Horizons for the Enactive Mind: Values, Social Interaction, and Play

Ezequiel A. Di Paolo, Marieke Rohde, Hanneke De Jaegher

CSRP 587

April 2007

ISSN 1350-3162

University of Sussex

Cognitive Science
Research Papers

Horizons for the Enactive Mind: Values, Social Interaction, and Play

Ezequiel Di Paolo, Marieke Rohde and Hanneke De Jaegher

Centre for Computational Neuroscience and Robotics (CCNR)
Centre for Research in Cognitive Science (COGS)
University of Sussex, Brighton BN1 9QH, U.K.

{ezequiel,m.rohde,h.de.jaegher}@sussex.ac.uk

To appear in *Enaction: Towards a New Paradigm for Cognitive Science*, J. Stewart, O. Gapenne, and E. A. Di Paolo (Eds), Cambridge, MA: MIT Press, forthcoming.

Abstract.

What is the enactive approach to cognition? Over the last 15 years this banner has grown to become a respectable alternative to traditional frameworks in cognitive science. It is at the same time a label with different interpretations and upon which different doubts have been cast. This paper elaborates on the core ideas that define the enactive approach and their implications: *autonomy*, *sense-making*, *emergence*, *embodiment*, and *experience*. These are coherent, radical and very powerful concepts that establish clear methodological guidelines for research. The paper also looks at the problems that arise from taking these ideas seriously. The enactive approach has plenty of room for elaboration in many different areas and many challenges to respond to. In particular, we concentrate on the problems surrounding several theories of value-appraisal and value-generation. The enactive view takes the task of understanding meaning and value very seriously and elaborates a proper scientific alternative to reductionist attempts to tackle these issues by functional localization. Another area where the enactive framework can make a significant contribution is social interaction and social undertanding. The legacy of computationalism and methodological individualism is very strong in this field. Enactivism allows us to see embodied social interaction and coordination at many different levels in an integrated manner, from the emergence of autonomous temporal structures that regulate

confined to the ‘lower levels’ of human cognition. This is the ‘reform-not-revolution’ interpretation. For instance, embodied and situated engagement with the environment may well be sufficient to describe insect navigation, but it will not tell us how we can plan a trip from Brighton to La Rochelle. Or enactive phenomena the situation will not change.

We dedicate this paper to clarifying the central tenets of enactivism and exploring some of its currently under-developed themes. In this exercise, following the logic of the central ideas of enactivism can sometimes lead to unexpected hypotheses and implications. We must not underestimate the value of a new framework in allowing us to *formulate the questions in a different vocabulary*, even if satisfactory answers are not yet forthcoming. Implicitly, the kind of non-reductive naturalism

constitution, can only follow the laws contained in its design, no matter how plastic, adaptive, or life-like its performance. In order for a system to generate its own laws it must be able to build itself *at some level of identity*. If a system 'has no say' in defining its own organization, then it is condemned to follow an externally given design like a laid down rail track. It may be endowed with ways

2

2.2 Sense-making

2.3 Emergence

The overarching question in cognitive science is: How does it work? For the enactive approach the connected concepts of autonomy and sense-making already invoke some notion of emergence in addressing this question. Autonomy is not a property of a collection of components, but the consequence of a new identity that arises out of dynamical processes in operational closure. Meaning is

but

saying

is

in abstracto

There is a further twist to the role played by the body in the case of human cognition – one that could explain the resilience of Cartesian modes of thinking. Even though our bodies are not puppets, to say that we control our bodies is, in a sense, not entirely wrong. We certainly do. But we do so in subtle ways that activity.

An embodied perspective results in serious attention being paid to isomorphisms between mechanisms and experience. Varela (1999) and van Gelder (1999) provide different, but related, dynamical systems accounts of mechanisms that might underlie the protentive and retentive structure of time consciousness as described by Husserl. Kelly (2000) considers neural models of pointing and grasping that run parallel to Merleau-Ponty's concepts of the intentional arc and maximal grip. Wheeler (2005) explores isomorphic relationships between embodied-embedded accounts of situated action and Heideggerian categories such as the ready-to-hand, breakdowns, and present-at-hand. What is interesting in many of these accounts is that the process of circulation is not one of assimilating scientific hypotheses into phenomenology, but may itself inform phenomenology. This is as it should be in a proper dialogue and such is the methodology advocated by first-person methods in the joint study of experience and brain-body activity (Varela, 1996; Lutz, 2002).

