Modeling brain reorganization after hemispherectomy

Luís F Seoane*[1,2,3] and Ricard Solé[4,5,6]
[1]Departamento de Biología de Sistemas, Centro Nacional de Biotecnología (CSIC), C/ Darwin 3, 28049 Madrid, Spain.
[2]Grupo Interdisciplinar de Sistemas Complejos (GISC), Madrid, Spain.
[3]Instituto de Física Interdisciplinar y Sistemas Complejos, IFISC (CSIC-UIB), Campus UIB, 07122, Palma de Mallorca, Spain
[4]ICREA-Complex Systems Lab, Universitat Pompeu Fabra (GRIB), Dr Aiguader 80, 08003 Barcelona, Spain
[5]Institut de Biologia Evolutiva, CSIC-UPF, Pg Maritim de la Barceloneta 37, 08003 Barcelona, Spain
[6]Santa Fe Institute, 1399 Hyde Park Road, Santa Fe NM 87501, USA

Brain reorganization after hemispherectomy (i.e. the removal of a whole hemisphere) is perhaps the most remarkable example of large-scale brain plasticity. Most often patients survive and recover their skills. Functional traits located in the lost side (e.g. language areas) can sometimes be completely reassembled in the remaining hemisphere, which seamlessly take on the additional processing burden. This demands drastic rearrangements, perhaps involving the readaptation of functionally and structurally diverse neural structures. We lack mathematical models of how this happens. We introduce a very simple model, based on self-organized maps, that provides a rationale to the clinical aftermath of the intervention, putative windows for recovery, and the origins and nature of observed thresholds for irreversible function loss. The implications for brain symmetry and potential scenarios in simulated pathologies, including efficient suggested treatments, are outlined.

Keywords: hemispherectomy, self-organized maps, neural plasticity, somatosensory maps, language

I. INTRODUCTION

The well-known potential of brains to cope with damage and aging is one of their most remarkable attributes. Both degradation and regeneration happen on multiple scales, from neurons to brain tissues, and over time. A remarkable example of this regenerative potential is the capacity to withstand brain hemispherectomy (BHEM). This is an extreme (but largely successful) neurosurgical procedure used to treat epileptic disorders that involve severe, life-threatening episodes. BHEM entails removal or disconnection of a whole cerebral hemisphere. Surprisingly, in the ensuing "half a brain" scenario [1–3] patients can recover cognitive, motor, and somatosensory functions very close to normal [4–6].

The case of language illustrates some challenges after BHEM. In complete brains, language is spontaneously lateralized following the assembly and maturation of several complex, tightly interlinked cortical regions [7–13], most often in the brain side. Children below three who have their left hemisphere removed can still recover language by having these structures reorganize at their right side. Interventions at early age correlate with more successful recovery, while processes later in life most often result in complete language loss. This suggests the existence of *window periods* (around six years of age [14–17]) after which function cannot be regained. But evidence is inconclusive, with counterexamples of language recovery at older age showing that the limits are not rigid [17–20].

The successful clinical cases reveal that we can survive and thrive with only one hemisphere, but little is known about how the brain reorganization itself takes place. A recent fMRI study registered neural activity from BHEM patients [21]. It is found that brain parcellations derived from intact, healthy brains are consistent with activity in hemispherectomized brains. Overall, connections within task-specific (e.g. resting state, motor, attention) networks increase with respect to normal brains. Commonly, connections across networks remain as in complete brains. Incidental evidence suggests that more connections between networks (i.e. *misplaced* links) result in worst cognitive performance. Thus functional centers appear to take up preferably affine tasks.

Brain reorganization has been proved, among others, by experiments in macaques [22–27], clinical procedures [28], and recovery after stroke [29]. Simple mathematical models [30–34] illustrate smooth, local processes when topographic maps need to expand (invading adjacent areas) or change their nature slightly. Less consensus emerges at larger scales. Some experiments indicate a spatial limit to plasticity, suggesting that some representations cannot be smoothly transferred to distant parts of a same cortical area [23, 24, 26]. Others demonstrate deep rearrangements of topographic maps prompted by drastic transformations of the external anatomy [22, 25, 27].

How does this remapping take place? Does it proceed smoothly, or does it need a drastic rewiring of the intact cortical structures? If so, how drastic, and how does it happen? Are new cognitive centers developed, or do existing areas double up their efforts? Is the computational load distributed according to each neural circuit's specific capabilities, or are the existing parts reconfigured to take

*corresponding author

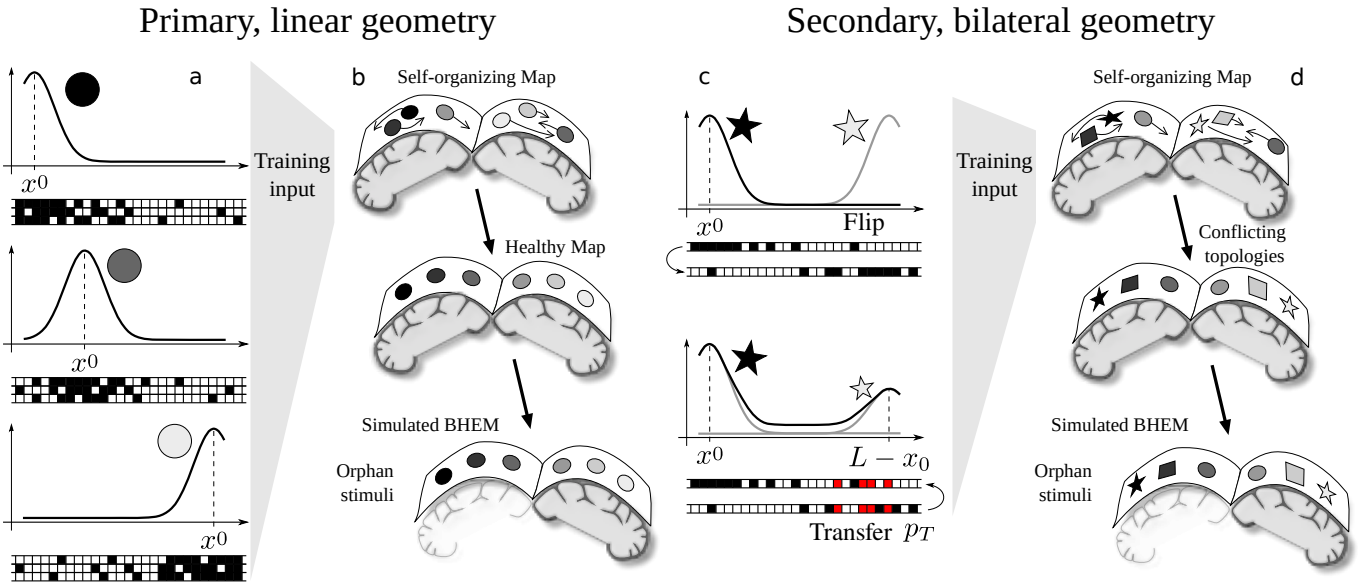


FIG. 1 **Modeling hemispherectomy with self-organized maps.** (a) Input signals consist of a string of bits. Probability of a bit taking value 1 (black boxes) peaks at a different position $x^0 \in [0, L]$ for each signal, which gives the input set an underlying one-dimensional geometry. A gray gradient (within circles) further illustrates this contiguity. Three possible signals are shown for each x^0. (b) Self-organized maps infer the underlaying geometry of the data and acquire an internal geometric structure of their own which might condition reorganization after simulated BHEM (c) We model more complex inputs with the primary linear geometry (gradient-coded) and an additional bilateral symmetry. Bilateral correspondence is illustrated by geometric shapes, which conflict with the simple contiguity of linear inputs. (d) Attempts to learn the conflicting geometries might result in SOMs shifting between alternative representations (Supporting Figure 3).

on alien functions? Can we capture the complex features and scales involved in BHEM in a simple model?

In this paper we seek to use the minimal key ingredients for brain reorganization to provide a rationale for the clinical and functional features observed after BHEM. We represent higher brain functions abstractly, as the ability of a Self-Organized Map (SOM) [35] to build globally-coherent and detailed representations out of local input cues. Below, we detail how we model these input stimuli, which might display two conflicting underlying geometries. We simulate BHEM and SOM reorganization under the simplest representations of lateralized stimuli and investigate how this changes as bilaterality becomes increasingly relevant. We further study a priming processes (of biological relevance) that enables our SOMs to achieve finer representations. We find that this priming conflicts with strongly bilateral stimuli, resulting in pathological representations. Reorganization of primed SOMs after simulated hemispherectomy strongly suggest that this intervention can be preferred for pathological SOMs.

These insights concern mostly somatosensory maps. Next we turn our attention to the language ability, which involves lateralized structures [7–13] that develop from bilaterally symmetric substrates in newborns [8, 13, 36]. Therefore we model redundant representations seated at opposing hemispheres, and how they fare before and after BHEM. Our model strongly suggests that lateralized stimuli can challenge the inborn bilateral symmetry, forcing the development of a dominant side. Keeping bilaterality risks a deterioration of both candidate representations. We observe how the dominated hemisphere degrades through the lateralization process, and how this degradation is key to explaining window periods within which language recovery is possible – and beyond which it becomes rare. Our results are in accordance with empirical observations and propose a mechanism that explains phenomenology around these window periods.

II. METHODS

Topographic maps in the somatosensory cortex represent body parts, which appear sorted (from the mouth to the toes) along the distal-medial axis of each hemisphere. This parsimonious representation along a line (which we refer to as a *linear* geometry) mimics the simplest, idealized arrangement of body parts along the body itself. A second relevant geometric component is the bilateral symmetry between parts at each side. Our somatosensory cortices have captured these and other geometrical details of our body through our evolution, through the development of each individual, and by means of plastic responses to distinct stimuli within a lifetime. Figure 1 summarizes how we model input stimuli with one or two conflicting geometries, and how SOMs attempt to infer

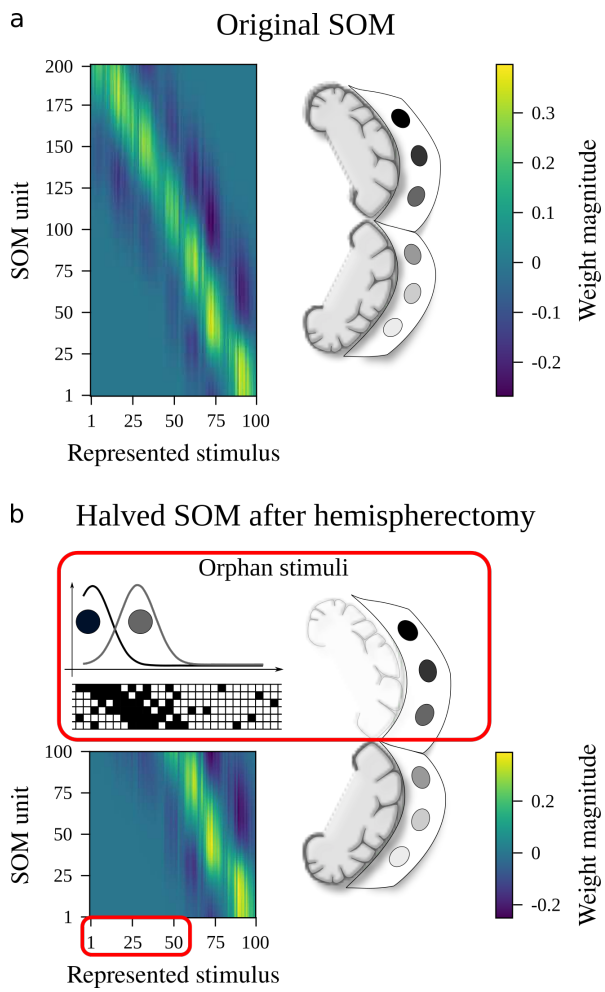


FIG. 2 **In silico hemispherectomy.** (a) A self-organized map infers the geometry of its input signals. An optimal representation of linear inputs results in a roughly linear correspondence between x^0 and the active SOM units, which is also reflected in the SOM's weights. (b) We simulate hemispherectomy by removing half a SOM. This results in a set of "orphan inputs" whose natural representation is missing.

those geometries through unsupervised learning. In the following we lay out the mathematical details.

