Binding Problem for Input vs. Output Representations and the Role of the Thalamus in Its Solution

As presented at the Computational Neurobiology Lab at Salk: June 14, 2005

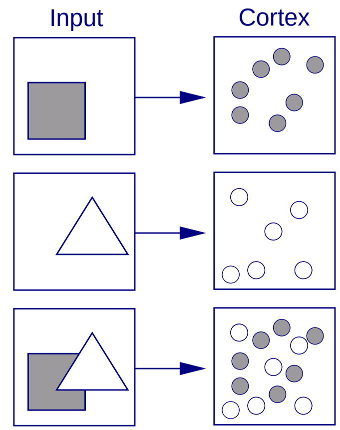
Yoonsuck Choe

Department of Computer Science

Texas A&M University

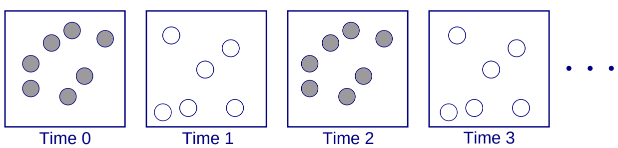
`http://faculty.cs.tamu.edu/choe`

Motivation: The Binding Problem



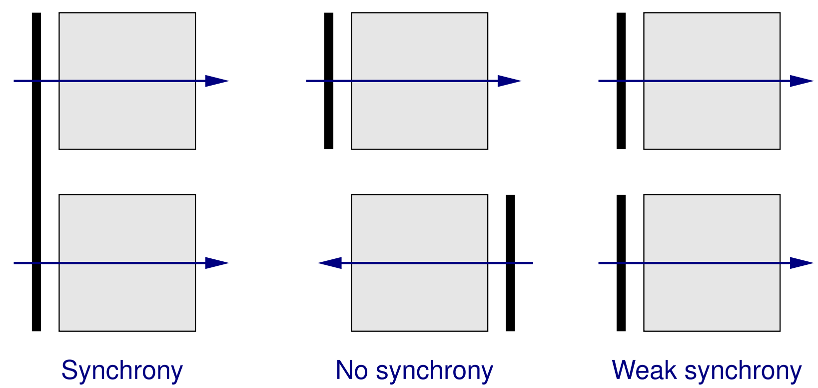
- Distributed representations lead to the **superposition catastrophe** (von der Malsburg 1986).
- How does the brain **piece together** partial representations to form a whole?
- Which feature should go along with which?

Potential Solution to the Binding Problem



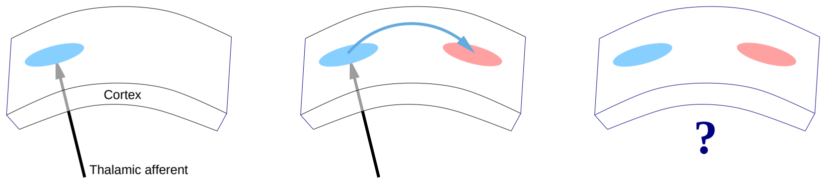
- **Timing** may be important in solving the problem.
- **Interleave** the activity pattern **over time** (von der Malsburg 1986).

Evidence for Temporal Coding



- Gray et al. (1989) and Eckhorn et al. (1988) (and many thereafter) showed that neural representations of coherent object features are **synchronized.**
- **But, that may not be the end of the story!**

The Main Research Question

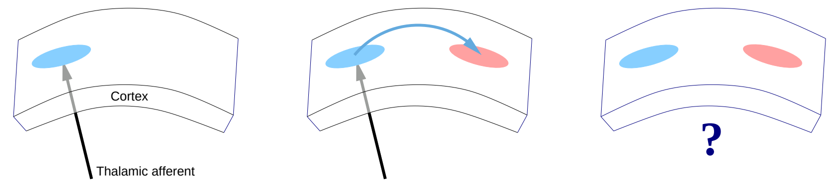


How does the brain **distinguish** between cortical activities that represent:

1. Questions posed to the cortex, and
2. Answers to those questions?

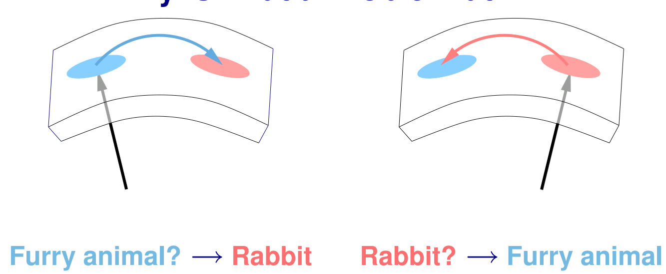
That is, **how can the input and the output of cortical computation be distinguished?**

Input–Output Binding Problem (IOBP)



Similar to the original binding problem, but not between input representations, **but between input and output representations.**

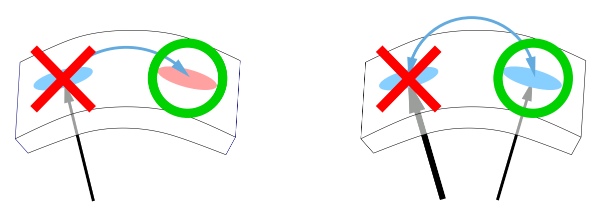
Why Is That a Problem at All?



The problem is nontrivial because:

- The same representation can serve as **both question and answer** at different times, under different contexts.
- The source and the target cortical region will maintain almost **simultaneous activation** while the source region is active.

Possible Answer: Simply Promote the Output

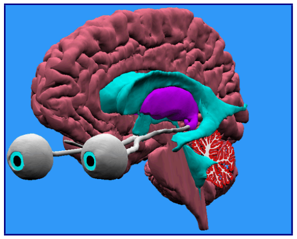


Promote (or propagate) cortical activity that are:

1. **Not input-driven**, or
2. Relatively **less input-driven**.

But, **how (and where)** does the brain achieve this?

Possible Neural Basis: The Thalamus



Some clues:

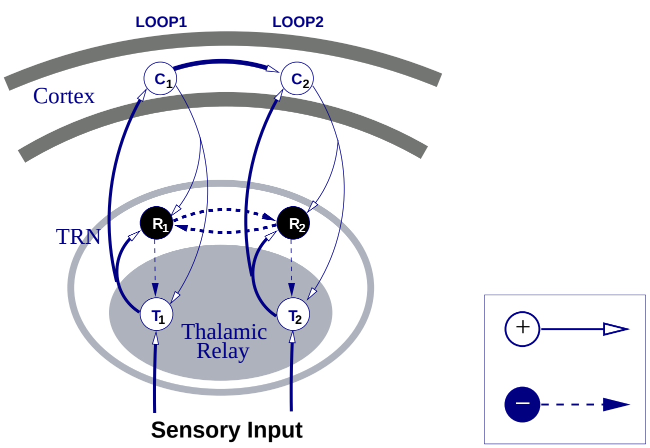
- **Heavy feedback** from the cortex.
- Covered by an inhibitory shell, the **Thalamic Reticular Nucleus (TRN).**

Image Source: http://mail.biocfarm.unibo.it/aunsnc/3dobjb.html

Related Work on the Thalamus

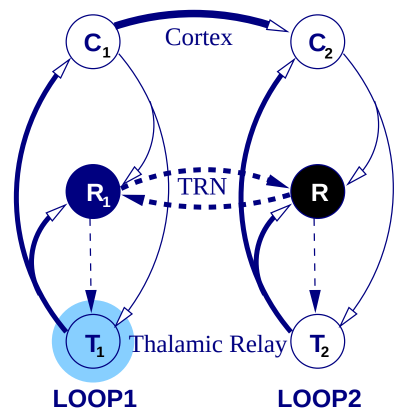
- Sensory relay (see Sherman and Guillery 2001 for a review).
- Sleep rhythms (Destexhe and Sejnowski 2001; Steriade and McCormick 1993; McCormick and Bal 1997) / Epilepsy.
- Synchrony (Llinás and Ribary 1994; Sillito et al. 1994).
- Mediating cortical communication (Guillery and Sherman 2002).
- Cross-modality switching (Crabtree and Isaac 2002).
- Attention (LaBerge 1995; Crick 1984).
- Active blackboard (Mumford 1995; Harth et al. 1987)
- Global workspace (Newman et al. 1997).
- Consciousness (Crick 1984; Taylor 1998).