Experience may also serve the role of clarifying our commitments. Hans Jonas (1966) looks into the world of living beings and sees that life is a process with interiority. Metabolism has all the existential credentials of concernful being. It is precarious, it separates itself from non-being, it struggles to keep itself going and preserve its identity, and it relates to the world in value-laden terms. However, the inward aspect of life cannot be demonstrated using our current scientific tools. This does not make it any less factual for Jonas. He knows that all life is connected along an evolutionary continuum, and he knows that we ourselves are embodied living creatures with an inner life. This is how we can then know that living beings are forms of existence and that they also have an inner life.

the enactive approach clearly rejects, e.g., homuncularity, boxology, separability between action and perception, and representationalism. In this section we will revisit some of these themes in a more focused manner.

In everyday life we experience the world in value-loaded terms. This fact is hard to avoid and has been the subject of much philosophical debate throughout the ages. For enactivism, value is simply an aspect of all sense-making, since sense-making is, at its root, the evaluation of the consequences of interaction for the conservation of an identity. Perhaps as a reaction to the subjective overtones of this issue, traditional cognitive science has not dwelled much on the explicit mechanisms involved in value judgement as an inherent aspect of cognitive activity. In general, questions about value have or natural purposes have been dealt with separately, preferably with reference to evolutionary history (Millikan, 1984): everything living beings do is ultimately reduced to survival strategies in situations their ancestors encountered, or to the urge to spread their genes as widely as possible. In a more traditional modeling framework this idea translates to values being 'built-in' by evolution; phylogenetically invariant yardsticks against which actual lifetime encounters are measured and structured, and from which cognitive mechanisms that are themselves independent of these values deduce the meaning of situations, actions and perceptions.

Explanations of this kind are in tension with the principles of enactivism, in

If values are built-in, they need to have some form of priority over the living acting creature, either temporally or logically. Typically, claims about biological traits being built-in are about them being part of the genetic package. ‘Values’ is a term that describes the meaning of organismic behavior, not one of its physiological or mechanistic properties, like, for instance, the blood type. Therefore, the idea of built-in values relies on some kind of *a priori semantics*: parts of the genetic code are thought to execute according to pre-programmed rules and, thereby, generate values. This automated ‘sense look-up’ is not the same as sense-making, which we identified as one of the central concepts of the enactive approach. Similarly, we are dealing with *pre-factum* evolutionary teleonomy, not with autonomy. Instead of emergence, we find a direct reduction

some tendencies that bring them closer to the boundary of viability and acting to counter these tendencies can be actual 'sense-makers', not just robust to perturbations. A similarly subtle distinction is the one between adaptive organisms and interactive regulators (Moreno & Etxeberria, 2005): whilst the former act to counter hostile tendencies by changing their internal organization, the latter act on the environment and thereby exhibit the most fundamental form of agency. An example of a just-adaptive organism is the sulphur bacterium that survives anaerobically in marine sediments whereas bacteria swimming up a sugar gradient would, by virtue of their motion, qualify for minimal agency.

The further stages on the scale are largely adopted from Jonas's work. Animals, through their motility, exhibit the capacity to act and perceive as well as desire or fear something distal. And humans, through capacities such as image-making and ultimately of constructing a self-image, gain the ability to regard situations

> case of humans, that extend beyond the strict confines of the body into the socio-linguistic register" (Varela, 1991, p. 102).

It is certainly true that levels of value generation can be in conflict: how can it be that your body will fight for its life despite the deliberate attempt to end autopoiesis through an overdose of sleeping pills? Or, the other way around, and structural division between behavior-generating mechanisms and mechanisms of value-based adaptation is at the core of this type of architecture.