A. Modeling stimuli with linear geometry and bilateral symmetry

Consider a one-dimensional, discretized space of length L with boxes centered on $x_i \in \{1, \ldots, L\}$. These boxes can be empty or filled (Figure 1(a)), thus each contains one bit. The resulting bit string is an input signal, I, for our SOM. The value of the box at position x_i is labeled I_i and takes value 1 with probability:

$$p(I_i = 1) = P^0 \exp\left(-\frac{\left(x_i - x^0\right)^2}{2\sigma_s^2}\right). \tag{1}$$

This is a Gaussian normalized by $P^0 = 1$ and peaked around $x^0 \in [1, L]$ with standard deviation σ_s. Equation 1 represents an input elicited at x^0 (think, a touch at a given length along the arm) that has a chance of stimulating areas nearby. The parameter σ_s controls the extent of this spatial correlation. Some stochasticity is present, hence different stimuli centered at x^0 result in distinct binary patterns (each x^0 in Figure 1(a) is accompanied by three examples). Somatosensory inputs vary over time and originate at changing locations. In our experiments we generate T input signals ($I(t)$, $t = 1, \ldots, T$), each peaked at a different (random, uniformly drawn) location $x^0(t) \in [1, L]$.

Real somatosensory inputs are much more complex. Of all the ingredients that enrich our sensory experience, bilateral symmetry plays a key role. Bilaterality is present in the overall structure of the brain [37–39] and in somatosensory stimuli – e.g. as both hands engage a same activity simultaneously. To model a stimulus with bilateral symmetry (Figure 1(c)), given x^0 we generate two bit strings as above. We flip one of them, then each of its bits equal to 1 is transferred to the unflipped string with probability p_T. This results in a double-peaked probability for $p(I_i = 1)$ with the parameter $p_T \in [0, 1]$ controlling the level of bilaterality.

We refer to the signals without bilateral symmetry as having a *linear* geometry. Signals with some $p_T > 0$ are referred to as *bilateral*. Two relevant model parameters have been introduced so far: σ_s, controlling correlations between nearby input positions; and p_T, controlling correlations between mirror-symmetric positions.

B. Self-Organized Maps models

Self-Organized Maps [35] are a class of neural networks able to infer, in an unsupervised fashion, the geometrical structure underlaying a given data set. Different SOM versions have been used to model local plasticity in topographic maps after an injury [30–34]. Our implementation is the simplest possible, faithful to the original scheme.

Our self-organized network (Supporting Figure 1(a)) consists of a set of N neurons (labeled $j = 1, \ldots, N$) that gather external stimuli. Each unit has L weights $\{\omega_i^j, i = 1, \ldots, L\}$, labeled as the positions of bit boxes in our input signals. In general, $L \neq N$. At time t, the input $I(t) \equiv \{I_i(t), i = 1, \ldots, L\}$ reaches all units. This alters their activity, which becomes:

$$s^j(I(t)) \;=\; \sum_{i=1}^{L} \omega_i^j I_i(t). \tag{2}$$

One unit (labeled k in Supporting Figure 1(a)) has a greater activity than any other. We say that it becomes "activated", while all others are silenced. We say that the active unit *represents* the given stimulus $I(t)$, and

note it:

$$u\left(I(t)\right) \;\equiv\; \arg\max_j \left(\{s^j(I(t))\}_{j=1,\ldots,N}\right) = k. \quad (3)$$

If the network is in training mode, weights are updated according to:

$$\omega_i^j \rightarrow \frac{\omega_i^j + \eta\Psi(j,k)I_i(t)}{||\omega_i^j + \eta\Psi(j,k)I_i(t)||}. \quad (4)$$

This is, each unit's weights take a step in the direction of $I(t)$, thus becoming better at responding to this signal in the future. This step is bounded by a learning rate (here, $\eta = 0.1$). A further, local modulation is introduced by:

$$\Psi(j,k) \;=\; \left[1 - \frac{(j-k)^2}{\sigma_t^2}\right] e^{-\frac{(j-k)^2}{2\sigma_t^2}}. \quad (5)$$

This is the second derivative of a Gaussian centered in the active unit, k, and with standard deviation σ_t (Supporting Figure 1(b)). $\Psi(j,k)$ is positive in a neighborhood of k. Weights of units in this neighborhood move towards $I(t)$ – thus also becoming better at responding to this signal. This implements a Hebbian learning step. Outside this neighborhood, $\Psi(j,k)$ is negative – thus the corresponding units move away from $I(t)$ and become worst at responding to it. This implements an anti-Hebbian learning step. Updated weights are normalized to prevent them from blowing up.

These training dynamics introduce a new model parameter, σ_t, that encodes spatial correlations between SOM units. These correlations are different from the ones within input signals. In general $\sigma_t \neq \sigma_s$.

C. Optimal representations of linear and bilateral stimuli

Equation 3 defines a SOM's representation of a specific signal. We are interested in SOM's responses to classes of signals – specifically, to linear or bilateral inputs as a function of x^0. We denote these representations $u(x^0) \in \{1,\ldots,N\}$ (with $x^0 \in [1,\ldots,L]$), and evaluate them empirically by generating inputs peaked at x^0 and finding the active units as this parameter changes.

A good map will capture the signal's intrinsic geometry (whether linear or with bilateral symmetry). Hence, the best representation of fully linear stimuli traces a straight line over the space of active SOM units as a function of x^0 (Figure 2(a)). Two equally good representations exist:

$$u_L^+(x^0) = \frac{N}{L}x^0, \qquad u_L^-(x^0) = N - \frac{N}{L}x^0. \quad (6)$$

The active SOM unit is a discrete variable, so we can at best approximate $u_L^\pm(x^0)$. We compute how much does a representation deviate from the closest optimum:

$$d_L \;=\; \min_\pm \left\{\frac{1}{L}\sum_{i=1}^{L}\left[\frac{u_L^\pm(i) - u(i)}{N}\right]^2\right\}. \quad (7)$$

Good representations are i) exhaustive and ii) globally coherent. The first property means that as much space over $x^0 \in [1,L]$ is covered using as many SOM units as possible. This is achieved if each unit represents a stretch of length L/N. But we might arrive to exhaustive representations that are fragmented (Supporting Figure 2(a-c)), meaning that two nearby positions along $x^0 \in [1,L]$ are mapped to units which are far appart in the SOM (i.e. there is some x^0 such that $||u(x^0) - u(x^0 + \varepsilon)|| \gg 1$ for $\varepsilon \rightarrow 0$). In globally-coherent maps, instead, $u(x^0)$ changes monotonously as x^0 varies (Figure 2(a) and Supporting Figure 2(d-f)).

Optimal representations of stimuli with full bilateral symmetry ($p_T = 1$) respond equally to inputs peaked at x^0 and $L - x^0$ (i.e., to stimuli peaked at a certain location and to their mirror image). Again two optimal functions solve this problem:

$$\begin{aligned} u_{BL}^+(x) \;&=\; \begin{cases} \frac{2N}{L}x, & \text{if } x \leq L/2, \\ 2N - \frac{2N}{L}x, & \text{if } x > L/2; \end{cases} \\ u_{BL}^-(x) \;&=\; \begin{cases} N - \frac{2N}{L}x, & \text{if } x \leq L/2, \\ -N + \frac{2N}{L}x, & \text{if } x > L/2. \end{cases} \end{aligned} \quad (8)$$

These are piece-wise straight lines shaped into a 'V' (Supporting Figures 3 and 4(d-f)). A SOM's distance to its closest optimal bilateral representation reads:

$$d_{BL} \;=\; \min_\pm \left\{\frac{1}{L}\sum_{i=1}^{L}\left[\frac{u_{BL}^\pm(i) - u(i)}{N}\right]^2\right\}. \quad (9)$$

To evaluate a SOM's distance to the bilateral representation, we activate the network with stimuli *without* bilateral symmetry to ascertain that it is responding to each component separately.

D. Priming globally consistent, yet detailed representations

The representations seen so far limit the input and SOM sizes that we can explore. If L, N become too large, fixed σ_s, σ_t eventually look small. Then, correlations within inputs or between units are not enough to establish a global coherence in the emerging representations, resulting in fragmented maps. If σ_s, σ_t are kept large, each unit's response to input stimuli becomes very broad and ambiguous, overlaps with the representations of other units, and the emerging maps do not exploit all resources optimally (Supporting Figure 5). Ideally, we would like to work with small σ_s, σ_t that would show fine-grained representations, yet still produce globally coherent maps.

To solve this we *prime* our networks with a hint of the correct representation for linear stimuli. So far, SOMs are initialized with random weights $\omega_{ij} \in \omega^R$, where $\omega^R \in [-1,1]$ is a stochastic, uniformly distributed variable. Now consider a set of target weights ω^L such that

$\omega^L_{ij} = 1$ if $u^+_L(i) = j$ and $\omega^L_{ij} = 0$ otherwise. This embodies one of the optimal representations in Equation 7. By initializing our SOMs with:

$$\omega_0 \;=\; (1 - p_P) \cdot \omega_R + p_P \cdot \omega_L, \tag{10}$$

the parameter p_P trades between unprimed and fully primed SOMs. As p_P increases, a situation with fragmented representations (due to small σ_s, σ_t) transits into a globally coherent map (Supporting Figure 5). Since we keep $\sigma_s \ll L$ and $\sigma_t \ll N$, a unit's response does not become coarse.

