012). Further, an investigation of spatial layout learning in the context of a racing game compared participants trained on a single track (who could learn the layout) with participants trained on multiple different tracks (matched for all other aspects of the training) (Sagi et al., 2012). At the same time, it is important to maintain a critical outlook. For example, while subjects training for the LSAT appeared to show changes in measures of diffusivity in frontal and parietal cortex, the lack of effects in a VBM analysis fails to replicate earlier findings in a similar study of students studying for a medical exam (Draganski et al., 2006). Further, it is still appropriate to question whether changes in MRI measures that are so widespread (~5000 voxels, Mackey et al., 2012) or over a very short duration (90 min of training, Sagi et al., 2012) are believable or whether they might reflect the impact of confounding factors (Erickson, 2012).

Fields argues that, “Publication bias toward positive results is appropriate and necessary because negative findings provide no logical conclusion”, but a bias towards positive results may be detrimental for scientific progress and may come at an exorbitant cost (Bennett et al., 2009; Ioannidis, 2011) (see http://www.psychfiledrawer.org). We are proponents of removing barriers to publication (Kravitz and Baker, 2011) and publishing both positive and negative findings, but a crucial part of such a vision is an active scientific community that critically evaluates the literature.

Ultimately, the goal of our review was to critically assess the strength of the existing evidence for training-dependent structural plasticity, subjecting all studies, whether positive or negative, to the same level of scrutiny that Fields (2012) employs to one of the ‘negative’ studies. We concluded our review with the suggestion that the strongest evidence concerns structural changes in the hippocampus. This is not only because the study we highlighted (Erickson et al., 2011) adopted rigorous design and analyses but because the findings are complemented by converging evidence from cross-sectional human MRI studies (Maguire et al., 2000, 2006; Pereira et al., 2007; Thomas et al., 2012) as well as animal studies using multiple techniques including combinations of both histology and MRI (Blumenfeld-Katzir et al., 2011; Lerch et al., 2011). Focusing on the strength of the evidence and pursuing careful experimental design, rigorous data analysis, and

some healthy skepticism can help fortify our understanding of adult structural plasticity.

Acknowledgments

This work was supported by the Intramural Research Program of NIMH. Support for this work also included funding from Department of Defense in the Center for Neuroscience and Regenerative Medicine. We thank Dr. Carlo Pierpaoli for helpful comments and discussion.

