Global Organization from Local Signals in Neural and Artificial Networks:

Summary of Project Goals & Results (Final Report; Proj # 322033, acronym ‘LocalToGlobal’)

This project combined experimental work and theoretical analysis to investigate the computational and neural principles underlying perceptual and cognitive processing, specifically the vital brain process of *perceptual organization*. Individual neurons are biophysically complex, but computationally they are rather simple: they receive input from other neurons, and generate action potentials (or not) based on the magnitude of collated inputs. A major challenge of computational neuroscience is therefore to understand how networks of many such rather simple elements, neurons, operate in ways that give rise to complex behaviours. A prototypical example of this challenge is to understand how the brain transforms the *local* signals it receives from sensory neurons – each being highly restricted in space and time – into the *global* representations of objects and events that we experience perceptually. Performing local-to-global transformations is highly non-trivial because there is a host of ambiguities inherent to each local signal and those lead, in turn, to there being multitudes of different global scenarios which could have, in principle, given rise to the very same sensed stimulation. To make this more concrete, we focus on the sense of vision, and more specifically on visual motion perception. The local nature of early neural visual information starts in the retina – individual photoreceptors are tiny – and persists well into visual cortex. This is a good thing – it gives us visual acuity – but it leaves the brain with the task of reconstructing objects’ global motion from local motion cues. Consider a spinning wheel: locally, each point on the wheel generates a motion signal of a different direction and speed. The perception of a unitary, rigidly-moving object thus requires a process of grouping the disparate local signals, along with finding the underlying global motion (a 3D vector, in this case) that gave rise to them. Next, consider that the wheel can be in the vicinity of another one, which may be spinning in a different direction (eg, two cars crossing in traffic). Here, the grouping must be applied separately to the subsets of local motion signals that arise from the separate objects, namely it needs prior (or concurrent) motion segmentation. But this is difficult without any prior knowledge on which signals belong together – ie, on grouping. And while machine-vision research is still struggling to overcome this “chicken and egg” problem, psychophysical experiments show that humans typically perceive the correct segmentation-and-grouping from a “mess” of local motion signals even when there are no other cues for segmentation such as luminance, color or shape.

How does the brain achieve this computational feat, overcoming the ambiguities inherent in local signals to arrive at a stable, coherent interpretation of the underlying global scene? We addressed the question by utilizing a class of stimuli that produce something of an exception to the rule. When a stimulus has two distinct, ecologically valid interpretations, it can give rise to *perceptual bi-stability*: intermittent, irregular alternations between periods of perceiving one or the other interpretation. Previous work in our group made use of our expertise in motion perception to study the dynamics of bi-stable perceptual alternations during prolonged viewing of deeply ambiguous global motion displays. We found systematic commonalities with the dynamics of binocular rivalry, indicating that neural competition plays a central role in global motion perception. This suggested that neural competition may be far more ubiquitous in the brain than previously thought, in turn raising the possibility that it may have functional significance also for the processing of stimuli that are (perceptually) unambiguous. But what this function may be was left an open question. The important contribution of the present project is that it offers a novel, principled rationale for widespread use of competitive neural architecture in the brain, as well as a concrete network mechanism to implement it.

The experimental foundation of our results is the discovery of a hitherto unknown, quantitative characteristic of the dynamics of alternations, which is common to all bi-stable phenomena. We found that the fraction of time observers spend in one of the two bi-stable percepts predicts the mean durations of bi-stable epochs for *both* competing percepts. The functional relation is given by $\mathrm{T_A}(f) = \mathrm{T_{eq}}\sqrt{\frac{f}{(1-f)}}$ and $\mathrm{T_B}(f) = \mathrm{T_{eq}}\sqrt{\frac{(1-f)}{f}}$ where $\mathrm{T_A}$ and $\mathrm{T_B}$ denote the mean durations of competing percepts A and B, f denotes the fraction of time spent in A (ie, $f \equiv f_A$ and thus $f_B = 1-f$), and T_{eq} denotes the mean duration at $f=0.5$ (the 'equi-dominance' point, where $T_A = T_B \equiv T_{eq}$).

Note that the only *a priori* constraint is that $T_A/[T_A+T_B]$ must equal f (by definition) for any $f \in [0,1]$, meaning that it would be possible (arithmetically) for T_A and T_B to vary independently of each other. In the field of binocular rivalry, the influential work of Levelt (1968) has led to a long-held assumption that independent control of the mean dominance duration of each eye is possible also *experimentally*, by separately manipulating the 'stimulation strength' of each monocular image (eg, their contrast). But our research shows this has been a misimpression: a systematic re-analysis of all available (published) data on dominance durations in binocular rivalry, from numerous studies of over a dozen different research groups spanning nearly sixy years, showed a striking quantitative fit with the curve predicted by the function given above.

These results have far-reaching implications for our understanding of the mechanism(s) underlying perceptual bi-stability, and potentially also for how the brain resolves sensory ambiguities more generally. They cast doubt on the notion that the system represents, internally, the strength of each percept in manner independent from those of the competing percepts (with the putative 'absolute strengths' reflected in the measures of 'absolute time', T_A and T_B). Rather, they suggest that it is the *relative strength* of the competing percepts which is represented by the underlying system (and reflected in the measure of relative time, f).

This hypothesis provides an important link to a host of perceptual and cognitive phenomena that, thus far, have been studied with very different approach and tools, namely this of *Bayesian* theories. This is a powerful approach that has much to offer in terms of understanding the relation between the statistics of the outside environment and the behaviour of organisms within that environment – in other words, if offers deep insights as to *why* organisms behave as they do. But at present it is limited with regard to the question *how* these behaviours come about, ie in terms of offering mechanistic (rather than descriptive) models for the underlying system.

In the final, theoretical part of this project we have shown that probabilistic neural representations emerge naturally in network architectures that implement mutual exclusivity among competing attractor states via shared inhibitory feedback. A central issue here is to explain the precise noise mechanisms that give rise to the variability observed in behaviour. To mimic experimental findings, previous models have posited shared noise sources, either from the external stimulation or due to `background' stochastic input. But we show that variability like that observed at the behavioural level ('global') can emerge from spiking noise that is generated independently in each neuron (`local'). This is a well-established source of neural noise, naturally arising from synaptic processes such as ion channel gating and quantal release (in contrast to the putative `background' or stimulus noise which entail ad-hoc assumptions). Sampling behaviour arises from global state transitions between the different attractor patterns embedded in the system. The identity of the active attractor is not fixed, nor is the activity determined solely by which gets the stronger stimulation. Instead, the competing populations alternate over time (or trials), each getting to ``win'' a fraction determined by the relative strength of its stimulation.

With just three free parameters, the model reproduces and explains a wide range of experimental findings in bi-stable perception, including two longstanding observations that have been, thus far, difficult to reconcile. Importantly, the model also provides a theoretical framework to understand choice behaviour more generally. It shows that probabilistic sampling, while being a robust phenomenon, breaks down when external stimulation is so strong as to overcome the inhibitory feedback and drive the winning attractor to saturation activity. This offers a novel, principled rationale for the prevalent finding of balanced excitation and inhibition in cortical networks. Moreover, this insight could have practical impact on clinical work with patients, by shedding new light on links previously observed between `sensory overload' symptoms and inflexible, obsessive behaviours in certain neurological disorders.