Neuroinformatics

ISSN 1539-2791/04/163–168/$25.00

Original Article

Fair Localization of Function Via Multi-Lesion Analysis

Alon Keinan,[1] Alon Kaufman,[2] Nadia Sachs,[3] Claus C. Hilgetag,[3] and Eytan Ruppin[*,1,4]*

[1]School of Computer Science, Tel Aviv University, Tel Aviv, Israel; [2]Center of Neural Computation, Hebrew University, Jerusalem, Israel; [3]School of Engineering and Science, International University Bremen, Bremen, Germany; [4]School of Medicine, Tel Aviv University, Tel Aviv, Israel. (The first two authors contributed equally to the paper.)

Abstract

Acknowledging that causal localization of function in a processing network requires a multi-lesion analysis, this paper presents a rigorous and efficient method for defining and calculating the functional contributions of network elements as well as their interactions. The method's applicability to biological networks is demonstrated in the investigation of spatial attention in cats via lesion and reversible deactivation experiments.

Index Entries: Localization of function; multi-lesions; Shapley value; contributions analysis; interactions; multi-perturbations.

Introduction

One of the fundamental challenges in understanding neural information processing is identifying the individual roles of a neural network's elements. Localization of specific tasks in the nervous system is typically done by recording neural activity during behavior and correlating the elements' activation with behavioral and functional observables.

However, this correlation does not necessarily identify causality. To enable a causal localization of function, lesion studies have been employed in which functional performance is measured after disabling different elements of the system. While lesion studies have classically played an important role in function localization, most of them have been single lesion studies, in which only one element is disabled at a time. Such approaches are limited in their ability to reveal the significance of interacting elements (Aharonov et al., 2003). One obvious example is provided by two elements that exhibit a high degree of redundancy in their function: Lesioning either element alone will not reveal its significance.

*Address to which all correspondence and reprint requests should be sent.
E-mail: ruppin@post.tau.ac.il

Acknowledging that single lesions are insufficient for localization of function in neural systems, a multi-lesion analysis approach is needed. This article presents a new method, the Multi-perturbation Shapley value Analysis (MSA), which addresses the challenge of defining and calculating the contributions of network elements from a data set of multiple lesions (or, more generally, multi-perturbations) and their corresponding performance scores. In this framework, we view a set of multiple lesion experiments as a coalitional game, borrowing concepts and analytical approaches from the field of game theory. Specifically, we define the set of contributions to be the Shapley value (Shapley, 1953), which stands for the unique fair division of the game's worth (the network's performance score when all elements are intact) among the different players (the network elements). The contribution of an element to a function measures its importance, that is, the part it plays in the successful performance of that function. While in traditional game theory the Shapley value is more a theoretical tool that assumes full knowledge of the behavior of the game at all possible coalitions, we have developed methods to compute it approximately with high accuracy and efficiency from a relatively small set of multiple lesion experiments (*see* Keinan et al., 2004 for a more detailed account of the MSA). Specifically, in the results to follow, a predictor is trained using a given subset of multi-lesion experiments to predict the performance levels of all possible multi-lesion experiments (various predictors were used, including projection pursuit regression and multiple expert neural networks).

The Shapley value stands for the average marginal importance of an element in the performance of an investigated function. For complex networks where the importance may depend on the state (lesioned or intact) of other elements, a higher order description is required to capture the characteristics of the function's performance. Focusing here on a two-dimensional analysis, we define the interaction between a pair of elements as how much larger (or smaller) the average marginal importance of the two combined elements is, compared with the sum of the average marginal importance of each of them separately when the other one is lesioned. Intuitively, the interaction states how much "the whole is greater than the sum of its parts" (synergism) or vice versa (antagonism), where the whole is the pair of elements. Furthermore, the MSA classifies the type of interaction between each pair based on comparison of the average marginal importance of an element when its pair is always intact ($\gamma_{i,j}$) and when its pair is always lesioned ($\gamma_{i,j}^{-}$): A positive value of both measures denotes that element *i*'s contribution is always positive, irrespective of whether element *j* is lesioned or intact. When both are negative, element *i* always hinders the performance, irrespective of the state of element *j*. In cases where the two measures have inverted signs, we define the contribution of element *i* as *j*-modulated. The interaction is defined as positive modulated when $\gamma_{i,j}$ is positive while $\gamma_{i,j}^{-}$ is negative, causing a "paradoxical" effect. We define the interaction as negative modulated when the former measure is negative and the latter is positive. The interaction of *j* with respect to *i* may be categorized in a similar way, yielding a full description of the type of interaction between the pair.