III. RESULTS

We run a series of numerical experiments of simulated hemispherectomy. In each of them, we first train a SOM on either linear or bilateral stimuli (Figure 2(a)), then remove half of its units (Figure 2(b)), then retrain the remaining network on stimuli of the same kind. For training, we fix model parameters for the sensory input (L, σ_s, and p_T) and the SOM (N, σ_t, and p_P). Then we produce T_t stimuli examples to be presented to the SOM, each peaked at a random, uniformly distributed $x^0 \in [0, L]$. We measure the distance of the resulting representation, $u^0(x^0)$, to either the optimal linear, d^0_L, or bilateral, d^0_{BL}, representations. The superindex 0 indicates that this evaluation happens before the SOM is halved.

To simulate a BHEM, we copy half (either the first or last) contiguous units into a smaller SOM (Figure 2(b)). We evaluate this halved SOM right after hemispherectomy, $u^h(x^0)$, by measuring the distances d^h_L or d^h_{BL}. Note that these networks would have missing proper representations for some of the input stimuli. We term these stimuli 'orphan'. Finally, we retrain the halved SOM for T_{rt} time steps with new input strings. After retraining, we evaluate the SOM's representation, $u^{rt}(x^0)$, again with respect to linear, d^{rt}_L, or bilateral, d^{rt}_{BL}, optimal representations.

The ability to establish finely-tuned and globally-coherent representations depends on σ_s and σ_t (Supporting Figures 2 and 5). If $\sigma_s \ll L$ and $\sigma_t \ll N$ (Supporting Figure 2(a-c)), linear representations form locally, but not globally. As the training processes begins, locally coherent representations start forming, then grow like a crystal as nearby-peaked stimuli are shown. This happens at several locations over the range of SOM units. At each seed there is a choice between the two optimal representations, $u^{\pm}_L(x)$. Local choices are random and might collide with others as they expand, resulting in a globally fragmented $u(x^0)$. From a somatosensory perspective, this would be equivalent to representing two adjacent patches of skin in distant cortical regions.

Larger σ_s and σ_t result in more global coherence (Supporting Figure 2(d-f)). But the emerging representations are fairly coarse-grained (compare the distributions of weights in Supporting Figure 2(b) vs. 2(e)). This means that individual units of these globally-coherent maps are activated by a broad range of stimuli. This corresponds to a somatosensory representation that responds to stimulation of distant, unrelated patches of skin. Study of *primed* SOMs circumvents this problem.

A. Simulated hemispherectomy of unprimed SOMs

We want to study BHEM in relatively good maps, which model somatosensory representations in intact brains. We do not care about fragmented or otherwise pathological maps in his work. It is challenging to obtain optimal linear representations with unprimed SOMs. We explored the model visually to select fair unprimed SOMs (Supporting Figure 5).

Supporting Figure 7 illustrates BHEM in an unprimed ($p_P = 0$) SOM trained with fully linear stimuli ($p_T = 0$). Immediately after hemispherectomy, the SOM representation ($u^h(x^0)$, light gray in Supporting Figure 7(a); weights in Supporting Figure 7(c)) is missing its natural response to half of the stimuli. This "orphan domain" extends from $x^0 = 0$ to $x^0 = L/2$. SOMs always categorize any input, even if it is new or unexpected. Here, orphan inputs close to the center ($x^0 \lesssim L/2$) are parsimoniously mapped to the representation of the first non-orphan stimulus – because Hebbian learning guarantees that a neighborhood of units responds to a neighborhood of inputs. As we move towards the extreme of the orphan domain ($x^0 \to 0$), stimuli become represented at the opposite end of the halved SOM. This is due to the anti-Hebbian mechanism, which represses the response to moderately distant signals more strongly than to very distant ones. Orphan stimuli at the extreme ($x^0 \to 0$) are actually being assigned to *least indifferent units*.

New representations emerge after $T_{rt} = 10^3$ ($u^{rt}_3(x^0)$, medium gray in Supporting Figure 7(a)) and $T_{rt} = 10^4$ ($u^{rt}_4(x^0)$, dark gray in Supporting Figure 7(a)) retraining steps (weights in Supporting Figure 7(d) and (e) respectively). They grow from the original map, $u^h(x^0)$. Intact representations are gradually squeezed into smaller SOM regions to accommodate orphan stimuli (Figure 4(a)). The process seems smooth, as units at the edge of the (shrinking) orphaned domain are parsimoniously incorporated. This is consistent with earlier studies of lesions in topographic maps based on SOM [30–32, 34].

This parsimonious rearrangement might seem convenient, but it requires the relocation of all intact representations. In neuro-anatomical terms this seems possible to correct small injuries, for which orphaned function can be *stuffed* into relatively similar, adjacent tissue. But after a hemispherectomy this would imply that well-established representations are displaced way out of their place. Motor control regions must keep track of these changes too. Thus, over large scales, the parsimonious rearrangement appears costly and unlikely.

Things get more interesting with bilateral stimuli. For a fixed set of parameters (see Figure 3(a-b)) we explored

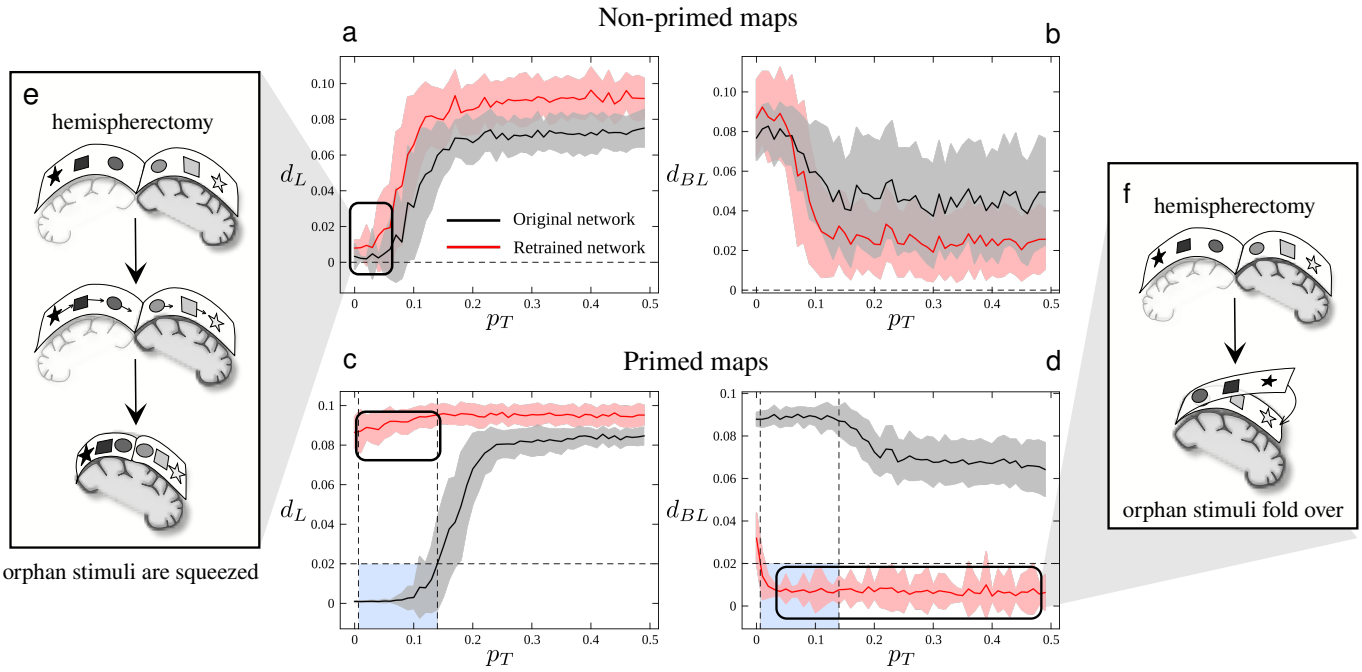


FIG. 3 **Optimality of emerging maps before and after hemispherectomy.** (a-d) SOMs were trained for $T_t = 10^4$ steps with signals of fixed bilaterality ($p_T \in [0, 0.5]$, horizontal axes). After BHEM, retraining proceeded for $T^{rt} = 10^3$. Plots display distances to the optimal linear (a, c) and bilateral (b, d) representations right before (black) and after hemispherectomy and retraining (red). Shading indicates standard deviation over 50 repeats. (a-b) Unprimed SOMs with $L = 100$, $\sigma_s = 3$, $N = 200$, and $\sigma_t = 30$. For low bilaterality, both complete and halved SOMs approach optimal linear representations. Hence, after BHEM, orphan input signals are parsimoniously squeezed into the intact half (e). For higher bilaterality, linear maps fail to form in complete SOMs. Emerging representations approach the optimal bilateral map – even more so in halved SOMs. (c-d) Primed SOMs ($p_P = 0.5$) with $L = 100$, $\sigma_s = 2$, $N = 100$, and $\sigma_t = 10$. Optimal linear representations form in complete SOMs and low p_T, but not on the halved SOM (black square, very large d_L^{rt}). The *squeezing* plasticity pathway is blocked (e). Above a bilaterality threshold, both linear and bilateral representations fail to form in complete SOMs ($d_{BL}^0 \gg 0$ for all p_T). But halved, primed SOMs fold very robustly into optimal bilateral representations (f).

a range of input bilaterality, $p_T \in [0, 0.5]$. Figure 3(a) shows d_L^0 (black) and d_L^{rt} (red). For small p_T ($\lesssim 0.1$), the emergent maps ignore the bilateral component and result in mostly linear representations ($d_L^{0/rt} \sim 0$). Most halved SOMs, again, reorganize themselves by squeezing orphan stimuli as before (Figure 4(b)). In average, complete SOMs achieve better maps than those after BHEM ($d_L^0 < d_L^{rt}$). For $p_T > 0.1$, these linear representations degrade quickly. In turn, maps closer to the optimal bilateral emerge (Figure 3(b)). Outstandingly, the halved SOM after retraining shows better bilateral maps than the full original network (Figure 4(d)). This suggests that BHEM might amend problems emerging on complete SOMs when neither optimal representation can emerge.

B. Simulated hemispherectomy of primed SOMs

Unprimed representations are coarse, which results in relatively large $d_L^{0/rt}$ and $d_{BL}^{0/rt}$. Primed SOMs render globally coherent yet fine-grained maps closer to a lateralized optimum. This mechanism is developmentally justified: phantom limbs in newborns with missing extremities [33] prove that somatosensory maps grow from innate blueprints acquired over evolutionary time.