Dorsal Thalamus-TRN-Cortex Network



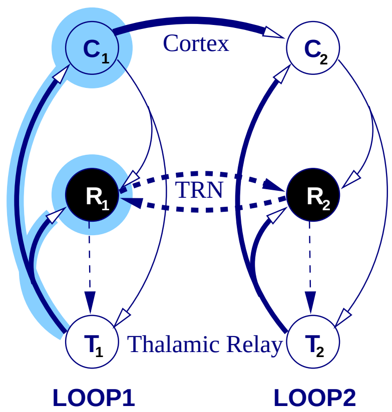
- A candidate circuit can be found in the dorsal thalamus-TRN-cortex circuit: **TRN plays a key role**.

Activation Sequence (1/6)



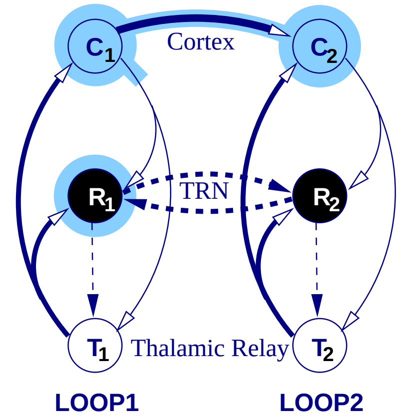
Initially, only T_1 receives an afferent sensory input.

Activation Sequence (2/6)



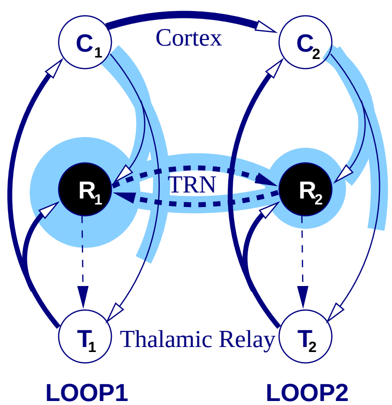
T_1 invokes R_1 and C_1.

Activation Sequence (3/6)



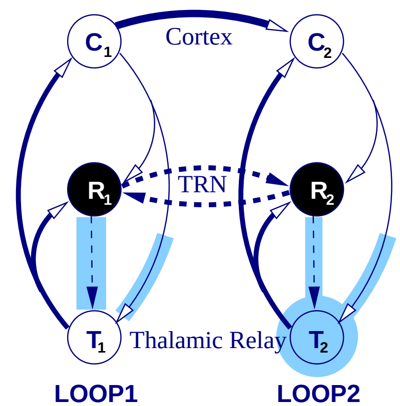
The cortical neuron C_1, through fast connections, invokes another cortical neuron C_2. C_1 also sends out feedback to R_1 and T_1, but these connections are slow. R_1 retains the level of excitation in the meanwhile.

Activation Sequence (4/6)



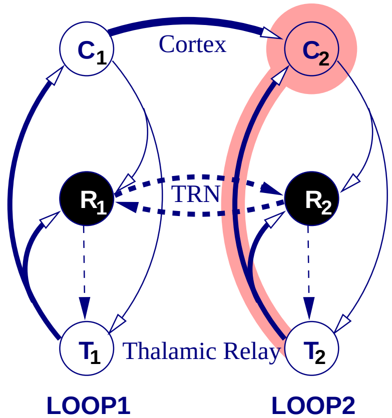
Cortical feedback from both C_1 and C_2 arrives at the TRN, and adds to the existing activity at TRN. Reticular neurons R_1 and R_2 inhibit each other through fast connections.