To demonstrate the applicability of our approach to the analysis of biological data, we investigated the localization of spatial attention to auditory and visual stimuli, using behavioral data from deactivation and lesion experiments of different brain regions in cats. The experiments tested auditory and visual stimuli detection and orientation responses, as a measure of spatial attentional behavior, in the left hemifield of the cat. While attentional mechanisms proceed efficiently and inconspicuously in the intact brain, perturbation of

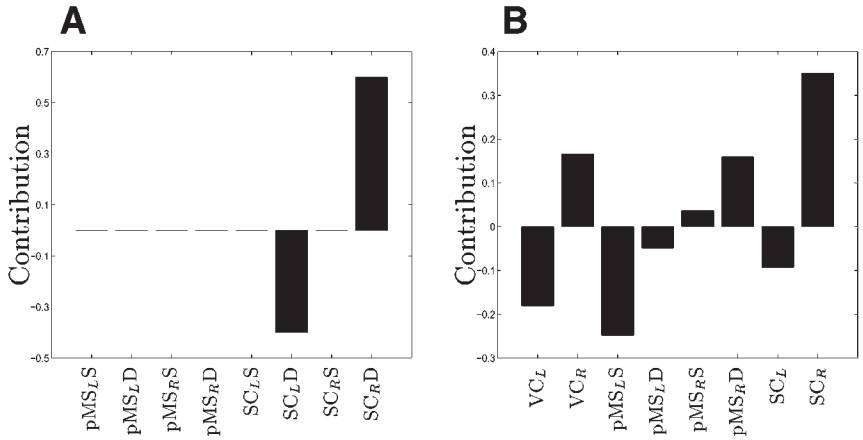

Fig. 1. MSA contributions of the different regions to auditory **(A)** and visual **(B)** spatial attention, in the left hemifield of the cat.

these mechanisms can lead to severe behavioral impairments. From the perspective of systems neuroscience, attentional mechanisms are particularly interesting because this function is known to be widely distributed in the brain. Moreover, lesions in the attentional network result in "paradoxical" effects, in which the deactivation of some regions results in a better-than-normal performance (Hilgetag et al., 2001), or reverses behavioral deficits resulting from earlier lesions (Sprague, 1966). Such effects challenge traditional approaches for lesion analysis and provide a testbed for novel formal analysis approaches such as the MSA.

Auditory Spatial Attention

The elements studied in the auditory experiments were left and right Superior Colliculus (SC) and left and right posterior Middle Suprasylvian (pMS) cortex. These structures were further subdivided into superficial and deep laminar compartments, so that there were altogether eight regions in the system's description. Figure 1A presents the MSA contributions of the different regions, calculated using prediction based on 33 available lesion experiments. The analysis revealed that only regions SC_L-deep and SC_R-deep play a role in determining auditory attentional performance in the left hemifield, which is in line with previous functional characterizations of the collicular system (Lomber et al., 2001). The contralateral deep layer of the SC has a positive contribution, suggesting that lesioning this region hinders performance. In contrast, the negative contribution of the ipsilateral SC indicates that lesioning of this region tends to improve the performance. Because of the restricted experimental access to deeper brain structures, when a deep layer of the SC was lesioned, the superficial one was lesioned as well. Nevertheless, the MSA framework successfully revealed that only the deep SC regions are the ones of significance.

We further performed a two-dimensional MSA to quantify the interactions between each pair of regions, finding only one significant interaction, between SC_L-deep and SC_R-deep. Observing that SC_L-deep has no contribution when SC_R-deep is intact, while it has an average negative contribution when SC_R-deep is lesioned, the MSA concluded that SC_L-deep exhibits a positive-modulated interaction with respect to SC_R-deep, uncovering the type of interaction assumed to take place in this function (Hilgetag et al., 2000; Lomber

et al., 2001). One should emphasize that a single lesioning analysis of the same data would not have revealed this "paradoxical" effect, and would have concluded that SC_R-deep is the only important region in this task.