In order to point out the difficulties that result from such a separation of value judgement (built in) from value execution (ongoing), we now present two examples of our own research in computational modeling. The deliberately simple first set of simulation experiments is described in more detail in (Rohde & Di Paolo, 2006) and illustrates the difficulties of embedding functional modules into another wise dynamic and embodied system. A mobile, two-wheeled agent is controlled by a neural network, which is generated automatically, using an evolutionary algorithm, such that the agent's behavior optimizes a formal find that it responds positively to activity on the left light sensor, but negatively to activity on the right light sensor, which, intuitively, does not make a lot of sense. The successful judgement can only be understood by taking the sensorimotor context into consideration, i.e., the agent's light seeking strategy (figure 3(B)). If the agent does not see the light, it turns to the right, until it senses the light with both sensors. It then approaches the light from the right, constantly bringing the light source in and out of range of the right sensor. In the end, the agent cycles around the light source in small circles, perceiving the light with the left sensor only. Knowing this, it is much easier to understand how the 'value system' achieves a correct estimation of the distance. The approach behavior only starts when the light is in range of the left light sensor, and this sensor remains activated from then on, which explains the positive response to left sensor

This deterioration of performance is hardly surprising, given the structure of the value system and the way it works. But it demonstrates that value system architectures as outlined are not guaranteed to work without taking on board further premises. It has to be ensured that a value system estimates performance independent of the presence of reciprocal causal links, feedback loops and semantic drift of local structures. If a value system is implemented in a rigid context, as it has been done in some robots with a limited behavioral domain (Verschure et al., 1995), the meaning of the signal can be preserved independent of the modulation of behavior, such that the proposed circuits of adaptation do indeed work. However, in order to be convincing as a biological theory, it is necessary to specify how such a rigid wiring and disembodiment of value systems is realized in a living organism that is in constant material flux. This is exactly the kind of problem that classical computationalist approaches have failed to answer satisfactorily. Indeed, we see value systems, because of their disembodied nature and top–down supervision of adaptation, as leftovers from a Cartesian mode of thinking. Such leftovers are not surprising; decades of exercising a computationalist methodology persist in the very language used to formulate questions.

An enactive approach, however, is based on the idea that functional invariants, such as values, self-organize and emerge from a constantly varying material substrate. They are not reduced to loca

And Ami Klin and his colleagues (2003) have also produced a so-called enactive approach to the social domain; however, it remains chiefly focused on perception.

4.1 Towards enactive social understanding

Before laying out our proposal for an enactive approach to social understanding, let us have a look at the gaps in traditional takes on social cognition. The underlying assumption of central paradigms such as Theory of Mind theory (ToM) and simulation theory is that minds are enclosed and opaque, and hence others are puzzles for us to solve. The proposal of ToM accounts as regards social understanding is that we cognitively figure out others: we understand others by

discusses ‘interpersonal engagement’. This is the intersubjective sharing of experiences, which infants are already good at and which forms the fertile

empirical study of ‘perceptual crossing’. Following from this, we will make the link to meaning generation in social interaction via the introduction of the notions of and .

One approach to the question of how coordination between social interactors be used strategically. Because of their random initial positions coordination between the agents must be achieved in order to facilitate a continuing interaction.

Successful agent pairs acquire a coordinated pattern of signaling in which empirical research conducted on human subjects that is driven by a similar

interaction rhythm

the individuals involved. We define this central capacity of social cognition as: flexibly temporally coordinating through the interaction with another person. Through such flexible coordination, the rhythm of an interaction can be adapted

essential to human cognition as opposed to other forms of animal cognition? Margaret Donaldson (1992) formulates the issue in a very useful way. She puzzles about the amazing human capability of constantly inventing new goals so that we invest them with value and submit passionately to them (sports, hobbies, record-breaking). An explanation of human intelligence should perhaps not concentrate so much on issues such as, say, how do we manage to do maths. It should bring to the centre the question of *why* we do those things at all? When did they become valuable for us?

Donaldson describes different ways to be a human mind. As a developmental psychologist, she concentrates on how transitions between these different modes occur throughout a lifetime. The question parallels how Jonas and others have treated the history of mind as transitions in scales of mediacy. Donaldson distinguishes four modes in which we function as minds depending on the focus of concern. This is amenable to the whole of our previous discussion. To have different foci of concern is no more or less than to have different modes of value-generation. The *point mode* deals with here-and-now coping (most animal activity, skilful practices in humans). The *line mode* expands the focus of concern to the immediate past and the possible future as well as to other spatial localities (understanding of immediate causes and consequences of events). The *construct mode* produces a de-centering of cognitive activity; concern focuses on events that have happened or may happen at some point in time or somewhere, and not necessarily involving the cognizer (induction, generalization). Finally, the *transcendent mode* has no locus; it deals with nowhere, no-time (abstract thought, metaphysics).

activity of play itself helps the child make this urge clearer.