We explore BHEM in primed SOMs ($p_P = 0.5$) with fixed parameters (see Figure 3(c-d)) for the same varying levels of input bilaterality ($p_T \in [0, 0.5]$). Representations in complete SOMs for low p_T ($\lesssim 0.1$) turn out much closer to the optimal linear map ($d_L^0 \simeq 0$, black plot in Figure 3(c)). But, crucially, priming precludes a reorganization into a similarly linear representation after hemispherectomy ($d_L^{rt} \gg 0$ for any p_T, red plot in Figure 3(c); Figure 4(e)) – even for stimuli with negligible bilaterality ($p_T \to 0$). Priming has induced a different, sturdy topological attractor in the halved SOM which, now, cannot be reorganized by parsimoniously squeezing orphaned stimuli into the intact substrate. Instead, orphan stimuli are readily mapped into their bilateral counterparts (Figures 3(f) and 4(f)).

This primed attractor also prevents the complete SOM from ever developing a bilateral representation ($d_{BL}^0 \gg 0$ for all p_T, black plot in Figure 3(d)). The interference with the primed map becomes hazardous as the bilateral component increases – for $p_T \gtrsim 0.1$ the linear representa-

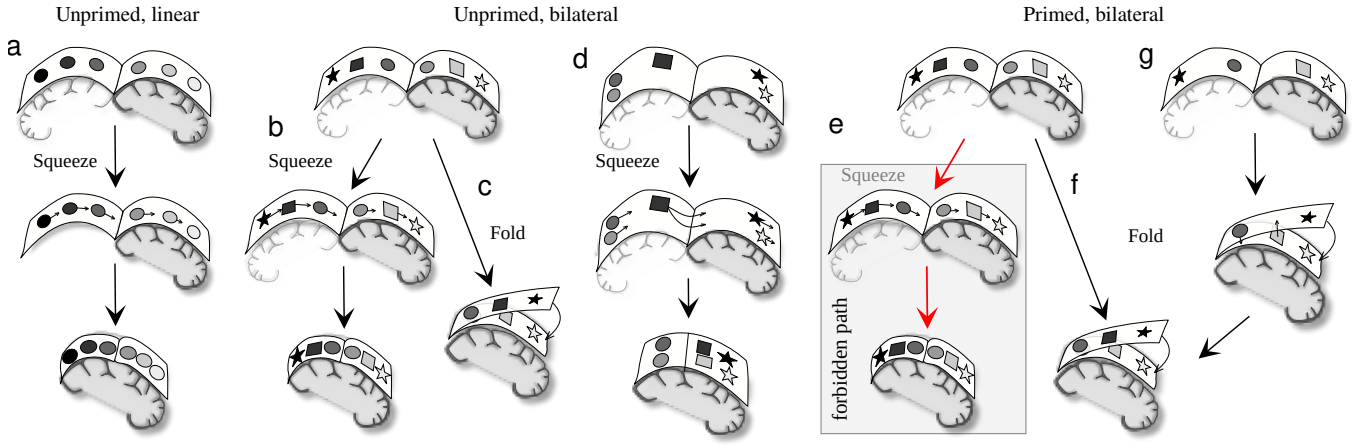


FIG. 4 **Enabled and forbidden pathways for brain reorganization.** (a-b) Reorganization by squeezing orphan stimuli into the intact neural substrate is the preferred form of rerganization for unprimed SOMs and signals with little or no bilateral component. (c) A small fraction (Supporting Figure 8(a-b)) of primed SOMs trained on moderately bilateral stimuli achieves a linear representation, but after BHEM it is reorganized by folding each orphan stimuli over to its bilaterally symmetric counterpart. (d) Most unprimed, complete SOMs trained on bilateral signals ($p_T \gtrsim 0.1$) capture the bilateral (rather than linear) geometry. After BHEM, halved SOMs reorganize into even better bilateral representations. We conjecture that squeezing is also the natural route for this kind of reorganization. (e) The *squeezing* route to reorganization is strictly prohibited for all primed SOMs. (f) In turn, a majority (Supporting Figure 8(c)) of healthy linear maps fold robustly and swiftly into bilateral representations. (g) Large input bilaterality in primed SOMs results in pathological maps without clear linear or bilateral geometry. This is readily fixed by BHEM, which folds the map into a sturdy bilateral representation.

tion is destabilized as well. Thus, an imbalance between the rigidity of inborn representations and excess input bilaterality results in pathological maps in complete SOMs. In turn, again, a striking side effect is that halved SOMs swiftly fold into robust bilateral maps ($d_{BL}^{rt} \simeq 0$ for all p_T except $p_T \to 0$, red plot in Figure 3(d); Figure 4(g)).

What does this mean in neuro-anatomical terms? Regarding the impossibility of reorganization into a linear map (Figure 4(e)): The priming process enables more refined representations (with smaller σ_s, σ_t), hence signals with close values of x^0 are more dissimilar than before. This hampers the gradual incorporation of orphan stimuli through the map's edge (opposed to what happened in Supporting Figure 7). This mechanism works locally at most, in agreement with earlier simulations of local injuries [30–32, 34].

On the upper side, stimuli are similar enough to their mirror-symmetric counterparts. This enables the robust folding of halved SOMs into bilateral maps (Figure 3(f) and 4(f)), which demands less overall reorganization since the intact representations are not dragged across the brain. This offers an insight on why BHEM works with relatively small cognitive impact. Furthermore, the bilateral maps of halved SOMs appear sturdy even when the linear representations of complete networks break apart (Figure 4(g)). This puts BHEM forward as a treatment for pathologies in which full, lateralized maps fail to develop. In our model this happens due to an interference between the two optimal topologies: a primed linear geometry versus a strong bilateral component.

C. Symmetry breaking of redundant structures – insights for language

The language capacity is housed in several cortical regions around the sylvan fissure of the dominant (usually left) hemisphere [7, 9–13]. The same areas in both brain halves start as symmetric at birth [8, 13, 36], but only the dominant side matures into a complete language organ. We model two roughly equivalent *left* and *right* SOMs that compete to elaborate a globally coherent representation of fully linear input signals (model parameters in Figure 5). Each SOM is independently primed with p_P^L and p_P^R respectively. This ensures that both sides are hinted towards the target representation (i.e. both contain similar precursory circuits), but a slight interhemispheric asymmetry is explored by keeping $p_P^R = 0.5$ and varying $p_P^L \in [0.2, 1]$.

We train the composed SOM (both sides concatenated), simulate a hemispherectomy, and retrain each independent half. During the initial training, each input has two candidate representations. Symmetry is broken at random but aided by the distinct primings ($p_P^L \neq p_P^R$). The responding side becomes more apt at representing the shown stimulus, thus amplifying the asymmetry. Figure 5(a) shows the distance to the optimal representation for left ($d_L^{rt,L}$, black) and right ($d_L^{rt,R}$, red) SOMs after hemispherectomy and retraining. For very symmetric priming ($p_P^L \simeq p_P^R$) both sides might fail to establish a functioning map, even after individual retraining (Figure 5(b)). Hence symmetry breaking and the emergence of a dominant side are preferred.

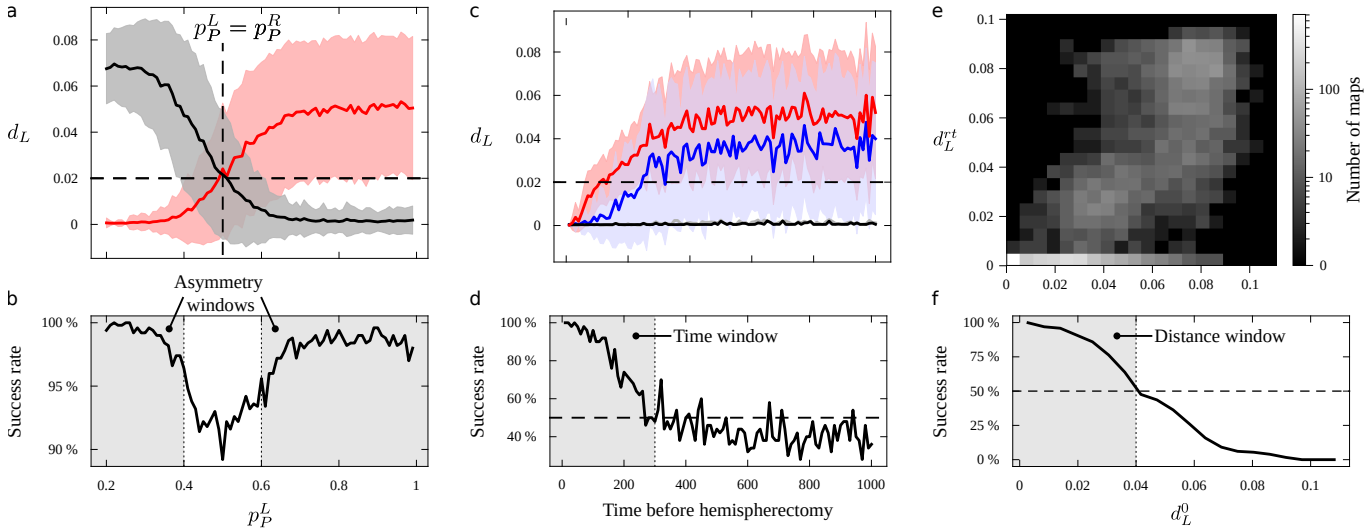


FIG. 5 **Hemispherectomy in redundant representations.** Two half SOMs with $\sigma_t = 10$ and $N_h = 100$ units were concatenated and trained with fully linear ($p_T = 0$) inputs with $L = 100$ and $\sigma_s = 2$. (a) Plots represent distance to the optimal linear map for left (black) and right (red) half SOMs after $T_t = 1000$ training steps, BHEM, and $T_{rt} = 100$ retraining steps. Shading indicates standard deviation over 50 repeats. The right SOM was primed with $p_P^R = 0.5$ and the left one with $p_P^L \in [0.2, 1]$ (horizontal axis). (b) Proportion of cases in which a healthy map develops (arbitrarily defined as reaching $d_L^{rt,R/L} < 0.02$). (c-f) We perform BHEM with fixed priming ($p_P^R = 0.5$ and $p_P^L = 0.8$, making left dominant) after varying training times $T_t \in [10, 1000]$ with $T_{rt} = 100$ retraining steps. (c) Plots show distance to optimal linear map for the left SOM before (black) and after (gray) BHEM and retraining (hardly distinguishable), and for the right SOM before (red) and after (blue) BHEM and retraining. Shading (omitted for the left SOM) represents standard deviation over 50 repeats. (d) Proportion of repeats in which the right (dominated) SOM reorganizes into a healthy linear map. (e) Heatmap shows amount of repeats whose right SOM presented a given d_L^0 before BHEM (horizontal axis) and reached a given d_L^{rt} after BHEM and retraining (vertical axis). (f) Proportion of cases that got reorganized into a healthy linear map as a function of d_L^0 right before BHEM.

