

# Democratic Integration: A Theory of Adaptive Sensory Integration \*

Jochen Triesch

The University of Rochester  
Computer Science Department  
Rochester, New York 14627

Technical Report 00.1

February 2000

**Keywords:** sensory integration, adaptation, neocortex, modular networks, correlations, synaptic plasticity, McGurk effect, Democratic Integration

## Abstract

The human brain has to integrate the inputs it receives from different sensory modalities into a coherent description of its environment. This integration is often adaptive, showing recalibration or suppression of discordant sensory modalities. This paper proposes a qualitative theory of sensory integration which relates these adaptation phenomena to the anatomy of the neocortex and a rapid reversible synaptic mechanism, as proposed in von der Malsburg's correlation theory of brain function [19].

## 1 Introduction

Our brains receive inputs from different sensory modalities such as vision, audition, olfaction, taste, touch, proprioception and so on. Each of these senses employs a large number of individual sensors. For instance, our retinas contain more than 100 million rods and cones each. This immense number of inputs has to be constantly integrated into a coherent description of the world since many tasks can only be achieved by paying attention to many of the modalities.

**Bower's taxonomy.** In his essay on "The Unity of the Senses" [1], Bower argues that the information supplied to our different senses is usually concordant. He gives the example of a buzzing bee on a fragrant rose which is heard and seen in the same place as the source of the odor. This concordance is a property of the physical world and is a prerequisite for sensory

---

\*This article is a slightly extended version of a chapter of [17].

integration. For the purpose of studying the nature of sensory integration Bower stresses the importance of situations with discordant information between the sensory modalities. For example, in an experiment on sensory integration a subject might look at a sound source through wedge prisms which visually displace the object. It is no longer seen in the same place as it is heard. In these cases the integration of information by the brain can show adaptive phenomena, i.e., the way in which information from different sensors is combined changes adaptively in response to the discordances in the current scene. The time scale of this adaptation can be of the order of a second [21]. Research in robotics suggests that discordances between sensors are the rule rather than an exception in sensory systems, as every sensor has only a finite precision and a non-vanishing probability of complete failure. Furthermore, the probability of both of these types of errors depends on the current context. For example, the cones in the human retina may fail completely when the light intensity falls below a certain threshold (due to a light bulb suddenly going out) whereas the more sensitive rods may still be operational. Another example is the use of ultrasonic sensors in a robot. They produce very noisy readings when used in an environment with many smooth, hard surfaces producing specular reflections while being quite reliable in other environments.

Bower distinguishes the following four *levels of unity* of the senses, which have recently been interpreted as different *modes of integration* by Murphy [12]:

1. *complete unity*: The sensors of different modalities are integrated without a possibility of detecting discordances between them. The result is a compromise between the different sensors.
2. *unity with awareness of discordance and the possibility of recalibration*: The sensors are integrated in a manner that allows the detection of discordances and a subsequent recalibration of some sensors.
3. *unity with awareness of discordance and a tendency towards suppression*: The discordances between sensors are coped with by suppressing the offending sensors.
4. *no unity at all*: The sensors are not integrated at all and hence there is no way of detecting any discordances. This trivial case will not be discussed in the following.

Although Bower's scheme is supported by some evidence from experiments with animals and humans [1, 12], future research in this area is needed.

**Anatomical Substrate.** The neocortex may be divided into different areas, which can be defined according to their cyto-architecture, their physiology, or their connections to other regions of the brain. For instance, in the visual system of the macaque brain, whose organization is very similar to that of the human brain, more than 30 areas have been distinguished [18]. They show a functional specialization, i.e., they are responsible for computing and representing particular aspects of the visual input. A hallmark of cortical organization seems to be the high degree of interconnectedness of these areas. Van Essen *et. al.* [18] report that almost one third of the theoretically possible connections between the areas in the macaque visual system have already been identified. This is supported by studies of Braitenberg and Schüz [2] about the statistics of synaptic connections in the