How is this possible within an accommodation/assimilation/equilibration contextual manner so that they can continue to play and fun does not run down by exhausting the possibilities of the game. Rules are made-up in play; they are solidified versions of norms. Fun is the exploration of the limits thus imposed on thinking is at home with linear causality, well-defined and unchanging systems, and reduction. The alternatives of emergent, many-layered, causally spread, non-linear systems in constant constitutive and interactive flux are very hard to manage conceptually. This is an important focus of resistance to many enactive ideas. It is here that synthetic modeling techniques may have their major impact: in producing novel ways of thinking and generating proofs of concept to show that some proposals may not be as coherent as they sound (as in our critical study on value system architectures) or to demonstrate that apparently hazy concepts find clear instantiations even in simple systems (as in the case of emergent coordination through social interaction processes). Methodological minimalism is, therefore, a key element contributing to the acceptability of enactive ideas.

Models that attempt to illuminate the enactive framework will have to take into account the core ideas of enactivism. A serious take on embodiment will depend on the extent to which a system's behavior relies non-trivially on its body and its sensorimotor coupling with the environment as opposed to input-output information processing. Emergent properties and functionality will contrast with misplaced localization in sub-agential modules. Autonomy, to the extent that it can be captured in simulation or robotic models, will depend on how the model instantiates the dynamics of self-constituted precarious processes that generate an identity and how such processes create a normativity at the interactive level that leads to sense-making. Enactive modeling must also relate to experience. As a scientific tool it belongs to the realm of third person methods and so the relation will have to find its place in the process of mutual constraining that has

Ashby, W. R. (1960). *Design for a brain: The origin of adaptive behaviour* (Second edition). London: Chapman and Hall.

Auvray, M., Lenay, C., & Stewart, J. (2006). The attribution of intentionality in a simulated environment: the case of minimalist devices. In *Tenth Meeting of the Association for the Scientific Study of Consciousnes, Oxford, UK, 23–26 June, 2006*.

Bach–y–Rita, P., Collins, C. C., Sauders, F., White, B., & Scadden, L. (1969). Vision substitution by tactile image projection. *Nature*, 221, 963–964.

.

De Jaegher, H. (2006b).
.
Unpublished DPhil Thesis, University of Sussex.

Dennett, D. C. (1993). Review of F. Varela, E. Thompson and E. Rosch, The Embodied Mind. , 106, 121–126.

Dewey, J. (1929). (Second edition). New York: Dover. (1958).

Di Paolo, E., Rohde, M., & Iizuka, H. (2007). Sensitivity to social contingency or stability of interaction? Modelling the dynamics of perceptual crossing. . Forthcoming.

Di Paolo, E. A. (2000a). Behavioral coordination, structural congruence and entrainment in acoustically coupled agents. , 8, 27 – 47.

Di Paolo, E. A. (2000b). Homeostatic adaptation to inversion of the visual field and other sensorimotor disruptions. In Meyer, J.-A., Berthoz, A., Floreano, D., Roitblat, H., & Wilson, S. (Eds.),
. Cambridge MA: MIT Press.

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. , 4, 429–452.

Donaldson, M. (1992). . London: Penguin Books.

Dreyfus, H. L. (2002). Intelligence without representation - Merleau-Ponty's critique of mental representation. The relevance of phenomenology to scientific explanation. , 1, 367–383.

Edelman, G. M. (1989). .
Oxford: Oxford University Press.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. , 2, 704–716.

Fagen, R. (1981). vior. Oxford: Oxford Univ

Studies, 41, 19–30.

Gallagher, S. (1997). Mutual enlightenment: Recent phenomenology in cognitive science. *Journal of Consciousness Studies*, 4, 195 – 215.

Gallagher, S. (2001). The practice of mind: Theory, simulation or interaction?. In Thompson, E. (Ed.),

Lutz, A. (2002). Toward a neurophenomenology as an account of generative passages: A first empirical case study. *Phenomenology and the Cognitive Sciences*, 1, 133–167.

Maynard Smith, J., & Szathmáry, E. (1995). *The major transitions in evolution*. Oxford: W. H. Freeman.

McGee, K. (2005). Enactive cognitive science. Part 1: Background and research themes. *Constructivist Foundations*, 1 (1), 19–34.

La science et l'hypothèse

Advances in Artificial Life: Sixth European Conference on Artificial Life

Nature

Cognitive science research paper 584, COGS, University of Sussex

Proceedings of the Fourth European Conference on Artificial Life

Stewart, J., & Coutinho, A. (2004). The affirmation of self: A new perspective on the immune system. *Artificial Life*, 10, 261–267.