Activation Sequence (5/6)



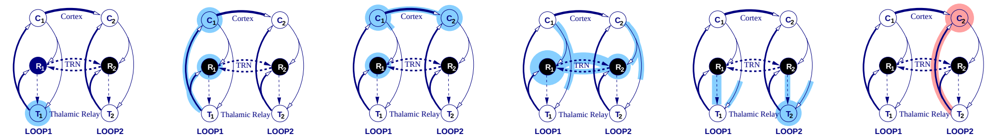
The reticular neurons exert inhibition on the thalamic relays. Feedback from C_1 is canceled out, while that from C_2 is not.

Activation Sequence (6/6)



Finally, only T_2 is allowed to fire again, reactivating C_2 for the second time.

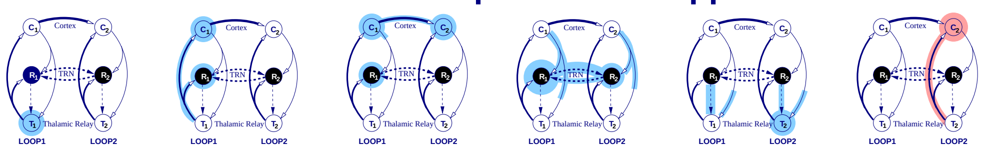
Functional Requirements



(a) (b) (c) (d) (e) (f)

1. TRN neurons must have slow a dynamic (b–d).
2. Inhibition between reticular neurons must be strong (e).
3. Either the cortico-cortical connections must be very fast or the corticothalamic feedback connections must be slow (or both), compared to each other (c–d).
4. Interaction between reticular neurons must be fast (d).

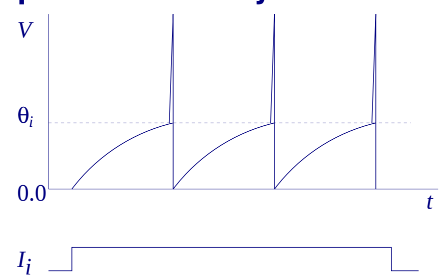
Direct/Indirect Experimental Support



(a) (b) (c) (d) (e) (f)

1. TRN neurons activate and deactivate on a slow timescale (Coulter et al. 1989; Huguenard and McCormick 1992).
2. TRN neurons are harder to depolarize (Huguenard and McCormick 1992): May be due to strong inhibition between TRN neurons.
3. Corticothalamic feedback connections are unmyelinated (i.e., very slow; Tsumoto et al. 1978).
4. Gap junctions found between TRN neurons (Landisman et al. 2002): Interaction may have to be rapid.

Computational Study: Neuron Model

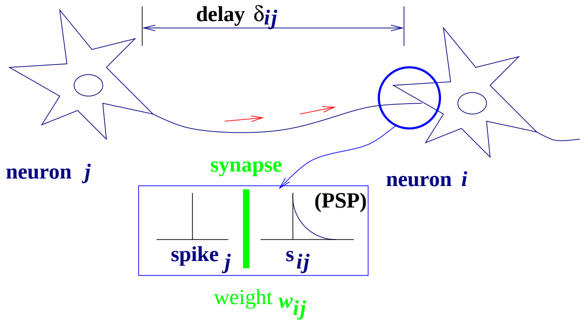


For each neuron i, the membrane potential V_i evolved according to the following dynamic equation:

$$C_i \frac{dV_i}{dt} = I_i(t) - \frac{V_i}{R_i}, \quad (1)$$

where C_i is the membrane capacitance, R_i the resistance, and $I_i(t)$ the input contribution to neuron i at time t. When V_i reaches a threshold value θ_i, a spike is generated and V_i is reset to 0.0.

Synaptic Potential

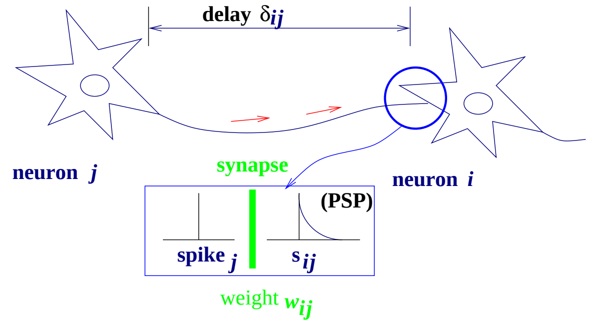


A spike generated by a presynaptic neuron j results in a postsynaptic potential (PSP) s_{ij} at a target neuron i, which is set to 1.0 at the moment the spike is received and is decayed over time as follows:

$$\frac{ds_{ij}}{dt} = -\frac{s_{ij}}{\tau}, \qquad (2)$$

where τ is the time constant of the PSP.

