Title
How does the sensory-motor brain integrate and give rise to cognition and learning?

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors
Schöner, Gregor
Trifonova, Iliyana
Spencer, John
et al.

Publication Date
2024

Peer reviewed

How does the sensory-motor brain integrate and give rise to cognition and learning?

Gregor Schöner (gregor.schoener@ini.rub.de)
Institute for Neural Computation, Faculty of Computer Science, Ruhr-University Bochum, Germany

Keywords: speech; word learning; language comprehension; associative recall, neural cognitive architectures; neural dynamics

The brain evolved as a sensory-motor machine that drives behavior while being linked to the world through sensors. Human cognition abstracts from these sensory-motor roots but retains intimate ties. The brain's structure reflects this history. How do neural processes at different distances from the sensory and motor surfaces integrate to achieve meaningful and grounded cognition? This is a challenge given the time-continuous and graded nature of sensory-motor processing, which enables continuous online updating. It is also a major challenge to understanding development and autonomous learning, in which the coupling across functional boundaries evolves under the influence of online activation patterns.

The symposium discusses neurally grounded perspectives on how integration across processes may emerge around a set of case studies. Talk 1 models word learning driven by time-continuous speech signals; Talk 2 links the continuous dimensions of speech to sentence meaning comprehension; Talk 3 accounts for the interaction between language label learning and executive control from a developmental perspective. These three talks use Dynamic Field Theory (DFT) as the framework for understanding integration. Talk 4 addresses associative recognition from a different theoretical perspective that uses spiking neural networks to cover the processing stream from visual feature extraction to the behavioral decision and predicts source-localized MEG and fMRI data.

Presenters

Gregor Schöner (moderator) holds the Chair of Theory of Cognitive Systems and is Director of the Institut for Neural Computation at RUB, Germany. His interdisciplinary research aims to understand how perception, motor behavior, embodied and higher cognition may emerge from integrated neural dynamic architectures framed within DFT.

Talk 1: Iliyana Trifonova is a Postdoctoral Research Associate in John Spencer's lab pursuing research in visual and spoken word recognition and statistical learning. **John Spencer** is a Professor of Psychology at the University of East Anglia in Norwich, UK. His research examines the development of word learning, working memory, attention, and executive function using brain imaging and DFT.

Talk 2: Maria Piñango is a Professor of Linguistics and Director of the Language and Brain Lab, **Jason Shaw** is Associate Professor of Linguistics and Director of the Phonetics Lab, and **Michael Stern** is a PhD candidate in linguistics, all three at Yale University. Their collaboration focuses on modeling language in terms of neural dynamics coupling sensory-motor and conceptual systems defined on continuous dimensions.

Talk 3: Aaron Buss is an Associate Professor of Psychology at the University of Tennessee. His research aims at an integrated understanding of the behavioral and neural dynamics of executive function over development. **Alexis McCraw** is a doctoral student in his lab. **Larissa Samuelson** is a Professor of Psychology at the University of East Anglia, UK, whose research examines processes of cognitive development with a focus on early word and category learning.

Talk 4: Jelmer Borst is an Associate professor of Computational Cognitive Neuroscience at the Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence at the University of Groningen. His research develops computational cognitive models at different levels of abstraction, from higher level process models in the cognitive architecture ACT-R to low-level spiking-neuron models.

Talks

Talk 1: A neural dynamic model of auditory word learning and recognition

Natural speech is usually continuous without silent gaps between words, which poses a challenge to word learners. Successful word learning requires that continuous auditory signals be segmented in real time into its constituent word units. Infants use a variety of cues from the speech input to achieve word segmentation and learning, including the probability with which adjacent syllables co-occur. How do infants integrate the in-coming sensory stream with higher-level processes to yield auditory word representations? We present a novel neural dynamic model framed within DFT that is sensitive to the distributional patterns important for speech segmentation and word learning. Real-time speech input was created by converting auditory waveform inputs into a time-dependent sequence of discrete phoneme categories using Wav2Vec. The input feeds

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society*.

into a series of dynamic neural fields that encode the presence of phonemes via peaks of activation along a phoneme dimension (a discrete dimension with 44 categories). The model tracks both the current phoneme and the previous phoneme, forming a successor map where peaks represent the current-previous sequence. Associations are then formed between this sequence representation and the activation of word units via Hebbian learning. These associations comprise the representations of the auditory words.