Sutton-Smith, B. (1966). Piaget on play: a critique. *Psychological Review*, 73, 104–110.

Sutton-Smith, B. (1997). *The ambiguity of play*. Cambridge, MA: Harvard University Press.

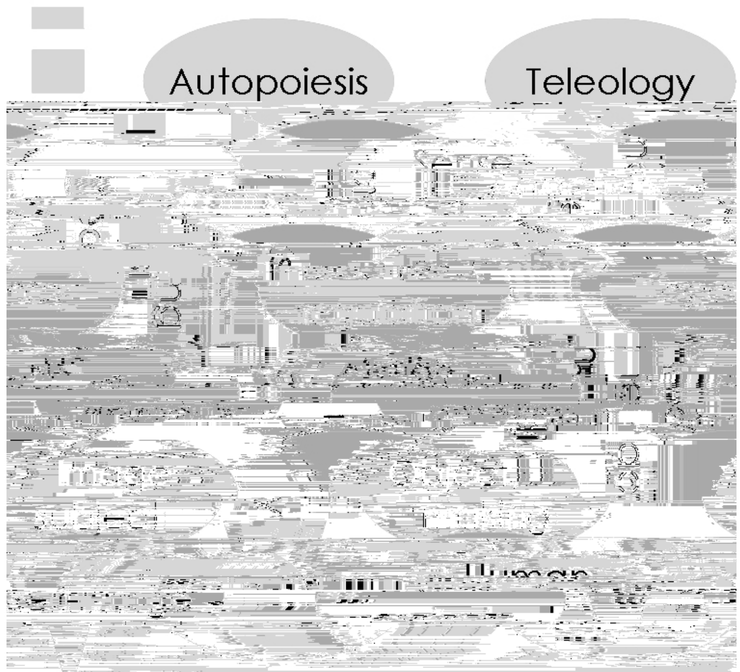
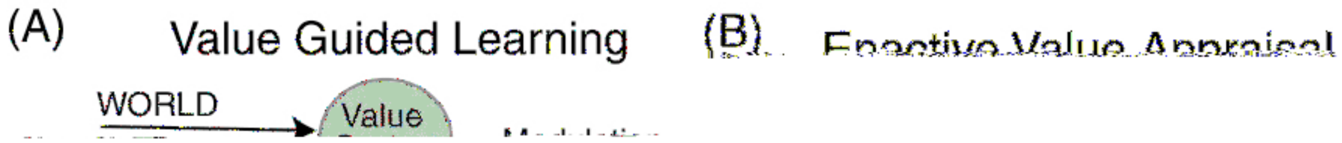


Figure 1: Life-mind continuity and the scale of increasing mediacy, (see text).