Input Contribution: $I_i(t)$



The input contribution $I_i(t)$ to a neuron i at time t is defined as follows:

$$I_i(t) = \sum_{j \in \mathcal{N}_i} w_{ij} s_{ij}(t - \delta_{ij}), \qquad (3)$$

where $\mathcal{N}_i$ is the set of neurons sending spikes to neuron i.

Model Parameters

Table 1: Neuron Parameters

Parameter	Thal. Relay (T_i)	TRN (R_i)	Cortex (C_i)
Capacitance C_i	0.3	**0.6**	0.3
Resistance R_i	3.0	3.0	3.0
Threshold θ_i	0.25	0.25	0.25
PSP time constant τ_i	0.05	0.05	0.05

Table 2: Connection Parameters

Weight w_{ij}	T_i	R_i	C_i
T_j		1.0	1.0
R_j	**2.0**	**10.0**	
C_j	1.0	1.0	**0.9**

Delay δ_{ij}	T_i	R_i	C_i
T_j		2.0	2.0
R_j	2.0	**0.2**	
C_j	**4.0**	2.0	**0.2**

Overview of Results

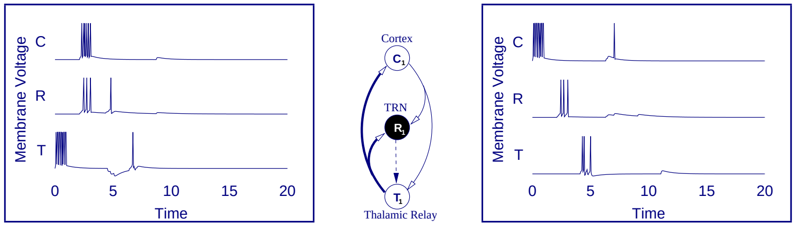
Core results:

- Experiment 1: **Direct stimulation** of thalamus or cortex.
- Experiment 2: Selecting **not input-driven** cortical activity.
- Experiment 3: Selecting **less input-driven** cortical activity.

Predictions under disruptions:

- Experiment 4: When TRN is fast.
- Experiment 5: When R⟶T inhibition is weak.
- Experiment 6: When C⟶C is slow.
- Experiment 7: When R⟶R is slow.

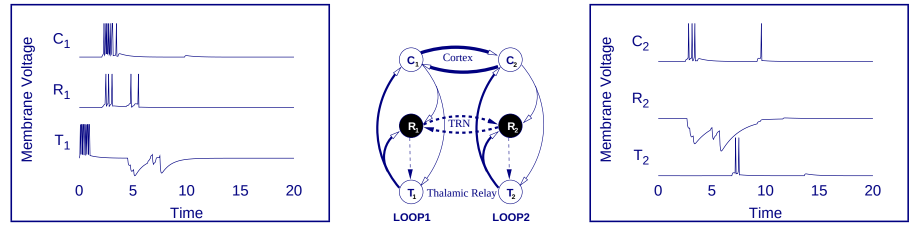
Exp 1: Thalamic vs. Cortical Stim.



Thalamus-driven **Cortex-driven**

- Thalamic stimulation: **No reactivation** of the cortex.
- Cortical stimulation: **Cortical reactivation** through the thalamo-cortical loop.

