On the benefits of emergent communication for threat appraisal

Kevin Godin-Dubois, Sylvain Cussat-Blanc and Yves Duthen

University of Toulouse, IRIT - CNRS UMR 5505, 2 rue du Doyen Gabriel Marty, 31042 Toulouse, France
{kevin.dubois, sylvain.cussat-blanc, yves.duthen}@irit.fr

Designing virtual agents that can positively interact with human users is a long-standing objective of Artificial Intelligence with numerous applications e.g. reducing elders isolation (Razavi et al., 2019) or improving individual health (Nakajima and Niitsuma, 2020). However, much of this research investment is focused on a top-down human-centered approach, whereas we argue that emergent communication in which actors collaborate on a common protocol (Lazaridou and Baroni, 2020) might serve similar purposes. In this abstract, we expose a framework where virtual embodied agents can communicate, vocally, with each other and the outside world through elementary signals thereby allowing arbitrary complex message structure. As a proof of concept, we illustrate their capabilities by investigating how both active (vocalization) and passive cues (noise, appearance) can be integrated by the agents to differentiate friends from foes.

Methodology

Such reliance on communication to correctly discriminate between a conspecific and a potential threat is the result of intricate interactions between multiple modeling levels. With concision in mind, we briefly introduce the main components of these agents and refer the reader to (Godin-Dubois et al., 2021) for a more detailed description. Genetically speaking, each individual is composed of both morphological (eye placement, body shape ...) and neurological parameters. In the latter case, these take the form of a Composite Pattern Producing Network (CPPN), an n-dimensional function used by ES-HyperNEAT (Risi and Stanley, 2012) to describe connectivity patterns.

From a semi-fixed topology of inputs/outputs (corresponding to the receptors/effectors depicted in figure 1), applying the ES-HyperNEAT algorithms results in a (potentially recurrent) neural network where hidden nodes are instantiated at locations of high informational density. Using a single, evolvable, component to encode for the creature's Artificial Neural Network (ANN), makes it possible to study the emergence of structures in response to specific evolutionary bias. Indeed, all of an agents' perceptions (color-based, multi-cell vision, and frequency-based hearing) and actions (dual motors and frequency/volume vocalization) are directly controlled by the activity of dedicated output neurons. In this manner, an individual's intrinsic behavior is solely the result of the CPPN that encoded its ANN.

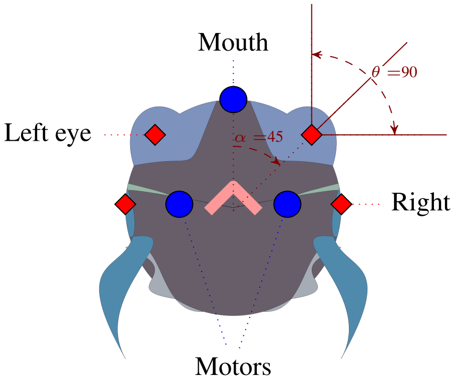
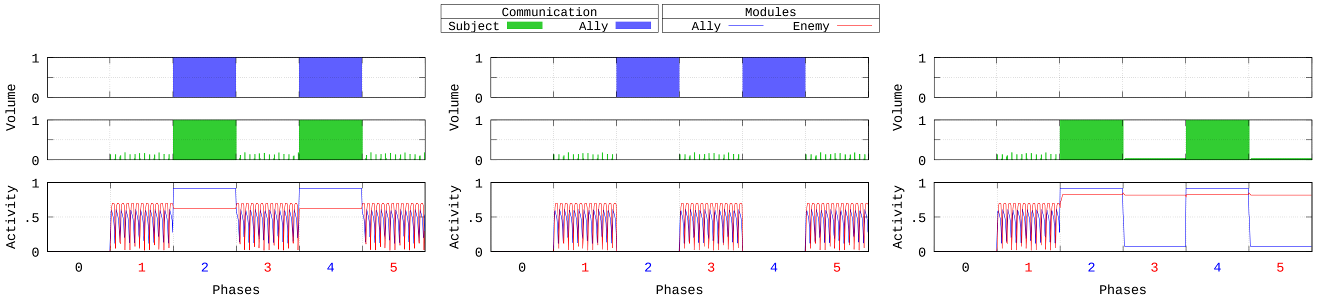


Figure 1: Anatomy of a creature showing the receptors (red) and effectors (blue)

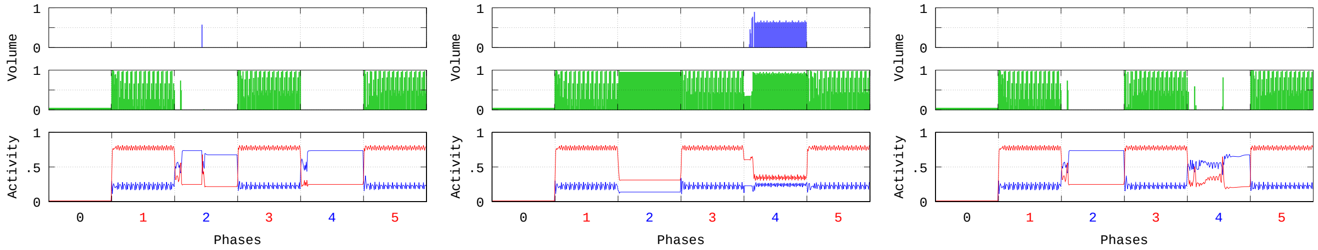
We focused our efforts on the neural basis of fear-like behavior through a prey/predator evolutionary game: collecting the food produces a large reward, being captured by the predator induces a small penalty, and running out of time is frugally rewarded. The ANNs of the resulting champions were then subjected to a clustering procedure to determine whether functional aggregates could be extracted from the undifferentiated mass of internal neurons. These evaluations were performed in controlled conditions with the subject being exposed to a given stimulus (or lack thereof) for 10 simulated seconds. Each neuron's activity was then correlated to the various stimuli, thereby allowing for the detection of task-specific sub-networks, hereafter called modules.