We next executed BHEM after a varying training times T_t, modeling surgery at different ages, making 'left' the dominant side. The left SOM achieves consistently a good representation (Figure 5(c)). The right SOM's map degrades progressively before BHEM (5(c)) and remains faulty after retraining (5(c)). Degradation is accentuated the latter that BHEM takes place. There is a time window ($T_t \lesssim 300$) within which the dominated SOM recovers successfully in around 50% of attempts (Figure 5(d)). But this is not strict – successful recovery in delayed BHEM remains high ($\sim 40\%$ and fluctuates above 50% repeatedly, even for long T_t). A more parsimonious behavior is observed as a function of the dominated map's quality right after BHEM (Figure 5(e-f)). There is a clear degradation threshold ($d_L^{h*} \simeq 0.04$) beyond which the likelihood of recovery drops below 50% and vanishes steadily. Hence the main contribution towards a critical period does not stem from lost plasticity in our model, but from an ongoing degradation of precursory structures in the dominated side. These drift apart because they are idle or adapting to other functions.

IV. DISCUSSION

The origin of the drastic anatomical transformations after hemispherectomy present a considerable theoretical challenge. BHEM is not only the most remarkable example of brain plasticity: it also offers a unique opportunity to interrogate the origins of the topological mapping between the neural substrate and the emerging representations of input stimuli.

By modeling reorganization of somatosensory cortices, we find two preferred ways in which a healthy topographic representation is reorganized into a smaller, also healthy, map after BHEM:

- **Linear-to-linear** (Figure 4(a-b)): The intact representations of an optimal linear map are *squeezed* into a smaller region of the neural substrate to make room for orphan stimuli, which are incorporated gradually and in order. This solution appears parsimonious, since the topology of the map is preserved; but it entails a system-wide rewiring of both intact and orphan representations.

- **Linear-to-bilateral** (Figure 4(c, f-g)): Each orphan, lateralized stimulus is assimilated by the SOM unit that represents its mirror-symmetric counterpart. The topological nature of the map is changed, but the actual rewiring needed is less: intact representations are not displaced and only local adjustments are required.

We prime our networks to help them acquire better

representations. This is biologically justified: newborns already have a blueprint of their somatic maps, which matures further as they grow. This blueprint was built over the course of evolution. We find that priming adds a harsh topological constraint: the linear-to-linear reorganization pathway becomes ruled out, but the linear-to-bilateral path is very robustly enforced. Furthermore, primed maps conflict with strongly bilateral input stimuli, resulting in pathological representations before BHEM. The robustness of the bilateral attractor after hemispherectomy suggests this intervention to amend defective maps – beyond their current use.

This translates into neuroscientific insights: The lateral-to-bilateral remapping seems a robust, ready-made solution, suggesting why the drastic BHEM has relatively minor cognitive and behavioral impact. Recent fMRI scans indicate that hemispherectomized brains tend to reorganize following this path [21] – not only for somatosensory networks, but for others such as vision, attention, etc. The same study finds evidence that deviations from this preferred remapping result in worst cognitive performance.

Before BHEM, stimuli with some bilateral symmetry elicit responses in both contra- and ipsilateral areas. Control takes place on the contralateral hemisphere, as activation in the ipsilateral one does not exceed a threshold. In our model, this bilateral activity works as a second priming during training, predisposing far away areas to respond. In the absence of bilaterality, the preferred form of reorganization is a gradual shift of intact representations. A conflict between local (following the linear geometry) versus bilateral reorganization has been observed after stroke [29]. Immediately after localized lesions, the ipsilateral (unaffected) hemisphere takes over control of an orphaned limb (e.g. a hand). Patients in which the local map around the injury regains control (presumably by locally reintegrating the orphan limb) recover better. If, instead, the ipsilateral cortex keeps control, recovery is less likely. Tellingly, restriction of movement of the healthy limb and transcranial magnetic stimulation of the unaffected hemisphere (both aimed at switching off the extended ipsilateral control) are promising therapies. Note that this case is different from BHEM: only a patch of map needs reorganization, and the conflict arises from two functioning geometries struggling to take over.

The neural basis of language inspired us to explore how two redundant, competing representations fare. Language develops from bilaterally symmetric circuitry at birth [8, 13, 36], usually into a fully lateralized set of interrelated centers [7, 9–13]. In our model, a failure to break the initial symmetry can prevent both candidates from maturing, thus losing functionality. This is in agreement with clinical evidence that observes a correlation of symmetry in some language areas across hemispheres and language pathology [8, 40].

After removal of the language-dominant hemisphere, adults tend to lose that capacity for good [37]. Young children can regain function by regrowing key structures in the intact, non-dominant side [5, 17, 41–44]. This suggests critical *window periods* (of around six years [14–17]) for function recovery. A frequent interpretation is that plasticity wanes as the brain matures, eventually preventing the re-growth of language centers. But counter-examples exist of older children, even adolescents, regaining full or partial language capacity [17–20, 41]. It has also been counter-argued that the adult brain retains its plasticity [27, 29]. A complete, convincing alternative explanation of language loss and recovery mechanisms is missing.

Our results support a promising hypothesis. The key would not lie in plasticity, but in a decay of precursory structures present at birth in the dominated side. Language loss would become irreversible once an integrity threshold is crossed. Crucially, in our experiments the decay of dominated structures is irregularly linked to age: some runs see a swift decline, others keep a fine integrity throughout time. This translates in inconsistent time-window periods – as observed in the clinical record. After long pre-hemispherectomy training, successful re-growing of the dominated map becomes less likely, but it is still a non-vanishing possibility. Stochastic fluctuations make it very probable in odd cases. Our hypothesis depends on the existence of an innate (yet not fully developed) language capacity, as proposed by Chomsky and others [13, 45, 46].

Our work relies on two key mechanisms: First, stimuli need to be able to reach every cortical center so that, after BHEM, signals can find a path to candidate representations in the opposite side. We propose a *brain-wide broadcast* of certain neural signals, thus aligning our framework with other hypotheses such as the *global workspace theory* of consciousness [47]. Second, candidate representations must compete with each-other [48]. Our model also concurs with a view that plastic brain reorganization is aided by unmasking of previously existing (yet inhibited) responses and guided by cortical activity rather than subcortical rewiring [29, 49]. Despite its minimal set of assumptions, our model provides deep insights about brain reorganization after hemispherectomy, failure thereof, and potential treatment in certain pathological situations. More realistic models based on our premises might allow us to explore these possibilities in clinical setups.

Acknowledgments

The authors thank the Complex Systems Lab members for fruitful discussions. Special thanks to Alvah Bassie for his inspiring ideas. This work was supported by the Spanish Ministry of Economy and Competitiveness, grant FIS2016-77447-R MINECO/AEI/FEDER, and an AGAUR FI 2018 grant. Seoane was funded by the Span-

ish National Reseach Council (CSIC) and the Spanish Ministry for Science and Innovation (MICINN) through a Juan de la Cierva fellowship (grant IJC2018-036694-I) and by the Institute of Interdisciplinary Physics and Complex Systems (IFISC) through the María de Maeztu Program (grant MDM-2017-0711). This work was done in the Fall of 2019 at the Santa Fe Institute, where the magic happens.

References

[1] Borgstein J, Grootendorst C. Clinical picture: half a brain. *Lancet* **359**, 473 (2002).

[2] Devlin AM, Cross JH, Harkness W, Chong WK, Harding B, Vargha-Khadem F, Neville BG. Clinical outcomes of hemispherectomy for epilepsy in childhood and adolescence. *Brain* 126, 556-566 (2003).

[3] Battro A. *Half a brain is enough: The story of Nico.* Odile Jacob (2003).

[4] Pulsifer MB, Brandt J, Salorio CF, Vining EP, Carson BS, Freeman JM. The cognitive outcome of hemispherectomy in 71 children. *Epilepsia* **45**(3), pp.243-254 (2004).

[5] Danelli L, Cossu G, Berlingeri M, Bottini G, Sberna M, Paulesu E. Is a lone right hemisphere enough? Neurolinguistic architecture in a case with a very early left hemispherectomy. *Neurocase* **19**(3), pp.209-231.

[6] Liegeois F et al. Speech and oral motor profile after childhood hemispherectomy. *Brain Lang.* **114**, 126-134 (2010).

[7] Geschwind, N. Language and the brain. *Sci. Am.* **1972**, *226*(4), pp.76-83.

[8] Galaburda, A.M. Anatomic basis of cerebral dominance. In *Brain asymmetry*; Davidson, A.J.; Hugdahl, K.; MIT Press: Cambridge, USA, 1995; pp.51-73.

[9] Catani, M.; Jones, D.K.; Ffytche, D.H. Perisylvian language networks of the human brain. *Ann. Neurol.* **2005**, *57*(1), pp.8-16.

[10] Fedorenko, E.; Kanwisher, N. Neuroimaging of language: why hasn't a clearer picture emerged?. *Lang. Linguist.* **2009**, *3*(4), pp.839-865.

[11] Fedorenko, E.; Nieto-Castanon, A.; Kanwisher, N. Lexical and syntactic representations in the brain: an fMRI investigation with multi-voxel pattern analyses. *Neuropsychologia* **2012**, *50*(4), pp.499-513.

[12] Fedorenko, E.; Thompson-Schill, S.L. Reworking the language network. *Trends Cogn. Sci.* **2014**, *18*(3), pp.120-126.

[13] Berwick, R.C.; Chomsky, N. *Why only us: Language and evolution.* MIT press: Cambridge, USA, 2016.

[14] Woods BT, Teuber HL. Changing patterns of childhood aphasia. *Ann. Neurol.* **3**(3), pp.273-280 (1978).

[15] Woods BT, Carey S. Language deficits after apparent clinical recovery from childhood aphasia. *Ann. Neurol.* **6**(5), pp.405-409 (1979).

[16] Marcotte AC, Morere DA. Speech lateralization in deaf populations: Evidence for a developmental critical period. *Brain Lang.* **39**(1), pp.134-152 (1990).

[17] Hertz-Pannier L, Chiron C, Jambaqué I, Renaux-Kieffer V, Moortele PFVD, Delalande O, Fohlen M, Brunelle F, Bihan DL. Late plasticity for language in a child's non-dominant hemisphere: A pre-and post-surgery fMRI study. *Brain* **125**(2), pp.361-372 (2002).

[18] Smith A. Speech and other functions after left (dominant) hemispherectomy. *J. Neurol. Neurosur. Ps.* **29**(5), p.467 (1966).

[19] Vargha-Khadem F, Carr LJ, Isaacs E, Brett E, Adams C, Mishkin M. Onset of speech after left hemispherectomy in a nine-year-old boy. *Brain* **120**(1), pp.159-182 (1997).

[20] Voets NL, Adcock JE, Flitney DE, Behrens TEJ, Hart Y, Stacey R, Carpenter K, Matthews PM. Distinct right frontal lobe activation in language processing following left hemisphere injury. *Brain* **129**(3), pp.754-766 (2006).