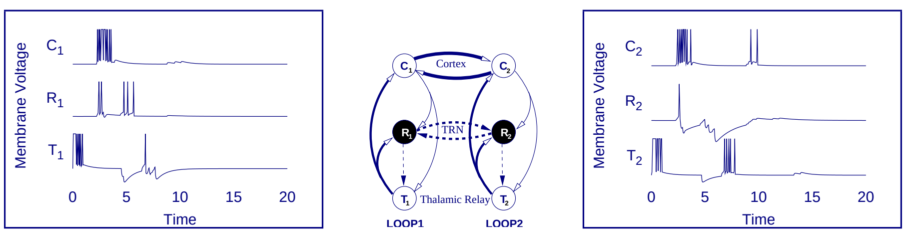
Exp 2: Input vs. No-Input



Loop1: **Input=1.0** Loop2: **Input=0.0**

- **Input-driven** cortical activity **does not reactivate.**
- **Cortically induced** cortical activity **reactivates** through the cortex-thalamus-cortex loop.

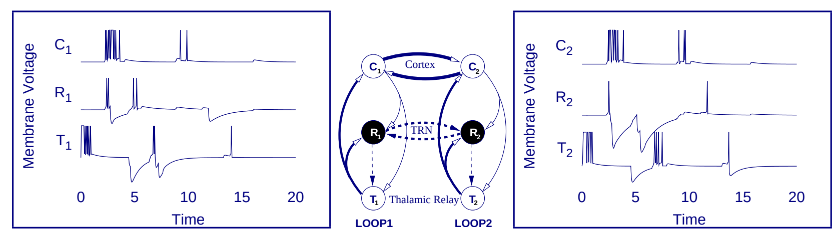
Exp 3: Strong vs. Weak Input



Loop1: **Input=2.0** Loop2: **Input=1.0**

- **Strongly input-driven** cortical activity **does not reactivate**.
- **Weakly input-driven** cortical activity **reactivates** through the cortex-thalamus-cortex loop.

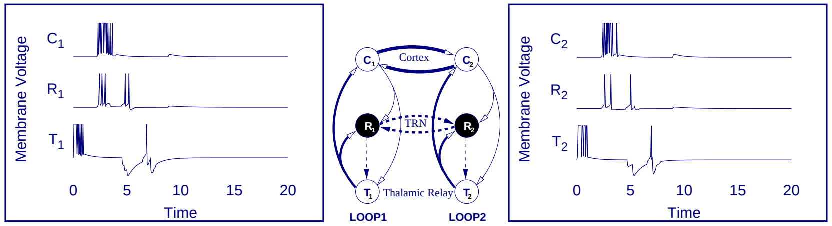
Exp 4: Fast TRN dynamics



Loop1: Input=2.0 Loop2: Input=1.0

- With faster TRN dynamics ($C_i = 0.5$), the reticular neurons fail to integrate the thalamic and cortical contributions, and thus timely inhibition is interrupted.

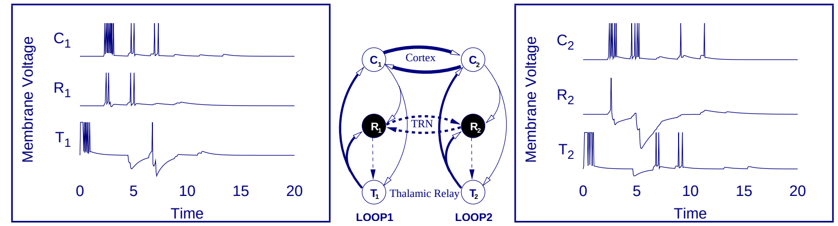
Exp 5: Weak TRN to Thalamus Inhibition



Loop1: Input=2.0 Loop2: Input=1.0

- With lowered R⟶T weight (2.0), due to the weaker disinhibition effect, loop2 reticular neuron generates more activity to suppress the thalamic relay. As a result, loop2 fails to reactivate the cortex.

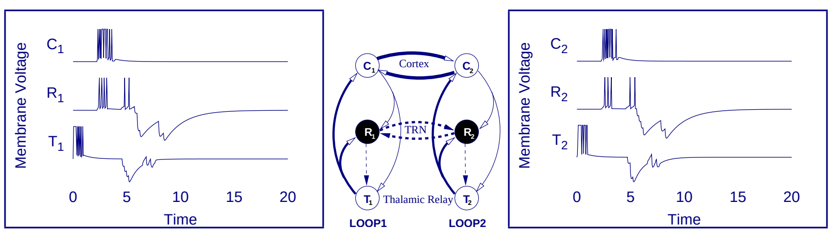
Exp 6: Slow Corticocortical Connections



Loop1: Input=2.0 Loop2: Input=1.0

- With longer C⟶C connection delay, the phases of loop1 and loop2 activities start to drift and become irregular.