References

Bennett, C.M., Wolford, G.L., Miller, M.B., 2009. The principled control of false positives in neuroimaging. Soc. Cogn. Affect. Neurosci. 4, 417–422.
Bennett, C.M., Baird, A.A., Miller, M.B., Wolford, G.L., 2011. Neural correlates of interspecies perspective taking in the post-mortem Atlantic salmon: an argument for proper multiple comparisons correction. J. Serendipitous Unexpect. Results 1, 1–5.
Blumenfeld-Katzir, T., Pasternak, O., Dagan, M., Assaf, Y., 2011. Diffusion MRI of structural brain plasticity induced by a learning and memory task. PLoS One 6, e20678.
Bueti, D., Lasaponara, S., Cercignani, M., Macaluso, E., 2012. Learning about time: plastic changes and interindividual brain differences. Neuron 75, 725–737.
Colom, R., Quiroga, M.Á., Solana, A.B., Burgaleta, M., Román, F.J., Privado, J., Escorial, S., Martínez, K., Álvarez-Linera, J., Alfayate, E., 2012. Structural changes after videogame practice related to a brain network associated with intelligence. Intelligence 40, 479–489.
Deoni, S.C.L., 2010. Quantitative relaxometry of the brain. Top. Magn. Reson. Imaging 21, 101–113.
Deoni, S.C.L., Rutt, B.K., Arun, T., Pierpaoli, C., Jones, D.K., 2008. Gleaning multicomponent T1 and T2 information from steady-state imaging data. Magn. Reson. Med. 60, 1372–1387.
Draganski, B., Kherif, F., 2012. In vivo assessment of use-dependent brain plasticity — beyond the "one trick pony" imaging strategy. NeuroImage.
Draganski, B., Gaser, C., Kempermann, G., Kuhn, H., Winkler, J., Buchel, C., May, A., 2006. Temporal and spatial dynamics of brain structure changes during extensive learning. J. Neurosci. 26, 6314–6317.
Draganski, B., Ashburner, J., Hutton, C., Kherif, F., Frackowiak, R., Helms, G., Weiskopf, N., 2011. Regional specificity of MRI contrast parameter changes in normal ageing revealed by voxel-based quantification (VBQ). NeuroImage 55, 1423–1434.
Erickson, K.I., 2012. Evidence for structural plasticity in humans: comment on Thomas and Baker (2012). NeuroImage.
Erickson, K.I., Voss, M.W., Prakash, R.S., Basak, C., Szabo, A., Chaddock, L., Kim, J.S., Heo, S., Alves, H., White, S.M., 2011. Exercise training increases size of hippocampus and improves memory. Proc. Natl. Acad. Sci. U. S. A. 108, 3017–3022.
Fields, R.D., 2012. Changes in brain structure during learning: fact or artifact? Reply to Thomas and Baker. NeuroImage.
Gryga, M., Taubert, M., Dukart, J., Vollmann, H., Conde, V., Sehm, B., Villringer, A., Ragert, P., 2012. Bidirectional gray matter changes after complex motor skill learning. Front. Syst. Neurosci. 6, 1–9.
Ioannidis, J., 2011. Excess significance bias in the literature on brain volume abnormalities. Arch. Gen. Psychiatry 68, 773–780.
Johansen-Berg, H., 2011. The future of functionally-related structural change assessment. NeuroImage 62, 1293–1298.
Jones, D.K., 2010. Precision and accuracy in diffusion tensor magnetic resonance imaging. Top. Magn. Reson. Imaging 21, 87–99.
Kravitz, D.J., Baker, C.I., 2011. Toward a new model of scientific publishing: discussion and a proposal. Front. Comput. Neurosci. 5. http://dx.doi.org/10.3389/fncom.2011.00055.
Kriegeskorte, N., Simmons, W., Bellgowan, P., Baker, C., 2009. Circular analysis in systems neuroscience: the dangers of double dipping. Nat. Neurosci. 12, 535–540.
Lerch, J.P., Yiu, A.P., Martinez-Canabal, A., Pekar, T., Bohbot, V.D., Frankland, P.W., Henkelman, R.M., Josselyn, S.A., Sled, J.G., 2011. Maze training in mice induces MRI detectable brain shape changes specific to the type of learning. NeuroImage 54, 2086–2095.
Mackey, A.P., Whitaker, K.J., Bunge, S.A., 2012. Experience-dependent plasticity in white matter microstructure: reasoning training alters structural connectivity. Front. Neuroanat. 6, 1–9.
Maguire, E., Gadian, D., Johnsrude, I., Good, C., Ashburner, J., Frackowiak, R., Frith, C., 2000. Navigation-related structural change in the hippocampi of taxi drivers. Proc. Natl. Acad. Sci. U. S. A. 97, 4398.