Visual Spatial Attention

We turn to analyzing data from similar lesion experiments studying the localization of spatial attention to moving visual stimuli. The analyzed data set was collected from various lesion and reversible deactivation experiments of striate and parastriate visual cortex (VC), SC and deep and superficial layers of the pMS cortex in the cat (e.g., Sprague, [1966]; Lomber and Payne, [1996]; Lomber et al., [2001]; Lomber et al., [2002]). The MSA contributions were calculated using prediction based on 21 available lesion experiments out of the 2^8 multi-lesion space. This small number of lesion experiments is not sufficient to predict precisely the full multiple lesion set and hence, the results might be inaccurate. However, any additional experimental results, as they become available, can be added in a straightforward manner to refine the prediction. Figure 1B shows the contributions of the different regions involved in the experiments. The analysis yielded opposite contributions of left- and right-hemispheric structures, with the structures on the ipsilateral side to the tested (left) field having negative contributions and the structures on the contralateral side having positive ones. The largest positive contribution made by SC_R is in line with the presumed central role of the SC in visual spatial attention in cats. Conversely, the largest negative contribution is made by the superficial layers of the ipsilateral pMS owing to their powerful role in reversing the impact of contralateral lesions. For example, deactivation of the ipsilateral superficial pMS is sufficient to reverse the impact of a complete deactivation (superficial and deep) of contralateral pMS (Lomber and Payne, 1996), or even a lesion of the entire contralateral VC including the pMS (Lomber et al., 2002). The role of cortical regions in visual spatial attention in the cat has been previously investigated in great detail for pMS. Our analysis also suggests important contributions of other cortical regions (VC). It remains a challenge for future deactivation experiments to test this prediction and identify additional regions within the cat visual cortex that specifically contribute to attentional behavior.

Detailed two-dimensional MSA reveals the functional interactions between the regions. Figure 2A illustrates the type of the most significant interactions. Figure 2B quantifies the contributions of lesioning various regions, with respect to a given lesion in the network. The contribution of lesioning region i while region j is already lesioned ($-\gamma_{i,j}^{-}$) is based on the average marginal contribution of such lesioning, taking into consideration other possible unknown lesions to other regions in the network. The figure clearly shows the significant reverse impact of the ipsilateral superficial layers of pMS. This graph is just one example of displaying the causal effects that can be revealed by the MSA, testifying to its usefulness in deducing the functionally important regions, as well as their significant interactions.

MSA is a novel method for causal localization of function, with a wide range of potential applications, not only for the analysis of lesion and reversible deactivation experiments, but also for the analysis of laser ablation studies of neurons and TMS-induced "virtual lesions." It can also serve in the analysis of existing neural models of biological systems, ranging from detailed models of organisms to large-scale integrated models of imaging and activation dynamics in different tasks.

Acknowledgments

We acknowledge the valuable contributions and suggestions made by Ranit Aharonov,