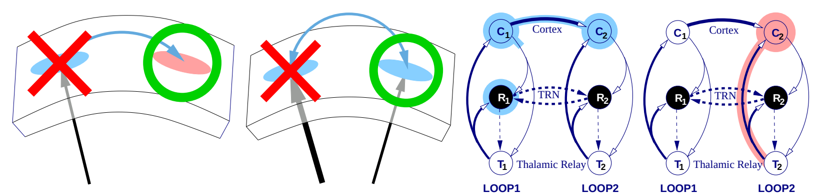
Exp 7: Slow intra-TRN connections



Loop1: Input=2.0 Loop2: Input=1.0

- With longer R⟶R connection delay (1.5), the disinhibition effect did not happen in time to allow loop2 to reactivate the cortex.

Summary of Results



- A thalamocortical model was implemented with parameters derived from functional, anatomical, and physiological considerations.
- The model was successful in detecting and promoting (1) **non-input-driven**, and (2) **less input-driven** cortical activity.

Discussion

- How particular **answers are generated from the quesions**?
  - Analogy, inference, association, etc.
- Why need such a round-about? **Why not do it in the cortex**?
- What about primitive animals **without the thalamus**?

Discussion (cont'd)

The model does not account for the following:

- Drivers vs. modulators innervating thalamic relays.
- Slowness of TRN is in I_T.
- Low-threshold firing in thalamic relay and TRN (burst, as opposed to tonic firing).
- Role of the interneurons in dorsal thalamic nuclei.
- Other inputs to TRN and dorsal thalamus (parabrachial region, brain stem, etc.).
- Higher-order relays: feedback is from layer V, not layer VI.
- Intricate circuitry in the cortex (layers IV, II/III, etc.).

Predictions

- Results from as Exp 1 to Exp 3 would be replicable in *in vivo* experiments.
- Not just I_T but other currents in TRN may turn out to have a slow dynamic.
- Intra-TRN connectivity will reflect that of its cortical counterpart (majorly in its extent, but maybe also in its broader pattern).
- The time-course of a unit of computation T_u in the cortex would follow:

$$T_u = \underbrace{T \rightarrow C}_{\text{feedforward}} + \underbrace{C \rightarrow C}_{\text{computation}} + \underbrace{C \rightarrow T}_{\text{feedback}} + \underbrace{T \rightarrow C}_{\text{reactivation}} .$$

Conclusion

- **Input–output binding problem** (IOBP) may need more attention.
- The **thalamo-cortical loop** may be able to solve the IOBP.
- It may be important to look at how pieces of circuit properties **fall into place in the puzzle**.

Acknowledgments

- Insightful comments by James A. Bednar, Bruce H. McCormick, Risto Miikkulainen, Lokendra Shastri, S. Murray Sherman, Ray Guillery, Jason Traesger, Dennis Glanzman, and Yingwei Yu helped clarify several points.
- The simulation results presented here are based on earlier results in (Choe 2002, 2004). All simulations were implemented in *XPPAUT* by Ermentrout (2002).
- This research was supported in part by Texas A&M University, by the Texas Higher Education Coordinating Board grant ATP#000512-0217-2001, and by the National Institute of Mental Health Human Brain Project grant #1R01-MH66991.

References

Choe, Y. (2002). Second order isomorphism: A reinterpretation and its implications in brain and cognitive sciences. In Gray, W. D., and Schunn, C. D., editors, *Proceedings of the 24th Annual Conference of the Cognitive Science Society*, 190–195. Erlbaum.

Choe, Y. (2004). The role of temporal parameters in a thalamocortical model of analogy. *IEEE Transactions on Neural Networks*, 15:1071–1082.