Maguire, E., Woollett, K., Spiers, H., 2006. London taxi drivers and bus drivers: a structural MRI and neuropsychological analysis. Hippocampus 16, 1091–1101.
Marenco, S., Rawlings, R., Rohde, G.K., Barnett, A.S., Honea, R.A., Pierpaoli, C., Weinberger, D.R., 2006. Regional distribution of measurement error in diffusion tensor imaging. Psychiatry Res. Neuroimaging 147, 69–78.
Mårtensson, J., Eriksson, J., Bodammer, N.C., Lindgren, M., Johansson, M., Nyberg, L., Lövdén, M., 2012. Growth of language-related brain areas after foreign language learning. NeuroImage 63, 240–244.
May, A., Gaser, C., 2012. Response to Thomas and Baker: the structural adaptation of the brain to training. Trends Cogn. Sci. 20, 1–2.
Pereira, A.C., Huddleston, D.E., Brickman, A.M., Sosunov, A.A., Hen, R., McKhann, G.M., Sloan, R., Gage, F.H., Brown, T.R., Small, S.A., 2007. An in vivo correlate of exercise-induced neurogenesis in the adult dentate gyrus. Proc. Natl. Acad. Sci. U. S. A. 104, 5638–5643.
Pierpaoli, C., 2011. Artifacts in diffusion MRI. Diffusion MRI. Oxford Publications, pp. 303–318.
Sagi, Y., Tavor, I., Hofstetter, S., Tzur-Moryosef, S., Blumenfeld-Katzir, T., Assaf, Y., 2012. Learning in the fast lane: new insights into neuroplasticity. Neuron 73, 1195–1203.
Schlegel, A.A., Rudelson, J.J., Tse, P.U., 2012. White matter structure changes as adults learn a second language. J. Cogn. Neurosci. 24, 1664–1670.
Scholz, J., Klein, M., Behrens, T., Johansen-Berg, H., 2009. Training induces changes in white-matter architecture. Nat. Neurosci. 12, 1370–1371.
Streitbürger, D.P., Möller, H.E., Tittgemeyer, M., Hund-Georgiadis, M., Schroeter, M.L., Mueller, K., 2012. Investigating structural brain changes of dehydration using voxel-based morphometry. PLoS One 7, e44195.
Tang, Y., Lu, Q., Geng, X., Stein, E., Yang, Y., Posner, M., 2010. Short-term meditation induces white matter changes in the anterior cingulate. Proc. Natl. Acad. Sci. U. S. A. 107, 15649–15652.
Tang, Y.Y., Lu, Q., Fan, M., Yang, Y., Posner, M.I., 2012. Mechanisms of white matter changes induced by meditation. Proc. Natl. Acad. Sci. 109, 10570–10574.
Thomas, C., Baker, C.I., 2012. Remodeling human cortex through training: comment on May. Trends Cogn. Sci. 20.
Thomas, C., Baker, C.I., 2012. Teaching an adult brain new tricks: A critical review of evidence for training-dependent structural plasticity in humans. NeuroImage.
Thomas, A., Marrett, S., Saad, Z., Ruff, D., Martin, A., Bandettini, P., 2009. Functional but not structural changes associated with learning: an exploration of longitudinal voxel-based morphometry (VBM). NeuroImage 48, 117–125.
Thomas, A.G., Dennis, A., Bandettini, P.A., Johansen-Berg, H., 2012. The effects of aerobic activity on brain structure. Front. Psychol. 3. http://dx.doi.org/10.3389/fpsyg.2012.00086.
Voss, M.W., Heo, S., Prakash, R.S., Erickson, K.I., Alves, H., Chaddock, L., Szabo, A.N., Mailey, E.L., Wójcicki, T.R., White, S.M., Gothe, N., McAuley, E., Sutton, B.P., Kramer, A.F., 2012. The influence of aerobic fitness on cerebral white matter integrity and cognitive function in older adults: results of a one-year exercise intervention. Hum. Brain Mapp. http://dx.doi.org/10.1002/hbm.22119.
Walker, L., Chang, L.C., Koay, C.G., Sharma, N., Cohen, L., Verma, R., Pierpaoli, C., 2011. Effects of physiological noise in population analysis of diffusion tensor MRI data. NeuroImage 54, 1168–1177.
Walker, L., Gozzi, M., Lenroot, R., Thurm, A., Behseta, B., Swedo, S., Pierpaoli, C., 2012. Diffusion tensor imaging in young children with autism: biological effects and potential confounds. Biol. Psychiatry 72 (12), 1043–1051.
Woollett, K., Maguire, E.A., 2011. Acquiring "the Knowledge" of London's layout drives structural brain changes. Curr. Biol. 21, 2109–2114.