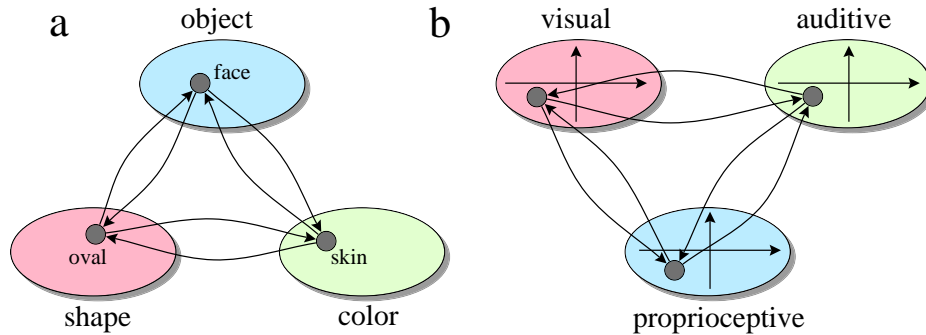


Figure 1: Complete integration. **a:** Three reciprocally connected areas which could appear in a visual system. Neurons with particular tuning characteristics are combined to an attractor due to excitatory connections. Usually, many attractors will exist which may compete and/or overlap. The activity in the FACE unit integrating shape and color information is a compromise between its two inputs. The OVAL and SKIN unit will compromise between their inputs from lower areas (not shown) and the top-down expectation from the FACE unit. **b:** Three areas from different modalities which code the location of a stimulus in body centered coordinates. Each neuron in one of the areas codes a specific position. Neurons referring to the same physical position are connected reciprocally as shown for one example position. Attractors referring to different physical locations are competing. If discordant stimulation is present, the system could respond by either showing some activity in multiple attractors or by running into a single attractor which receives the most support from the senses and suppressing the others by inhibition. Which activity type is present will depend on the relative strength of the sensory signals driving the neurons, the associative signals between them and the strength of the inhibition between attractors. However, in both cases the system compromises.

neocortex. They come to the conclusion that the neocortex is a “mixing machine”. Another important feature of neocortical organization seems to be that the connections between areas tend to be reciprocal, i.e., if an area A projects to an area B, the reciprocal connections from B to A are very likely to be present as well [18].

**Goal of the Paper.** The goal of this paper is to propose a qualitative theory which relates the adaptation phenomena to the gross architecture of the neocortex and a single fundamental principle of synaptic plasticity, which is an extension to the Hebb rule [9]. In a nutshell, the theory assumes that different cortical modules try to agree on a coherent description of the environment by exchanging their individual interpretations via the reciprocal connections. At the same time, the connections between individual modules are adapted in response to discordances, resulting in suppression and recalibration phenomena, which restore the system’s overall coherence.

## 2 Democratic Integration in Modular Networks

Consider the two networks in Fig. 1. Network **a** is a caricature of part of a visual system. Network **b** comprises areas representing locations of objects in body centered coordinates as perceived by different modalities. I do not claim that areas of this kind exist in the human cortex; they are to be seen merely as a metaphor. In both cases three different cortical areas with particular functional specializations are depicted and they are reciprocally connected. All of them are also connected to many other areas not shown. The areas contain a number of neurons (only one of which is shown in each case) which are tuned to particular stimulus properties. For example, a neuron in the “shape” area may be tuned to an oval-shaped stimulus. One may prefer to think of cortical columns instead of single neurons; the connections will then represent bundles of fibers between columns. Mechanisms for segmentation and attentional focusing are assumed, that ensure that the network deals with one object at a time (binding by temporal correlations) if several are in the field of view. Depending on their logical relation, i.e., whether their stimulus properties may be combined in natural scenes or not, neurons within one area will have either excitatory or inhibitory connections, which are assumed to be fixed for the sake of simplicity of the argument. (This point will be taken up in the discussion.) Neurons in different areas are connected reciprocally by excitatory connections if the features that they code for are regularly combined in natural stimuli. In Fig. 1a the set {FACE, OVAL, SKIN} is connected in this manner and forms an attractor of the activity dynamics. In Fig. 1b neurons referring to the same physical location constitute such an attractor. Note that due to the different time-delays of the intra- and interarea connections the networks will show complicated temporal dynamics, whose discussion is beyond the scope of this paper.