Coulter, D. A., Huguenard, J. R., and Prince, D. A. (1989). Calcium currents in rat thalamocortical relay neurons: Kinetic properties of the transient, low-threshold current. *Journal of Physiology*, 414:587–604.

Crabtree, J. W., and Isaac, J. T. (2002). Intrathalamic pathways allowing modality-related and cross-modality switching in the dorsal thalamus. *The Journal of Neuroscience*, 22:8754–8761.

Crick, F. (1984). Function of the thalamic reticular complex: The searchlight hypothesis. *Proceedings of the National Academy of Sciences, USA*, 81:4586–4950.

Destexhe, A., and Sejnowski, T. J. (2001). *Thalamocortical Assemblies: How Ion Channels, Single Neurons, and Large-Scale Networks Organize Sleep Oscillations.* Oxford, UK: Oxford University Press.

Eckhorn, R., Bauer, R., Jordan, W., Kruse, M., Munk, W., and Reitboeck, H. J. (1988). Coherent oscillations: A mechanism of feature linking in the visual cortex? *Biological Cybernetics*, 60:121–130.

Ermentrout, B. (2002). *Simulating, Analyzing, and Animating Dynamical Systems: A Guide to XPPAUT for Researchers and Students.* Philadelphia, PA: SIAM.

Gray, C. M., Konig, P., Engel, A., and Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 338:334–337.

Guillery, R. W., and Sherman, S. M. (2002). Thalamic relay functions and their role in corticocortical communication: Generalizations from the visual system. *Neuron*, 33:163–175.

Harth, E., Unnikrishnan, K. P., and Pandaya, A. S. (1987). The inversion of sensory processing by feedback pathways: A model of visual cognitive functions. *Science*, 237:184–187.

Huguenard, J. R., and McCormick, D. A. (1992). A novel t-type current underlies prolonged ca^{2+}-dependent burst firing in GABAergic neurons of rat thalamic reticular nucleus. *Journal of Neuroscience*, 12:3804–3817.

LaBerge, D. (1995). *Attentional Processing.* Cambridge, MA: Harvard University Press.

Landisman, C. E., Long, M. A., Beierlein, M., Deans, M. R., Paul, D. L., and Connors, B. W. (2002). Electrical synapses in the thalamic reticular nucleus. *Journal of Neuroscience*, 22:1002–1009.

Llinás, R. R., and Ribary, U. (1994). Perception as an oneiric-like state modulated by the senses. In Koch, C., and Davis, J. L., editors, *Large Scale Neuronal Theories of the Brain*, chapter 6, 111–124. Cambridge, MA: MIT Press.

McCormick, D. A., and Bal, T. (1997). Sleep and arousal: Thalamocortical mechanisms. *Annual Review of Neuroscience*, 20:185–215.

Mumford, D. (1995). Thalamus. In Arbib, M. A., editor, *The Handbook of Brain Theory and Neural Networks*, 153–157. Cambridge, MA: MIT Press.

Newman, J., Baars, B. J., and Cho, S.-B. (1997). A neural global workspace model for conscious attention. *Neural Networks*, 10:1195–1206.

Sherman, S. M., and Guillery, R. W. (2001). *Exploring the Thalamus.* San Diego, CA: Academic Press.

Sillito, A. M., Jones, H. E., Gerstein, G. L., and West, D. C. (1994). Feature-linked synchronization of thalamic relay cell firing induced by feedback from the visual cortex. *Nature*, 369:479–482.

Steriade, M., and McCormick, D. A. (1993). Thalamocortical oscillations in the sleeping and aroused brain. *Science*, 262:679–685.

Taylor, J. G. (1998). *The Emergent Mind.* Cambridge, MA: MIT Press.

Tsumoto, T., Creutzfeldt, O. D., and Legendy, C. R. (1978). Functional organization of the corticofugal system from visual cortex to lateral geniculate nucleus in the cat. *Experimental Brain Research*, 32:345–364.

von der Malsburg, C. (1986). Am i thinking assemblies? In Palm, G., and Aertsen, A., editors, *Brain Theory: Proceedings of the First Trieste Meeting on Brain Theory*. Berlin; New York: Springer.