[21] Kliemann D, Adolphs R, Tyszka JM, Fischl B, Yeo BT, Nair R, Dubois J, Paul LK. Intrinsic Functional Connectivity of the Brain in Adults with a Single Cerebral Hemisphere. *Cell Reports* **29**(8), pp.2398-2407 (2019).

[22] Paul RL, Goodman H, Merzenich M. Alterations in mechanoreceptor input to Brodmann's areas 1 and 3 of the postcentral hand area of Macaca mulatta after nerve section and regeneration. *Brain Research* **39**(1), pp.1-19 (1972).

[23] Merzenich MM, Kaas JH. Reorganization of mammalian somatosensory cortex following peripheral nerve injury. *Trends Neurosci.* **5**, pp.434-436 (1982).

[24] Merzenich MM, Kaas JH, Wall J, Nelson RJ, Sur M, Felleman D. Topographic reorganization of somatosensory cortical areas 3b and 1 in adult monkeys following restricted deafferentation. *Neuroscience* **8**(1), pp.33-55 (1983).

[25] Merzenich MM, Nelson RJ, Stryker MP, Cynader MS, Schoppmann A, Zook JM. Somatosensory cortical map changes following digit amputation in adult monkeys. *J. Comp. Neurol.* **224**(4), pp.591-605 (1984).

[26] Merzenich MM, Recanzone G, Jenkins WM, Allard TT, Nudo RJ. Cortical Representational Plasticity. *Neurobiology of Neocortex*, pp.41-67 (1988).

[27] Pons TP, Garraghty PE, Ommaya AK, Kaas JH, Taub E, Mishkin M. Massive cortical reorganization after sensory deafferentation in adult macaques. *Science* **252**(5014), pp.1857-1860 (1991).

[28] Rivera-Rivera PA, Rios-Lago M, Sanchez-Casarrubios S, Salazar O, Yus M, GonzÃ¡lez-Hidalgo M, Sanz A, Avecillas-Chasin J, Alvarez-Linera J, Pascual-Leone A Oliviero A. Cortical plasticity catalyzed by prehabilitation enables extensive resection of brain tumors in eloquent areas. *J. Neurosurg.* **126**(4), pp.1323-1333 (2017).

[29] Pascual-Leone A, Amedi A, Fregni F Merabet LB. The plastic human brain cortex. *Annu. Rev. Neurosci.* **28**, pp.377-401 (2005).

[30] Armentrout SL, Reggia JA, Weinrich M. A neural model of cortical map reorganization following a focal lesion. *Artif. Intell. Med.* **6**(5), pp.383-400 (1994).

[31] Sutton III GG, Reggia JA, Armentrout SL, D'Autrechy CL. Cortical map reorganization as a competitive process. *Neural Comput.* **6**(1), pp.1-13 (1994).

[32] Weinrich M, Armentrout S, Reggia J. A neural model of recovery from lesions in the somatosensory system. *J. Neurol. Rehabil.* **9**(1), pp.25-32.

[33] Spitzer M, Böhler P, Weisbrod M, Kischka U. A neural network model of phantom limbs. *Biol. Cybern.* **72**(3), pp.197-206 (1995).

[34] Lytton WW, Stark JM, Yamasaki DS, Sober SJ. Computer Models of Stroke Recovery: Implications for Neurorehabilitation. *The Neuroscientist* **5**(2), pp.100-111 (1999).

[35] Kohonen T. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **43**(1), pp.59-69 (1982).

[36] Perani, D.; Saccuman, M.C.; Scifo, P.; Anwander, A.; Spada, D.; Baldoli, C.; Poloniato, A.; Lohmann, G.; Friederici, A.D. Neural language networks at birth. *Proc. Nat. Acad. Sci.* **2011**, *108*(38), pp.16056-16061.

[37] Harrington, A. Unfinished business: models of laterality in the nineteenth century. In *Brain asymmetry*; Davidson, A.J.; Hugdahl, K.; MIT Press: Cambridge, USA, 1995; pp.3-27.

[38] Swanson, L.W. What is the brain?. *Trends Neurosci.* **2000**, *23*(11), pp.519-527.

[39] Seoane LF. Fate of Duplicated Neural Structures. *Entropy*, 22(9), p.928.

[40] Bishop, D.V. Cerebral asymmetry and language development: cause, correlate, or consequence?. *Science* **2013**, *340*(6138).

[41] Liégeois F, Connelly A, Baldeweg T, Vargha-Khadem F. Speaking with a single cerebral hemisphere: fMRI language organization after hemispherectomy in childhood. *Brain Lang.* **106**(3), pp.195-203 (2008).

[42] Ivanova A, Zaidel E, Salamon N, Bookheimer S, Uddin LQ, de Bode S. Intrinsic functional organization of putative language networks in the brain following left cerebral hemispherectomy. *Brain Struct. Funct.* **222**(8), pp.3795-3805 (2017).

[43] Barceló-Coblijn L, Serna Salazar D, Isaza G, Castillo, Ossa LF, Bedia MG. *Netlang*: A software for the linguistic analysis of corpora by means of complex networks. *PLoS one* **12**(8), p.e0181341 (2017).

[44] Piattelli-Palmarini M. Normal language in abnormal brains. *Neurosci. Biobeh. R.* **81**, pp.188-193 (2017).

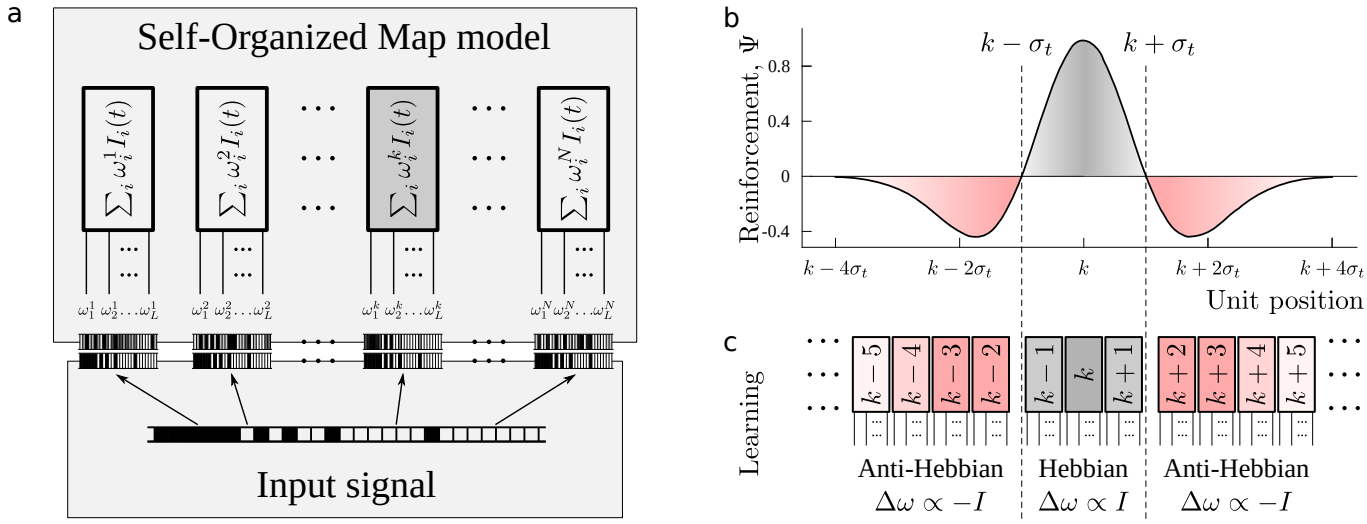
[45] Bickerton D. *Language and species*. University of Chicago Press, (1990).

[46] Bickerton D. *More than nature needs*. Harvard University Press, (2014).

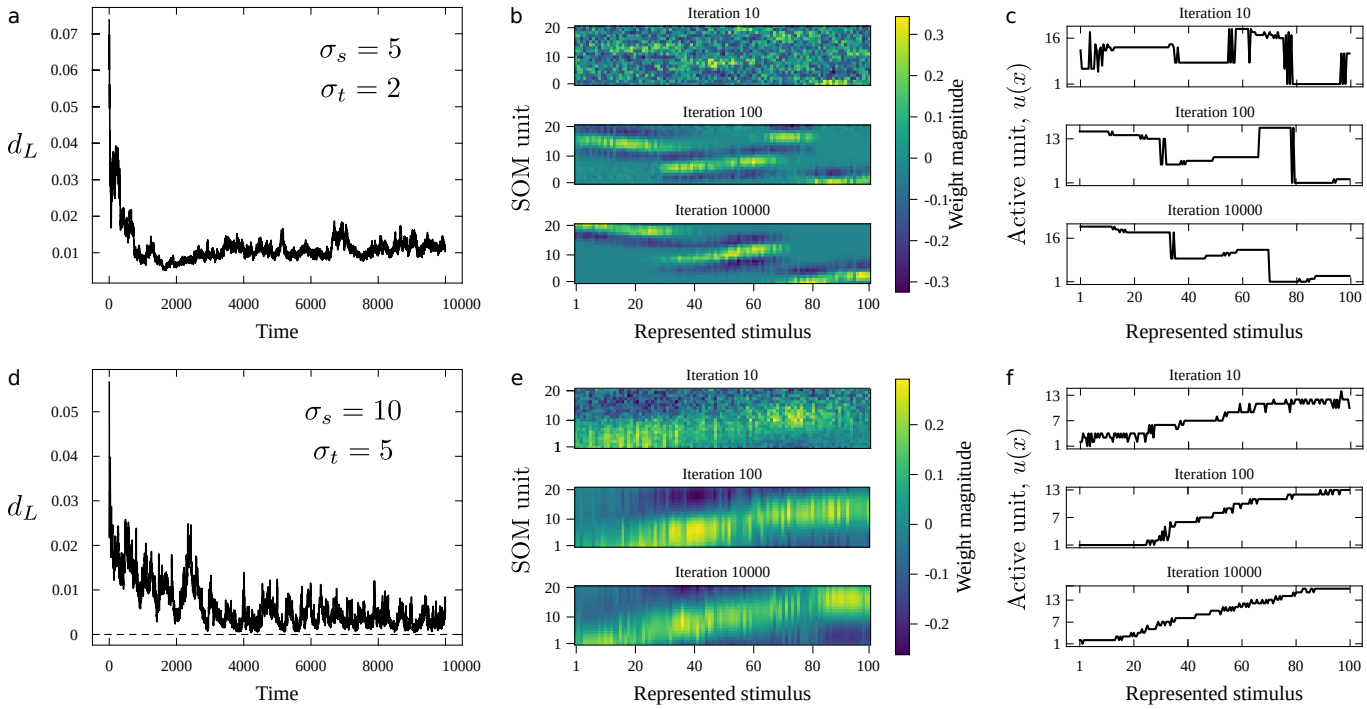
[47] Dehaene S. *Consciousness and the brain: Deciphering how the brain codes our thoughts*. Viking Press, New York City. (2014).