**Complete Integration.** The FACE neuron in the object area of Fig. 1a integrates the inputs from two different modalities specialized on shapes and colors respectively. It may be said to perform a complete integration of its two inputs. Its activity will be high if the OVAL neuron and the SKIN neuron are active simultaneously. If only one of its inputs is active, it will average and show moderate activity. (There is evidence for additive as well as multiplicative integration in the brain; this distinction, however, is not necessary for the purpose of our qualitative discussion.) In turn, the units in the lower areas compromise between their sensory stimulation and the lateral and top-down expectations from the other units. Due to the reciprocal connections the activity dynamics in the three areas favors states where the three neurons are coactive. For further stabilization of the percept the synapses between the units may be strengthened by a Hebbian learning rule [9], which increases those synapses from an active neuron, which take part in firing other neurons. Below, we will extend this learning rule in order to accommodate adaptation by suppression and recalibration.

**Suppression.** Consider the situation in Fig. 2a. We have two competing attractors of neurons: {FACE, OVAL, SKIN} and {GRAPEFRUIT, CIRCLE, YELLOW}. Now discordant information is introduced: for example, the system may be presented with a face painted in yellow color, resulting in initial activation of the set {FACE, OVAL, YELLOW} which does not correspond to one of the attractors. The system may use the following synaptic learning rule for adapting its weights to escape this dilemma and restore coherence:

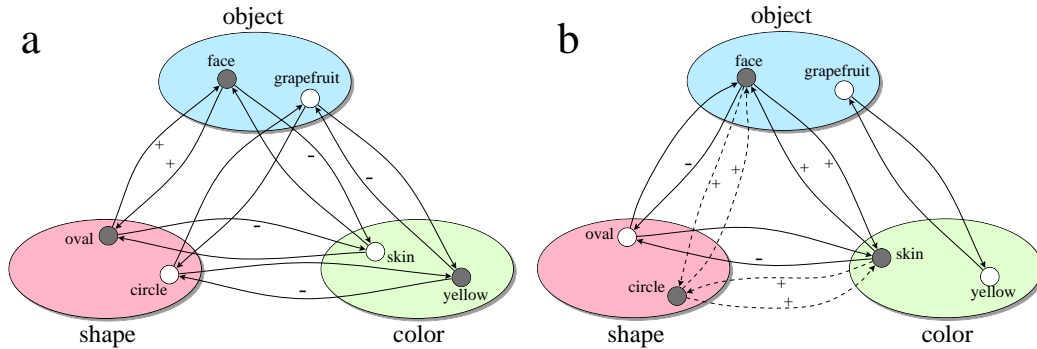


Figure 2: **a**: Integration with suppression. The system is comprised of two competing attractors and is confronted with discordant stimulation (active units are dark, inactive units are light). The system reacts by temporarily modifying its weights accordingly. Weights between active units are strengthened, weights from active to inactive units are depressed (plus and minus signs). The effect is a suppression or decoupling of the color area. **b**: Integration with recalibration. The system comprises the same attractors but also an additional weaker one, {FACE, CIRCLE, SKIN} (dashed connections), with partial overlap. (The connections from the CIRCLE unit to the GRAPEFRUIT and YELLOW unit have been omitted in the figure for clarity.) If the scene contains a round face activating the weaker attractor, the system can stabilize it with the same learning rule, thereby recalibrating itself to the current context (face is round rather than oval).

- temporarily strengthen the synapses between active units
- temporarily weaken the synapses from active to inactive units
- leave the other synapses unchanged

Additionally, small lasting effects are assumed to also strengthen or weaken the temporarily modified synapses permanently just as in von der Malsburg’s theory [19]. I suggest that synapses from inactive to active units are left unchanged. This is thought to keep the system more flexible and capable of switching to new connection patterns more quickly as soon as evidence in this direction appears. The result of such adaptation is a suppression of the discordant sensor. The weights from the YELLOW unit decay so that the color area does not disturb the other areas anymore. Its influence on them is suppressed. Conversely, the synapses from active units in the other areas, which project to inactive units in the color area, are also depressed, thereby no longer disrupting processing in the color area. The color area is decoupled from the rest of the system.

**Recalibration.** Now consider the situation in Fig. 2b. The situation is analogous to that discussed above but additional weaker synapses have been introduced (dashed connections) to reflect the fact that a face may sometimes be of circular shape. Given stimulation of the set {FACE, CIRCLE, SKIN}, the system can recalibrate to the new situation by applying the same synaptic modification rule as above. Due to strengthening of the weak {FACE, CIRCLE, SKIN} attractor and switching off disrupting connections the system recalibrates to the new pattern. More specifically, the FACE neuron is now tuned to a circular skin colored

input pattern. In contrast to the previous case all areas are still involved in the processing, but the synaptic modifications have formed a different attractor.