Simulations show that the model continuously and autonomously builds such word representations based on the statistical regularities of the input. In time, as patterns of phoneme sequences repeat, words are activated by a "bottom-up" match between the current input pattern and the association matrix. This drives a mechanism of lexical competition between consistent word units and subsequently leads to a recognition process. Thus, the model effectively integrates phoneme classification, word learning, and word recognition processes in a real-time dynamical system.

Talk 2: Neural dynamic architecture of language linking the sensory-motor to the conceptual domain via lexical items

Language can be characterized as a bidirectional mapping between sensory-motor and conceptual systems, both of which decompose to continuous dimensions. We describe two DFT models of language: one coupling lexical items to sensory-motor dimensions, and one coupling lexical items to conceptual dimensions. The sensory-motor model captures effects of lexical competitors on speech articulation, including both "trace effects" in speech errors and "contrastive hyperarticulation" in non-errorful speech. The model features a novel proposal for dynamically updating the polarity of input (excitatory vs. inhibitory) from lexical items to motor planning fields. The conceptual model captures effects of preceding linguistic context on lexico-semantic interpretation of the English word "have". This model features a novel proposal for between-field coupling that derives dependencies between conceptual dimensions. Using a common framework across domains allows mechanisms established in one domain to generate hypotheses in the other. In our case, dynamically-defined coupling parameters, motivated in the sensory-motor domain, provide new hypotheses about lexical interaction effects on conceptual interpretation; likewise, between-field coupling mechanisms from the conceptual domain provide hypotheses about phonotactics in the sensory-motor domain. More broadly, expressing the sensory-motor and conceptual systems underlying language in a common framework represents an important step towards a unified model.

Talk 3: Using a dynamic field model to ground executive function in dimensional label learning

Recent longitudinal neuroimaging data from 2 to 5 years of age show that neural mechanisms involved in learning labels such as red and color support the development of executive function (EF) skills. In this project, we used a recently published neurocomputational model that explains behavior in cross-situational word learning tasks to examine the grounding of EF in dimensional label learning. The model was given tasks designed to assess comprehension and production of labels for features and dimensions of color and shape. In the model, labels are learned via associations between labels and visual features (e.g., red with the red hue) and associations between labels and other labels (e.g., color with red). As these associations are strengthened, the model improves on measures of comprehension and production. We also assessed the ability of the model to attend to these visual dimensions in the context of a canonical measure of executive function, the dimensional change card sort task. In this task, children are instructed to first sort bivalent cards by either shape or color and then switch and sort by the other dimension. Typically, the majority of 3-year-olds fail to switch rules, but the majority of 4-year-olds have little difficulty. Model simulations showed a similar transition as dimensional label associations were strengthened. The model predominantly failed with weak associations and switching ability improved with stronger associations. These results show that high-level cognitive processes (rule-following and EF) can be grounded in perceptual associations between labels and visual features.

Talk 4: Connecting Cognition to the Brain: a Spiking-Neuron model of Associative Recognition

Most cognitive and mathematical models describe the human mind at a high level: for example, cognitive architectures like ACT-R use concepts like memory retrievals and chunks, evidence accumulation models focus on the decision process, and Bayesian models try to understand why the human mind functions as it does. However, if we want to understand how cognition emerges from the human brain, we need to develop cognitive brain models. In recent years, several neural models have been proposed. Often, these models are tremendously detailed (e.g., Markrams Blue Brain Project), but unfortunately fail to make the connection to actual behavior, or only focus on qualitative patterns. In addition, even if neural models are linked to functional behavior (e.g., Eliasmiths Spaun), such models are necessarily so complex that they are very hard to evaluate. To overcome these problems, we developed an integrative biologically-plausible brain model of associative recognition that connects the time-continuous nature of the brain to discrete cognitive decisions. To evaluate the resulting model, we compared its output to human behavior quantitively, and used its neural spikes to predict source-localized MEG and fMRI data. The results matched source-localized MEG data in occipital, temporal, prefrontal, and precentral brain regions; as well as a classic fMRI effect in prefrontal cortex. Integrating across different levels of description from neurons to cognition and from spikes to MEG/fMRI data resulted in a model that can explain how high-level cognition might be grounded in the spiking of neurons.