Sequence generation

Gregor Schöner
Institute for Neural Computation (INI)
Ruhr-University Bochum, Germany
dynamicfieldtheory.org
gregor.schoener@ini.rub.de
Sequences

- All behavior and thinking consist of sequences of physical or mental acts.
- Sometimes in a fixed order as in action routines, or highly trained action patterns.
- But potentially highly flexible ... as in language, thinking, problem solving ...
Probes of sequence generation

- serial order in memory
- Lashley: serial order as separate from other aspects of material
- implicit sequence learning
- sequential actions: timing
DFT challenge for sequences

- DFT postulates that all neural states underlying behavior/mental process are attractors that resist change…

- but generating sequences of such states require change of state! Conflicting constraints!

- answer: instabilities are induced systematically to enable switching to a next/new attractor
Sequence generation

- an illustrative example
- the neural/mathematical mechanism
Sequence of physical acts

- **task**: search for objects of a given color in a given order
  - 1 blue
  - 2 red
  - green

- stably couple to objects once they are detected

- ignore objects when their turn has not yet come (distractors)
Implementation as an imitation task

- learn a serially ordered sequence from a single demonstration
  yellow-red-green-blue-red

- perform the serially ordered sequence with new timing
  yellow-red-green-blue-red

[Sandamirskaya, Schöner: *Neural Networks* 23:1163 (2010)]
red a distractor    red a target

[Sandanirskaya, Schöner: Neural Networks 23:1163 (2010)]
Condition of Satisfaction (CoS)

The architecture for a sequential color-search task on a Khepera robot involves an active node of the ordinal dynamics projecting its activation onto an intention field, defined over the color dimension. The intention field is coupled to the space-color field, which also receives visual input from the robot's camera. An activation peak in the space-color field drives the navigation dynamics of the robot, setting an attractor for its heading direction. The condition-of-satisfaction field is also defined over the color dimension and is activated when the object of the currently active color takes up a large portion of the camera image.

Before an object that matches the current intention has been found, no peak exists in the space-color field. The heading direction does not receive input at that time from the space-color field and the vehicle's navigation dynamics is dominated by obstacle avoidance, which is implemented using a standard dynamic method (Bicho, Mallet, & Schöner, 2000). This results in the roaming behavior that helps the robot search for objects of the appropriate color.

During teaching, the visual input from the object shown to the robot is boosted enough to induce a peak in the space-color field. This peak projects activation backwards onto the intention field, where a peak is induced at the location that...

[Sandamirskaya, Schöner: Neural Networks 23:1163 (2010)]
Visual search

- 2D visual input color vs. horizontal space
- intensity of input from a color histogram within each horizontal location
current color searched provides ridge input into a color-space field
Perception for navigation

Perception for CoS

Color-space DF

Intention DF

CoS DF

Camera image
ordinal stack

condition of satisfaction (CoS)

intentional state

2D color-space field
Figure 11: One run of the robotic demonstrations. A: Time-courses of activation of five ordinal nodes during sequence learning and production. B: Time-course of activation in the action field. Positive activation in the field encodes the color currently searched for. C: Time-course of activation in the condition of satisfaction field. Arrows mark the times when condition of satisfaction signals were emitted. D: The projection of the perceptual color-space field onto the spatial dimension -horizontal axis of the image plane-. The arrows mark times when the object of interest in each ordinal position first appeared in the visual array of the robot. The "random search" behavior changed to "approach target" behavior at these points.
Mathematical mechanism

intention

condition of satisfaction

neural state

motor-world-sensor state

dimension x

dimension y

prediction
Sequence of instabilities

- the CoS is pre-shaped by the intention field, but is in the sub-threshold state
- until a matching input pushes the CoS field through the detection instability
- the CoS field inhibits the intention field that goes through a reverse detection instability
- the removal of input from the intention to the CoS field induce a reverse detection instability
- both fields are sub-threshold
one could think of the “prediction” implied in the CoS as being a form of efference copy that does act inhibitorily…

but it does so on the (motor)intention, not on the perception of the outcome that is predicted!
Generalization

- **match-detection** => *CoS*
- **mis-match (or change) detection** => *CoD*  
  (condition of dissatisfaction)

**Figure 2:** The feature matching sub-network. See the text for an explanation.

Mismatch within a single feature dimension is sufficient to activate the **condition of dissatisfaction (CoD)**. In contrast, the **condition of satisfaction (CoS)** node is only activated if all attended features match the search cue. Together with the intention node, these two nodes are used to autonomously generate sequences of neural processing steps (Sandamirskaya & Schöner, 2010).