**McGurk Effect.** If discordant information is introduced which stimulates several attractors equally, excitatory lateral connections between attractors can also lead to an intermediate “compromise” attractor being activated. This situation corresponds to the McGurk effect, where a subject is presented with discordant auditory and visual stimulation of syllables — the visual presentation of moving lips does not match the sound being played. In this condition subjects can reliably perceive a syllable matching neither the visual nor the auditory input [11].

### 3 Discussion

**Adaptive Integration.** The examples of integration in the theory show the following: complete integration of the senses may be modeled by simply connecting different modalities in the style of attractor neural networks with an associative mechanism. Crosstalk between attractors can accommodate McGurk like effects in this architecture. The theory is strongly related to many other works on learning sensory or sensorimotor integration with associative mechanism such as [13]. In [14] an additional gain control for plasticity called value-dependent learning is employed, which is similar to reinforcement learning. In these works, however, no reversible short-term adaptation is incorporated. Democratic Integration states that adaptive suppression and recalibration phenomena can be elegantly explained by endowing such networks with a rapid reversible synaptic learning rule, which is an extension to the Hebb rule as suggested by von der Malsburg [19]: synapses show a rapid reversible change of synaptic efficacy in response to signal correlations. Without this plasticity, it is difficult to imagine how classical attractor neural networks, e.g. Hopfield networks, could account for these adaptation phenomena. They simply appear to be too inflexible. Fast synaptic plasticity of a Hebbian nature has received some experimental support recently. Markram *et. al.* [10] report about the regulation of synaptic efficacy between neighboring layer 5 pyramidal neurons in response to carefully timed spike pairing of pre- and postsynaptic cell, which shows the desired Hebbian aspects. After a learning phase with paired activation the EPSPs produced by the first few spikes arriving at the synapse are significantly enhanced. Along similar lines Zhang *et. al.* [22] have shown that retinotectal synapses in the developing brain of frog exhibit potentiation and depression in response to carefully timed spiking of pre- and postsynaptic neuron. In particular, significant potentiation was shown to occur after less than 20 paired pre- and postsynaptic spikes irrespective of the frequency of stimulation.

Although there appears to be some evidence towards the hypothetical synaptic mechanism it may be useful to speculate about alternatives. The proposed mechanism is not necessarily a prerequisite of the theory. The important ingredient is only a rapid change of the functional coupling between the neurons, i.e., the impact of pre-synaptic spikes on the probability of the post-synaptic neurons firing.

**Working Memory.** The theory is related to the problem of working memory. The synaptic learning rule is a simple extension to the Hebb rule by working also on a fast time-scale.

It has been proposed earlier as the basis of short term or working memory and a manifestation of the short-term binding of features [19, 20, 16]. The context-dependent adaptive integration phenomena can be seen as a facet of working memory. What is “stored” during the adaptation is information about the current sensory context — which modalities should be suppressed or how they should be recalibrated. This leads to the question of time-scales of correlations and synaptic changes. I claim that the memory spans of recalibration phenomena should cover the range from fast sensory registers (1–2s) over short term memory time scales ( $\sim 20$ s) to lasting effects if the environment or the perceptual system itself have undergone a lasting change. The latter being implemented by the small long-lasting effects of the learning rule. This view fits nicely to the result of a recent review by Gilbert [7] on plasticity in visual perception and physiology of the visual system which suggests that the percepts shaped by the cortex are affected by previous visual experiences operating over a wide range of time scales. In particular, the receptive field properties of neurons may change as quickly as in about 1s.

The sensory input is generally processed in a hierarchy of stages with different levels of abstractions. I suggest that for each stage there exists a typical range of time-scales in which the representations can be assumed to remain constant and which in turn defines an appropriate time-scale for the adaptation. For example, in V1 patterns are usually fluctuating very fast in response to natural stimulation so that adaptation is only useful in V1 (if at all) if it is comparably fast. In contrast, in higher areas where, e.g., the identity of a face is represented, this representation should remain stable on the time-scale of the duration of typical encounters between persons.