[48] Kappel D, Nessler B, Maass W. STDP installs in winner-take-all circuits an online approximation to hidden Markov model learning. *PLoS Comput Biol* **10**(3), p.e1003511 (2014).

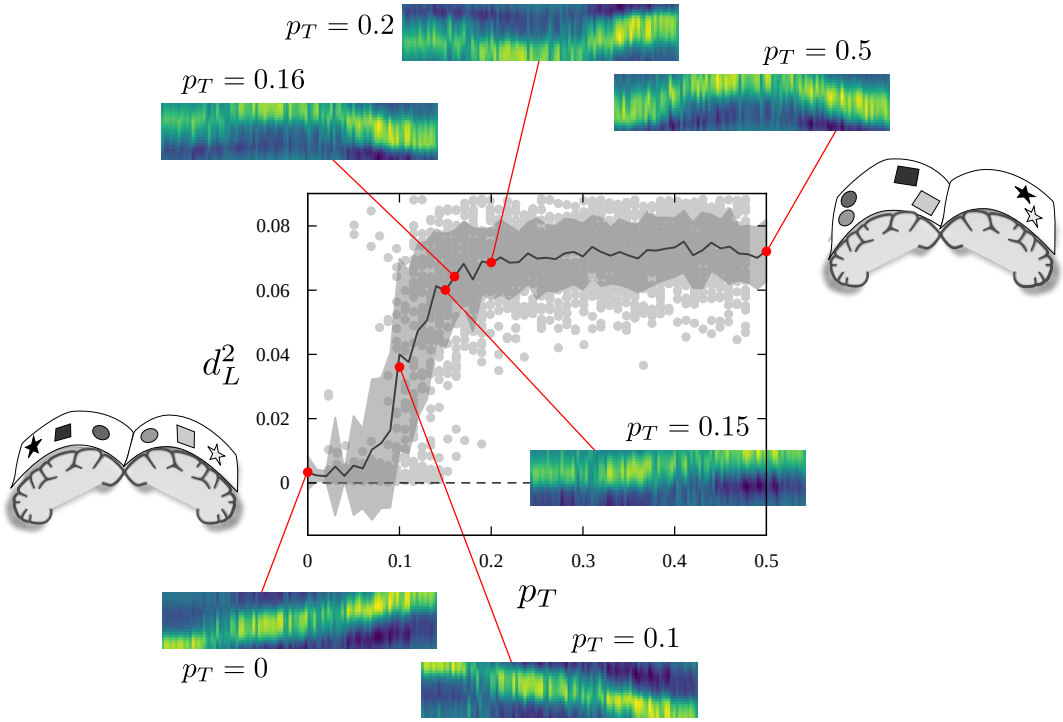
[49] O'Leary DD, Ruff NL, Dyck RH. Development, critical period plasticity, and adult reorganizations of mammalian somatosensory systems. *Curr. Opin. Neurobiol.* **4**(4), pp.535-544 (1994).



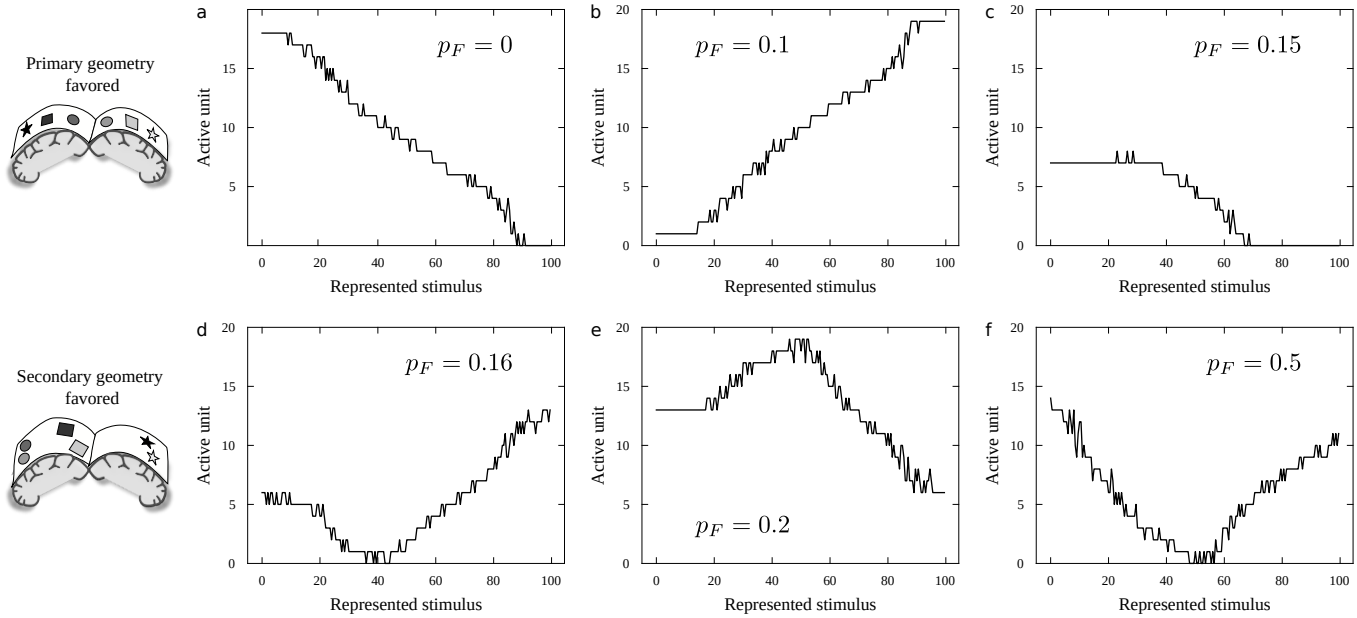
SUPORTING FIGURE 1 **Self-Organized Maps.** (a) A one-dimensional binary input is presented to all units of a SOM. Each unit has a set of weights, and one of these sets matches the input signal better than the others (shaded unit, labeled k). This unit becomes active and the others, inactive. (b, c) If the SOM is in learning mode, the active unit (k) and others in a neighborhood will change their weights in the direction of the current input. Units beyond the neighborhood will updates their weights in the opposite direction. The increase in each unit's weights is given by a Mexican hat function with width σ_t



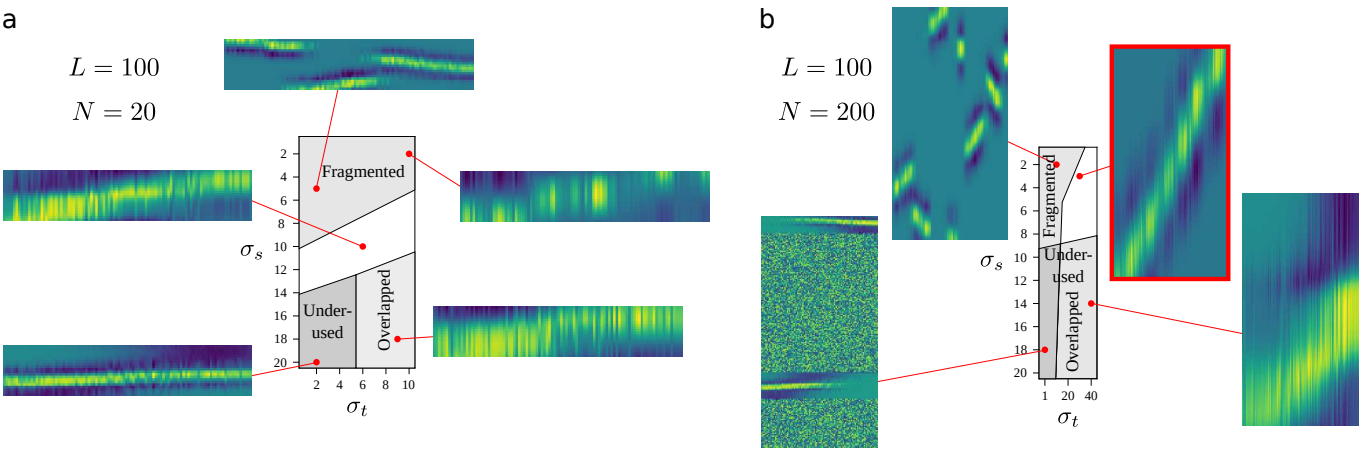
SUPORTING FIGURE 2 **Training Self-Organized Maps for $T_t = 10^4$ iterations.** (a-c) Convergence to a fragmented representation for $L = 100$, $N = 20$, $\sigma_s = 5$, and $\sigma_t = 2$. (a) Performance over time, as measured by the distance of the emerging representation to an optimal linear map d_L. Average performance stalls at above $d_L \simeq 0.01$. (b) These heatmaps show the weights for each SOM unit. The horizontal axis represents position in the stimuli space ($x^0 \in [0, L]$). The vertical axis represents units in the SOM. The small size of the reinforced neighborhood during training ($\sigma_t = 2$) favors that several good representations are nucleated simultaneously along the extent of the SOM. These locally good representations become difficult to flip into a globally coherent one, hence the SOM fails to converge to an optimal solution. (c) Those weights result in a unit that is active when the proper input signal is shown. Fragmentation in the weights results in active units far apart across the SOM for stimuli drawn from very close positions along the stimulus space. (d-f) Convergence to a fine representation for $L = 100$, $N = 20$, $\sigma_s = 10$, and $\sigma_t = 6$. (d) Performance over time falls below $d_L \simeq 0.01$, resulting in more optimal representations than the previous case. (e) This is possible due to the coherence brought about by larger correlations both within input signals and between SOM units. (f) Eventually, the SOM representation varies monotonously as stimuli shift smoothly over their range.