**Figure 3:** An overview of the neural dynamic process model.

Boxes represent sub-networks of fields and arrows their couplings. Green outlines highlight sub-networks changed with respect to the previous model.

Feed-forward feature maps and salience map

The bottom-up pathway of the model (and of human perception) is a parallel preattentive process purely driven by input. In the model, visual input may come from a live camera image (A) or, in the current case, from randomly generated search displays (A1) (Figure 4).

**Figure 4:** The bottom-up pathway of the model. See text for explanation. Green outlines highlight sub-networks changed with respect to the previous model.

Three features are extracted in parallel: **color**, **orientation**, and **shape**. Color is extracted from hue-space. Orientation is obtained by filtering the thresholded saturation with four elongated center-surround filters. To align with the experiments of Nordfang and Wolfe (2014), we swapped the size feature of our previous model (Grieben et al., 2020) to shape. Shape was obtained by template matching (normalized cross-correlation), a simplified account for preattentive recognition.

[Grieben, Schöner, CogSci 2021]
How is the next state selected?

Once the current state has been de-activated...

Three notions:

- Gradient-based selection
- Chaining
- Positional representation

An illustration
How is the next state selected?

- Once the current state has been deactivated…

- 3 notions (~Henson Burgess 1997)
  1. Gradient-based selection
  2. Chaining
  3. Positional representation
Gradient-based

- a field/set of nodes is released from inhibition once the current state is deactivated…
- a new peak/node wins the selective competition based on inputs…
  - e.g. salience map for visual search
  - e.g. overlapping input from multiple fields..
- return to previous states avoided by inhibition of return

[Grieben, Schöner, CogSci 2021]
Gradient-based

- this is used in many of the DFT architectures
  - visual search
  - relational grounding
  - mental mapping

[Grieben, Schöner, CogSci 2021]
Chaining

- for fixed sequences...
  - e.g. reach-grasp
  - fixed order of mental operations... e.g. ground reference object first, then target object

- less flexible (e.g., when going through the same state with different futures)

- could be thought to emerge with practice/habit from the positional system
Positional representation

- A neural representation of ordinal position is organized to be sequentially activated...
- The contents at each ordinal position is determined by neural projections from each ordinal node...

---

Ordinal nodes

Action field

CoS field

motor system

action perception

CoS perception

environment

[Sandamirskaya, Schöner: Neural Networks 23:1163 (2010)]
Positional representation

- essentially chaining with flexible contents
- good for fast learning of sequences...
  - e.g. imitation
  - a Hippocampus function?
- but: must have potential synaptic links to many representations...
- => such ordinal systems must exist for sub-representations... embodiment effects...
Tekülve et al., Frontiers in Neurorobotics (2019)

3. MODEL

The neural dynamic architecture described here is a network of neural fields that are coupled to a camera and a robotic arm. These links enable online connection to a changing visual scene and online control of the arm. Three sub-networks (Figure 2) autonomously organize sequences of activations to build visual representations, learn or perform serially or ordered sequences, and generate object-oriented movements.

The perceptual sub-network, connected to the camera, creates a working memory representation of the visual scene through autonomous shifts of attention. A motor sub-network drives an oscillator generating velocity commands for the robotic arm. The cognitive sub-network represents ordinal positions in a sequence and may autonomously shift from one ordinal position to the next. The ordinal system may be used in two different manners, sequence learning and sequence recall, controlled by the activation of one of two different task nodes. These task nodes activate behaviors by boosting fields' resting levels and enabling fields to generate task relevant attractor states.

The following sections describe for each sub-network the states that drive behavior and the mechanism for how the system switches between those states. The last section addresses the integration of all three sub-networks for the two tasks Learn and Recall.