**Lateral Connections.** In section 2 the lateral connections within one area have been treated as fixed for the sake of simplicity of the argument. The logical relation of the neurons’ preferred stimuli, i.e. whether they may be combined in natural scenes or not, determined whether the neurons were connected excitatorily or inhibitorily. *A priori* there is no reason to assume that these connections must be fixed. On the contrary, one might expect that the excitatory connections within one area are subject to a similar dynamics allowing for rapid reversible changes in response to correlations. The adaptive mechanism could thus produce recalibration and suppression phenomena even within one cortical area. This could be an advantage for experiments aiming at testing the theory with the micro-electrode, since recordings must be taken from one area only. An example worth being studied might be the following: in primary visual cortex there exist long range horizontal connections between columns of similar orientation selectivity which may be related to perceptual grouping [8, 6, 15]. Intracellular recordings are taken from two such connected neurons from different columns. When the system is stimulated with concordant inputs, i.e., a visual input pattern which activates both units, the EPSPs between the cells should become temporarily enhanced. Conversely, if only one of the neurons is brought to fire by the input pattern, the synapse from this neuron to the other should get depressed.

**Development.** An important problem for the integration of information stemming from different modalities is that of a common data format. The theory offered here does not require the areas to have an identical data format (contain neurons with identical preferred stimuli) but only requires connected areas to have developed neurons which regularly fire in a correlated way in response to natural stimuli. How are they to be created? I claim that

during ontogenetic development the areas form cells with these characteristics due to the small long-lasting effects of the learning rule, which shape a neuron's response characteristics in the long run. The system self-organizes the neurons' tuning *together* with the connections between them (network self-organization), leading to a network that favors coherent activity patterns in response to natural stimuli. As the tuning characteristics of neurons in cortex are defined by their afferent connections (including lateral and top-down influences), any change in connectivity produces a shift of a neuron's tuning characteristics. In the developing brain this mechanism forms appropriate neuronal response characteristics together with appropriate connections; in the adult brain learning continues such that the connectivity pattern represents a dynamic equilibrium. This proposition is consistent with the result of a recent review [3] by Cramer and Sur which suggests "significant similarities" between the activity-dependent formation of connection patterns during development and synaptic plasticity in the adult brain. In particular, they conclude that the refinement of connections during development is akin to a Hebbian process, with strengthening of synapses in response to correlated activation and weakening of synapses for uncorrelated activation.

The exploitation of cross-modal structure for learning has also been addressed by DE SA [4, 5]. She considers different modalities learning to solve a classification problem. She introduces the notion of "self-supervised learning", where output labels assigned by one modality are used to train a second modality and vice versa. Her results show that this type of learning outperforms unsupervised learning of the individual modalities and approaches the quality of supervised learning. The main difference to the theory proposed here is that in her architecture, no attempt is made to integrate the results of the two modalities to arrive at an even better classifier. Individual modalities process their input independently and the results are only exchanged for the purpose of learning.

An important problem is at what stage of processing different sensors should be integrated. The brain must make sure that only those areas can directly talk to each other which have something to say to each other, i.e., whose neurons show reliable correlations. For example, it might not be too useful to directly couple primary visual and primary auditory cortex, since the two show no reliable correlations among their neurons. The brain's solution to the problem may be read off the complex wiring patterns of inter-areal connections such as those presented in [18], whose make-up may be determined genetically to a large extent. In the engineering disciplines this problem is often referred to as the question of the level of sensor fusion and has triggered extensive research without a general theory having emerged. Even when the microscopic mechanisms of integrating information in the brain may be as simple as hypothesized above, it may well be that for building artificial perceptual systems, the real challenge is to figure out at what stage sensors are allowed to communicate their results to what other sensors.

**Conclusion.** This paper proposes a theory of adaptive sensory integration in the brain. In Democratic Integration, different brain areas try to agree on a description of the environment while at the same time adapting to the description agreed upon, thereby trying to maintain maximal coherence among the areas. The theory relates adaptive phenomena (suppression and recalibration of modalities) to fast synaptic plasticity as proposed in von der Malsburg's correlation theory of brain function and creates a framework for further investigation of this important problem.

## Acknowledgments

I am indebted to Christoph von der Malsburg and Dana H. Ballard for fruitful discussions and for providing excellent working environments. I would like to thank Jan C. Vorbrüggen for critical discussion of an earlier draft of the manuscript. The author is currently supported by NIH/PHS research grant P41 RR09283.