Benefits of communication

To highlight how an individual might leverage both passive (body color) and active (vocalization) communication, we compared the neural modules dynamics across three evaluation cases (standard, invisible and mute). In the standard procedure the subject was exposed to alternating stimuli: an empty environment in phase 0 and either a predator or an ally



(a) Subject 1 (standard) (b) S1 (invisible ally) (c) S1 (mute ally)



(d) Subject 2 (standard) (e) S2 (invisible ally) (f) S2 (mute ally)

Figure 2: Two instances of emergent communication. The bottom row of each subfigure shows the activity of the two modules correlated with threat and conspecific stimuli. Alternating stimuli (see text) are presented to the subject eliciting different types of responses. Top: excessive fear reaction when faced with a mute ally highlighting the necessity of both visual and auditory cues. Bottom: structured communication with different patterns depending on the current circumstances and history.

(harmless clone) in phases $\{1, 3, 5\}$ and $\{2, 4\}$, respectively. The subsequent variations around this pattern disable one communication medium to observe its impact on the overall behavior: in the second, while still capable of autonomous vocalization, the ally is invisible to the subject thereby precluding vision-based recognition and, in the last one, the ally is mute. We observed that a limited subset of the population exhibits increased activity of the predator module (20% and 8% for the invisible and mute cases, respectively. Wilcoxon test with p-value < 0.0005) when faced with partial inputs. Conversely, the neural module correlated with the ally stimuli has been shown to impact 78% of the population when faced with the invisible ally, as opposed to 14% for the mute alternative.

Indeed, in this experiment, communication was a desirable, albeit not explicitly enforced, partial solution to the prey/predator evolutionary game: numerous strategies are solely focused on sprinting towards the goal with no interaction either with the predator or ally. However, as highlighted by the previous tests, some individuals showed communicative behavior with varying degrees of complexity. Figures 2a-2c illustrate such a demeanor with the resulting differences in terms of verbal communication and neural activity being highly instructive on the underlying dynamics of this specific individual. Indeed in standard conditions both the subject and its ally agree on a binary mode of communication where simultaneously seeing and hearing a conspecific triggers an almost instant change of regime. However, one can also note that when presented with partial inputs the subject is either quiescent (fig. 2b) or suffers from an excessive (compared to its base regime, see fig. 2c) fear-like reaction: the neural module associated with threat-related stimuli stays activated, at peak value, throughout the remainder of the evaluation.

An alternative class of behavior is illustrated with subject 2 adopting different regimes based not only on the current conditions but also on its recent history: a single exchange is performed in phase 2 of figure 2d which seems enough to “reassure” it. In this instance, selectively disabling one communication medium instead results in different types of dialogues: either through a strong query of the neighbor’s identity (fig. 2e) or via sparser impulsions (fig. 2f).

It follows that this framework did exhibit both passive and active forms communication, though much of the conspecific identification was based on visual cues. This can be linked to the inherent problem of auditive neural inputs being harder to differentiate: while the red color can only be associated with the predator, motion-induced noise is not sufficient to discriminate between a friend and a foe. In future studies, we propose to address this issue by moving onto a 3-dimensional neural substrate which would facilitate the discovery of relevant independent components by ES-HyperNEAT. Moreover, targeted experiments on the incremental development of basic vocabulary, similarly to morse code, would provide a functional basis in low-level machine-human communication.

References

Godin-Dubois, K., Cussat-Blanc, S., and Duthen, Y. (2021). Spontaneous modular NeuroEvolution arising from a life/dinner paradox. In *ALIFE 21*, page in press, Cambridge, MA.

Lazaridou, A. and Baroni, M. (2020). Emergent Multi-Agent Communication in the Deep Learning Era.

Nakajima, K. and Niitsuma, M. (2020). Effects of Space and Scenery on Virtual Pet-Assisted Activity. In *Proceedings of the 8th International Conference on Human-Agent Interaction*, pages 105–111, New York, NY, USA. ACM.

Razavi, S. Z., Schubert, L. K., Kane, B., Ali, M. R., Van Orden, K., and Ma, T. (2019). Dialogue Design and Management for Multi-Session Casual Conversation with Older Adults.

Risi, S. and Stanley, K. O. (2012). An Enhanced Hypercube-Based Encoding for Evolving the Placement, Density, and Connectivity of Neurons. *Artificial Life*, 18(4):331–363.