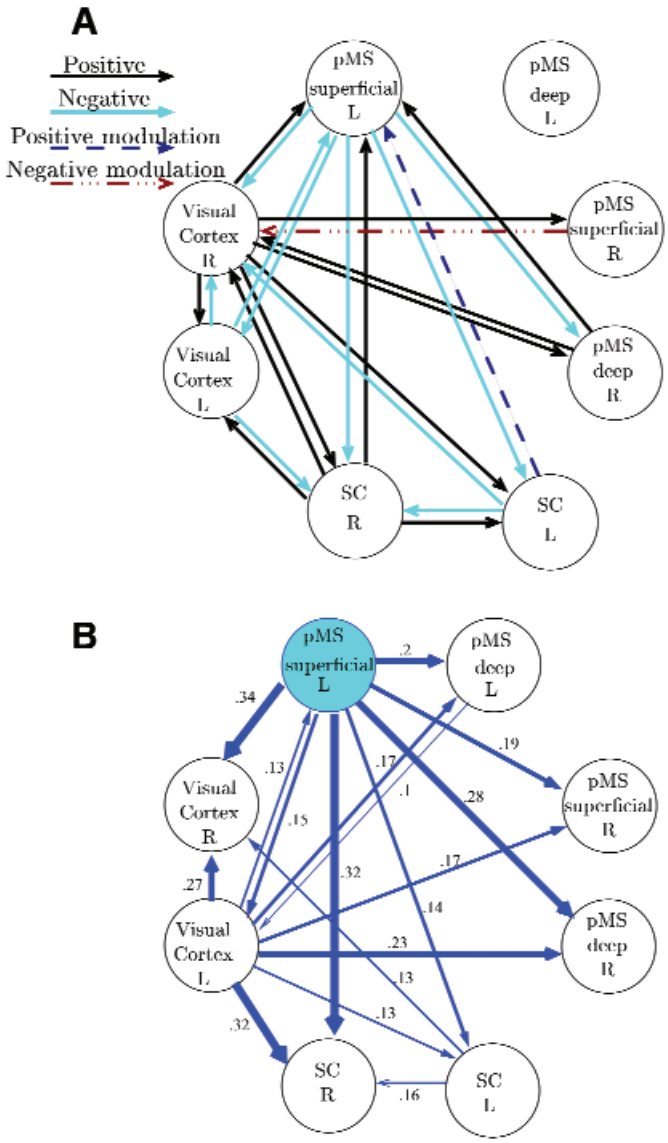

Fig. 2. Significant interactions in the visual experiments. **(A)** The type, in both directions, of the twelve most significant two-dimensional interactions. **(B)** The positive contributions of lesioning a region knowing a specific other region is lesioned ($-\gamma_{i,j}^{-}$). The figure shows the significant values, arrows representing the contribution of lesioning the arrow's base region (i), given that the arrow's end region (j) is already lesioned.

Dudi Deutscher, Ehud Lehrer, and Isaac Meilijson. This research has been supported by the Adams Super Center for Brain Studies in Tel Aviv University and by the Israel Science Foundation founded by the Israel Academy of Sciences and Humanities.

References

Aharonov, R., Segev, L., Meilijson, I., and Ruppin, E. (2003) Localization of function via lesion analysis. Neural Comput. 15 (4), 885–913.

Hilgetag, C., Theoret, H., and Pascual-Leone, A. (2001) Enhanced visual spatial attention ipsilateral to

rTMS-induced virtual lesions of human parietal cortex. Nat. Neurosci. 4(9), 953–957.

Hilgetag, C. C., Lomber, S. G., and Payne, B. R. (2000) Neural mechanisms of spatial attention in the cat. Neurocomputing 38, 1281–1287.

Keinan, A., Sandbank, B., Hilgetag, C. C., Meilijson, I., and Ruppin, E. (2004) Fair attribution of functional contribution in artificial and biological networks. Neural Computation 16(2).

Lomber, S. G. and Payne, B. R. (1996) Removal of two halves restores the whole: Reversal of visual hemineglect during bilateral cortical or collicular inactivation in the cat. Vis. Neurosci. 13 (6), 1143–1156.

Lomber, S. G., Payne, B. R., and Cornwell, P. (2001) Role of the superior colliculus in analyses of space: Superficial and intermediate layer contributions to visual orienting, auditory orienting, and visuospatial discriminations during unilateral and bilateral deactivations. J. Comp. Neurol. 441, 44–57.

Lomber, S. G., Payne, B. R., Hilgetag, C. C., and Rushmore, R. J. (2002) Restoration of visual orienting into a cortically blind hemifield by reversible deactivation of posterior parietal cortex or the superior colliculus. Exp. Brain. Res. 142, 463–474.

Shapley, L. S. (1953) A value for n-person games. In: Contributions to the theory of games, volume II. (Kuhn, H. W., Tucker, A. W., eds.) Princeton University Press, Princeton, NJ, pp. 307–317.

Sprague, J. M. (1966) Interaction of cortex and Superior Colliculus in mediation of visually guided behavior in the cat. Science 153, 1544–1547.