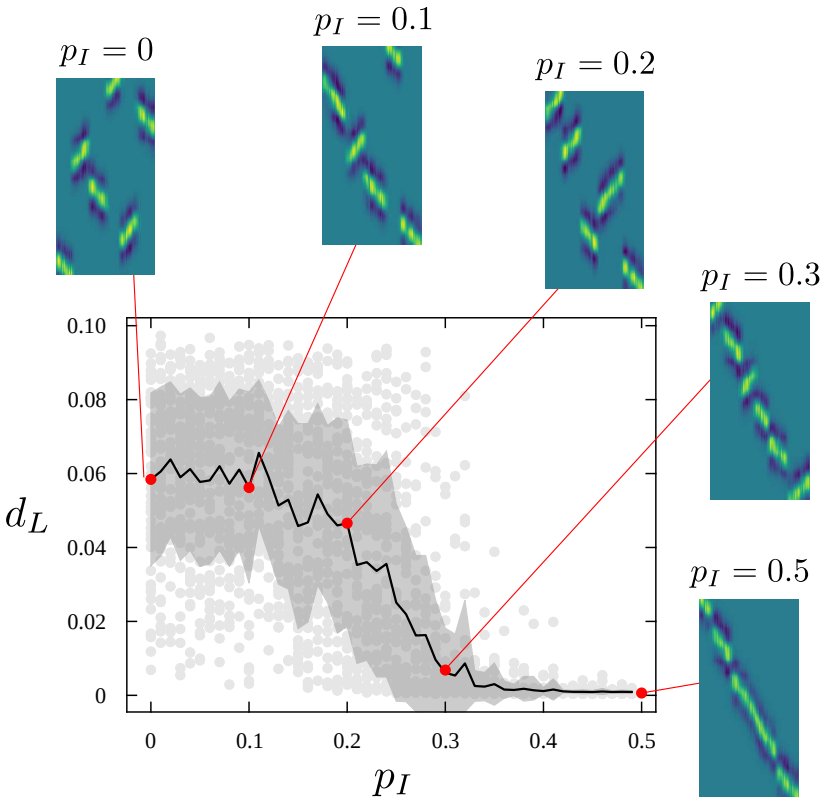
SUPORTING FIGURE 3 **Transitions between representations favoring, respectively, primary and secondary geometries.** Maps with $L = 100$, $N = 20$, $\sigma_s = 5$, and $\sigma_t = 2$ after 10^4 training steps. SOMs attempt to capture the underlaying geometry of its input signals. When the input can be characterized by a single kind of geometry, SOMs accomodate it more easily than when different geometric symmetries are at play. For low bilaterality in the input, SOMs ignore that component and linear maps emerge. As input bilaterality increases, a conflict arises. Eventually, SOMs ignore the linear geometry and converge to the optimal representation for fully bilateral signals – i.e. each SOM unit responds to signals peaked at some x^0 and to its mirror symmetric signal, peaked at $L - x^0$. Note that this is reflected, for each SOM unit, in weights with sensibility to two mirror symmetric input regions.



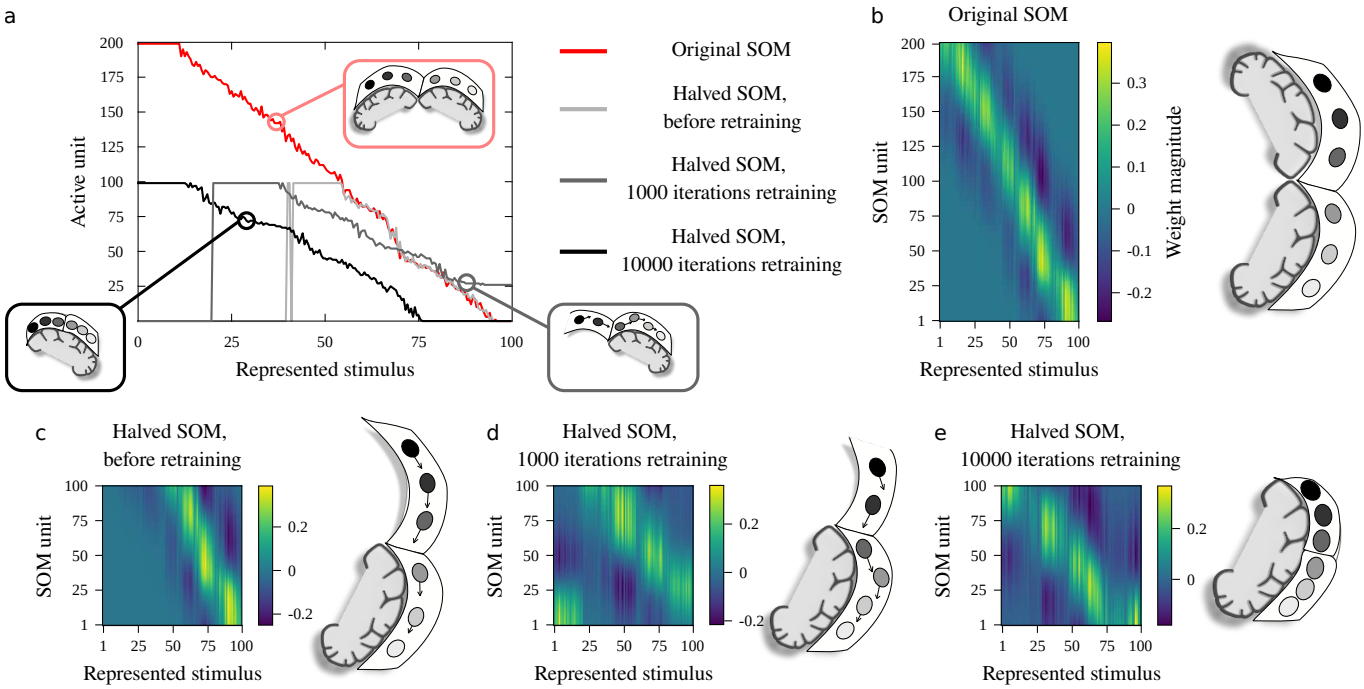
SUPORTING FIGURE 4 **Active units in transitions between primary and secondary geometries.** Maps with $L = 100$, $N = 20$, $\sigma_s = 5$, and $\sigma_t = 2$ after 10^4 training steps. The weights depicted in Supporting Figure 3 determine which will be the active unit when a stimulus peaked at a x^0 is presented. These active units follow the geometric hints that the SOM is capturing better. (a-b) The primary, linear geometry is mostly captured, resulting in a linear relationship between x^0 and the index of the active SOM unit. (c-e) At the transition between both geometries, poorer maps emerge. (f) As the bilateral geometry becomes dominant, it is reflected in the active units as well. Again, each unit responds to stimuli peaked at a position x^0 and its mirror symmetric counterpart $L - x^0$.



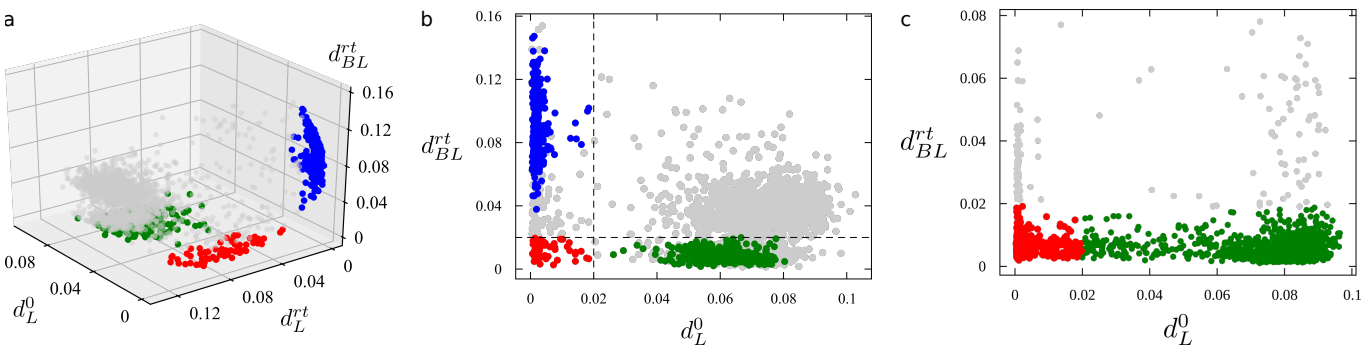
SUPORTING FIGURE 5 **Emerging representations as a function of σ_t and σ_s.** (a) $L = 100$ and $N = 20$. (b) $L = 100$ and $N = 200$. Relatively small σ_t and σ_s mean that correlations between adjacent input positions or nearby units are short range. This often fails to give the SOM a global coherence and results in fragmented representations. Large σ_t and σ_s can induce global coherence, but at the expenses of generating units with very broad receptive fields that greatly overlap with the receptive fields of other units. If the correlations between units are much smaller than correlations within signals ($\sigma_t \ll \sigma_s$), most units become unused. We generated these maps to inspect visually the quality of the emerging representations. Shaded areas (whose limits are drawn as an illustration) roughly delimit where each kind of faulty representation arises. A balance is possible under the right circumstances – e.g. $\sigma_s = 3$, $\sigma_t = 20$, framed in red. These model parameters were used for a first round of BHEM simulation.



SUPORTING FIGURE 6 **Transition from fragmented to fine-tuned representations.** SOMs with relatively low σ_t and σ_s result in fragmented representations because they cannot establish a global coherence (so that multiple, conflicting linear representations start to emerge – but no mechanism exist to align them all in the same direction). Usually, SOMs weights are initialized randomly before training. If, instead, we initialize them with a hint of the global order, a finely-tuned yet global representation emerges more easily. A phase like transition takes us from fragmented representations to globally coherent ones as p_P (the parameter controling the priming of initial weights) increases. Plot shows distance to optimal linear map. Shading represents standard deviation over 50 repeats.



SUPORTING FIGURE 7 **Retraining a SOM to recover a linear representation of the stimuli space.** (a) Active units as a function of x^0 in the complete SOM before BHEM (red), and in the halved SOM right after BHEM (light grey), after 10^3 retraining steps (medium grey), and after 10^4 retraining steps. (b-e) SOM weights at each of these stages: (b) complete SOM after training, and halved SOM (c) right after BHEM, (d) after 10^3 retraining steps, and (e) after 10^4 retraining steps. Each set of weights is accompanied by a cartoon representation of the brain. Note that, in order to keep active units as the variable that depends on the stimuli x^0, the cartoon representation is presented sideways.



SUPORTING FIGURE 8 **Quantifying what reorganization paths are taken.** Each circle represents a BHEM experiment. We depict the distance of the SOM before BHEM to the optimal linear map (d^0_L) and to the optimal bilateral map after BHEM and retraining (d^{rt}_{BL}). Additionally, in (a) we show the distance to the optimal linear map after BHEM and retraining (d^{rt}_L). (a-b) Unprimed SOMs with $L = 100$, $\sigma_s = 3$, $N = 200$, and $\sigma_t = 30$ after 10^4 training and 10^3 retraining steps. Gray circles started as linear (low d^0_L) and ended as linear d^0_L representation, meaning they got reorganized by squeezing. Green started as bilateral representations (low d^0_{BL}, not shown) and ended as bilateral representations as well (low d^{rt}_{BL}). Likely, these maps also reorganized themselves by squeezing – but they ignored the linear geometry from the beginning. Red cases started out as linear representations and ended up as bilateral representation. They got reorganized by folding. (c) Primed SOMs ($p_P = 0.5$) with $L = 100$, $\sigma_s = 2$, $N = 100$, and $\sigma_t = 10$. Same training and retraining times. In red, cases that started out as linear and ended up as bilateral representations (i.e. they reorganized themselves by folding). This attractor in halved SOMs is much more stable. Many more cases end up with good bilateral representations disregarding of their starting point (green).