(A)
Value Guided Learning
(B)
WORLD
Value

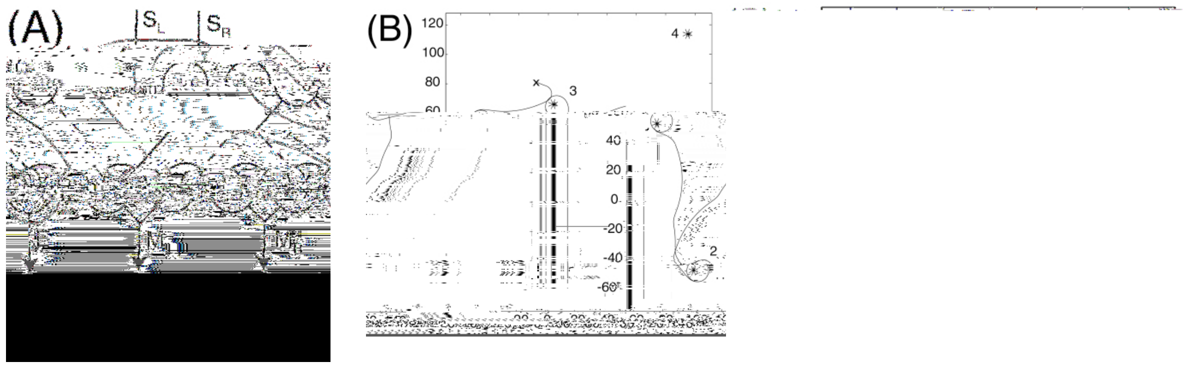
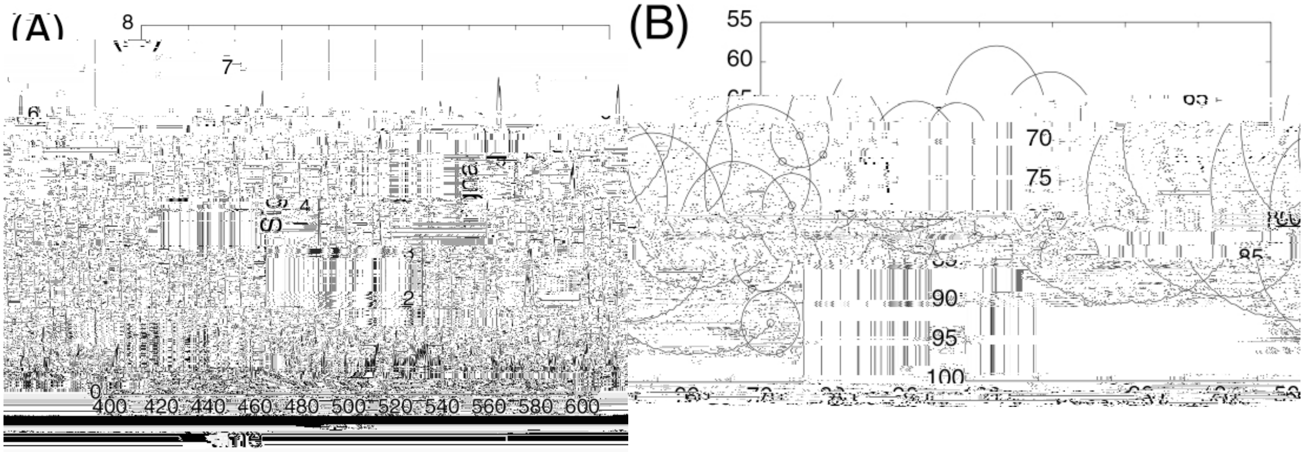


Figure 3: (A) The value judging and light seeking agent controller. (B) The successful light seeking behavior (C) The deterioration of light seeking through applications of the principles of neural Darwinism.


(A)
(B)

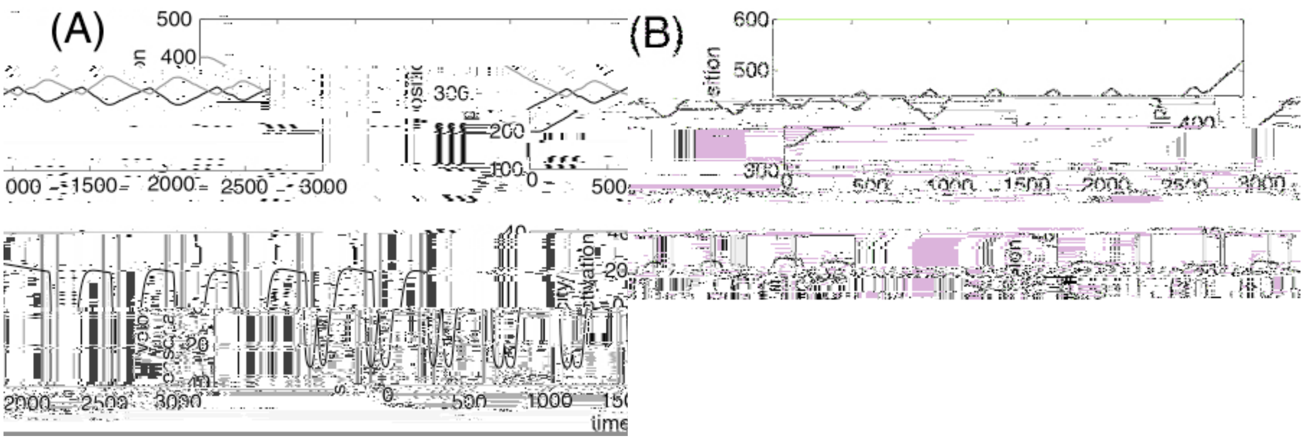


Figure 6: Perceptual crossing model. Top plots show the trajectories of agents over time; plots at the bottom show the motor commands (dark line) and sensor input (gray line). (A): Stabilized social perceptual crossing, (B): scanning of a fixed object.

Notes

1