## References

- [1] T. G. R. Bower. The evolution of sensory systems. In Robert B. MacLeod and Herbert L. Pick Jr., editors, *Perception: Essays in Honor of James J. Gibson*, pages 141–152. Ithaca, NY: Cornell Univ. Press, 1974.
- [2] Valentino Braitenberg and Almut Schüz. *Anatomy of the Cortex*. Springer, 1991.
- [3] Karina S. Cramer and Mriganka Sur. Activity-dependent remodeling of connections in the mammalian visual system. *Current Opinion in Neurobiology*, 5:106–111, 1995.
- [4] V. R. de Sa. *Unsupervised Classification Learning from Cross-Modal Environmental Structure*. PhD thesis, Department of Computer Science, University of Rochester, Rochester, New York, USA, 1994.
- [5] V. R. de Sa and D. H. Ballard. Category learning through multi-modality sensing. *Neural Computation*, 10(5), 1998.
- [6] D. J. Field, A. Hayes, and R. F. Hess. Contour integration by the human visual system: Evidence for a local “association” field. *Vision Research*, 33:173–179, 1993.
- [7] Charles D. Gilbert. Plasticity in visual perception and physiology. *Current Opinion in Neurobiology*, 6:269–274, 1996.
- [8] Charles D. Gilbert and Torsten N. Wiesel. Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *The Journal of Neuroscience*, 9(7):2432–2442, 1989.
- [9] Donald O. Hebb. *The organization of behavior*. Wiley, New York, 1949.
- [10] H. Markram, J. Lübke, M. Frotscher, and B. Sakmann. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, 275:213–215, 1997.
- [11] H. McGurk and J. MacDonald. Hearing lips and seeing voices. *Nature*, 264:746–748, 1976.
- [12] Robin R. Murphy. Biological and cognitive foundations of intelligent sensor fusion. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 26(1):42–51, 1996.
- [13] Helge J. Ritter, Thomas M. Martinetz, and Klaus J. Schulten. Topology-conserving maps for learning visuo-motor-coordination. *Neural Networks*, 2:159–168, 1989.
- [14] Michele Rucci, Giulio Tononi, and Gerald M. Edelman. Registration of neural maps through value-dependent learning: Modeling the alignment of auditory and visual maps in the barn owl’s optic tectum. *The Journal of Neuroscience*, 17(1):334–352, 1989.
- [15] K. E. Schmidt, S. Löwel, R. Göbel, and W. Singer. The perceptual grouping criterion of colinearity is reflected by anisotropies of connections in primary visual cortex. *European Journal of Neuroscience*, 9:1083–1089, 1997.

- [16] J. Triesch and C. von der Malsburg. Binding — a proposed experiment and a model. In *Proceedings of the International Conference on Artificial Neural Networks (ICANN) 1996*, Berlin, Heidelberg, New York, 1996. Springer Verlag.
- [17] Jochen Triesch. *Vision-Based Robotic Gesture Recognition*. PhD thesis, Institut für Neuroinformatik, Ruhr-Universität Bochum, Germany, 1999. Also published as book by Shaker Verlag, Aachen, Germany, ISBN 3-8265-6257-7.
- [18] David C. Van Essen, Charles H. Anderson, and Daniel J. Felleman. Information processing in the primate visual system: An integrated systems perspective. *Science*, 255:419–423, 1992.
- [19] Christoph von der Malsburg. The correlation theory of brain function. Internal Report 81-2, Dept. of Neurobiology, Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany, 1981. Reprinted in: *Models of Neural networks II*, edited by E. Domany, J.L. van Hemmen, and K. Schulten (Springer, Berlin, 1994) Chapter 2, pp. 95–119.
- [20] Christoph von der Malsburg. Binding in Models of Perception and Brain Function. *Current Opinion in Neurobiology*, 5:520 – 526, 1995.
- [21] H. Wallach and D. N. O’Connell. The kinetic depth effect. *J. Experimental Psychology*, 45:205–217, 1953.
- [22] L. I. Zhang, H. W. Tao, C. E. Holt, W. A. Harris, and M. M. Poo. A critical window for cooperation and competition among developing retinotectal synapses. *Nature*, 395(6697):37–44, 1998.