3.1. Perception: Scene Representation

The scene representation sub-network is based on Grieben et al. (2018) and creates three-dimensional (2D space and 1D color) working memory representations of objects in the visual scene captured by the camera. Each entry into the representation is created sequentially as the sub-network autonomously shifts attention across different objects in the scene.

[Tekülve et al., Frontiers in Neurorobotics (2019)]

Serial order demonstrated/enacted
At point $t_0$, the Exploration intention node provides a homogeneous boost to the Saliency Selection field leading to an activation peak at the location of the purple object. This causes the emergence of a three-dimensional peak in the Scene Selection field, of which the color dimension is shown in the third row. The Working Memory field contains no supra-threshold activation yet, but at the locations of the non-background objects, the resting level is increased across the whole color dimension.

Once the peak in the Scene Selection field has fully emerged at $t_1$, its color component is forwarded as a slice toward the Working Memory, where it overlaps with the tube originating from the Saliency Selection field and forms a three-dimensional peak. Subsequently a peak also forms in the Memory Spatial Selection field, which shares the same color as the peak in the Scene Space Selection causing an overlap in the Color Match field.

The peak forming in the Color Match field activates the CoS Explore node, which inhibits the Explore intention node. Thus the resting level boost is removed from the Saliency Selection field, which subsequently falls down to sub-threshold activation at point $t_2$. Only the self-sustained peak in the Working Memory field remains.

The absence of a peak in the Color Match field causes the CoS node to fall below threshold again, bringing the sub-network to its initial state. The following activation of the Explore intention node, depicted from $t_3$ until $t_5$, follows the same temporal activation pattern as the previous one with different feature values for spatial location and color. The spatial location in the Saliency Selection field differs due to the inhibitory influence from the Working Memory field. See Supplementary Video 3 for another example of autonomous build-up of visual working memory in continuous time.

4.2. Learning Demonstration

A particular color sequence is taught to the network in its learning regime by presenting objects of a certain color one after another. In Figure 5 activation snapshots of some points in time during an exemplary learning episode are shown. The top row depicts the temporal evolution of activation of the ordinal numbers.
Time course of attention selection and building of scene memory

The build-up of the scene working memory is an ongoing process that provides visual information to the network irrespective of the currently active task node. In Figure 4 we show activation snapshots of different points in time during working memory build-up in an exemplary scene containing three objects and the arm's end-effector.

FIGURE 4 | Time course of building a scene memory.
FIGURE 6 | Time course of recalling a three element sequence through pointing at colored objects.
FIGURE 7 | Online updating of the movement during sequence recall.

To representations of higher cognition (serial order) and to the motor system (pointing). The network architecture enables a robotic agent to autonomously learn a sequence of colors from demonstration and then to act according to the defined serial order on a scene. Both during learning and while acting out the sequence, the transitions between elements of the sequence are detected without the need for an external control signal (The switch between learning and recall mode is not autonomous, however, reflecting a similar need for task instructions when a human operator performs such a task).

In each of the three sub-networks responsible for scene representation, the representation of serial order, and movement generation, sequential transitions between neural activation states are brought about through the mechanism of the condition of satisfaction. Thus, visual attention shifts only once a currently attended item has been committed to working memory. A transition to the next element in the serial order occurs only once the robot has successfully acted on the current element. And an arm movement terminates only once the desired movement target has been reached. The mechanism of the condition of satisfaction thus reconciles the capacity to autonomously act according to learned or structurally determined plans with the capacity to be responsive to sensory or internal information about the achievement of goals.

5.1. What the Scenario Stands for

The scenario was simple, but meant to demonstrate the fundamental components of any neurally grounded autonomous robot. (1) A representation of the visual surround is the basis for any intelligent action directed at the world. It is also the basis for sharing an environment with a human user. We humans are particularly tuned to building scene representations which form the basis of much of our visual cognition (Henderson and Hollingworth, 1999). Scene representations need to...
Conclusion

- the principles of DFT
  - localist representations form stable states
  - that may made unstable in a controlled way
  - through the “condition of satisfaction”
- enable the autonomous generation of sequences of mental motor states
  - => a fundamental first step toward higher cognition