



Research article

AHA: A general cognitive architecture for Darwinian agents

Sergey Budaev*, Jarl Giske, Sigrunn Eliassen

Department of Biological Sciences, University of Bergen, Bergen, Norway



ARTICLE INFO

Keywords:

Cognitive architecture
Ecology
Emotion
Ethology
Evolution
Genetic algorithm

ABSTRACT

Unified theories of cognition have traditionally played a vital role in understanding the human mind. In the animal behavior field, however, acceptance of holistic views on the behavioral phenotype that includes diverse cognitive and behavioral traits is rather slow. Studying adaptation and evolution of behavior, especially complex cognition and decision making, requires integrative models applicable to a range of species. We describe a general cognitive architecture and a modeling framework for studying evolution and adaptation of behavior and cognition that we call Adapted Heuristics and Architecture (AHA). AHA is non-symbolic, rule-based and grounded in general neurobiological mechanisms. It integrates the whole organism with its genome, physiology, hormones, perception, emotion, motivation and cognition in an agent based model environment. The method lets us investigate various scenarios for the evolution of cognition, decision making and emergence of subjective phenomena. We illustrate the potential feasibility of the framework with a model of simple forms of self-awareness.

Introduction

All the fascinating complexities of life including the human mind are the results of millions of years of Darwinian evolution by natural selection. Evolution occurs when morphological, physiological, cognitive and behavioral traits have genetic basis and vary across individuals (Dawkins, 1986; Lewontin, 1974). Behavior is a major target and driver of evolutionary change (Duckworth, 2008), but studying behavioral evolution poses fundamental challenges. Most behavioral traits are causally distant from the genome, and display high degree of both plasticity and individual variability (Dingemanse, Kazem, Réale, & Wright, 2010; West-Eberhard, 1989).

Ever since Fodor (1983) mind, cognition and brain organization are viewed as modular (Bertolero, Yeo, & D'Esposito, 2015; Meunier, Lambiotte, & Bullmore, 2010; Sternberg, 2011). In classical ethology, behavioral organization has been viewed as hierarchically modular (Hogan, 2009, 2015; Toates, 1998). The behavioral phenotype is modular too. For example, many studies reveal correlations across various behavioral domains (Sih, Bell, Johnson, & Ziemba, 2004) which give rise to consistent personalities in animals of various taxonomic groups similar to human personality (Budaev & Brown, 2011; Dingemanse, Kazem, Réale, & Wright, 2010; Gosling, 1999; Reale, Dingemanse, Kazem, & Wright, 2010). These correlations can be caused by common genetic, physiological, developmental, neurobiological or cognitive mechanisms (Arnold, 1992; Budaev & Brown, 2011), but can

also be the result of ecological adaptation (Budaev & Brown, 2011; Reale et al., 2010; Wilson, Clark, Coleman, & Dearstyne, 1994). Consistent correlations across different cognitive domains can point to modularity in the underlying architecture of the cognitive processes (Rabaglia, Marcus, & Lane, 2011). Modularity constrains and facilitates evolutionary adaptation of cognitive and behavioral traits (Andersen, Jørgensen, Eliassen, & Giske, 2016; Clune, Mouret, & Lipson, 2013; Giske et al., 2014). This agrees with the emerging views on modularity in biological systems (Lorenz, Jeng, & Deem, 2011; Wagner, Pavlicev, & Cheverud, 2007), where evolved *adaptive architectures* are found across many organizational levels from the genome to cognitive mechanisms.

In spite of the growing empirical evidence for the importance of the integrated phenotype (McGlothlin & Ketterson, 2008; Murren, 2012), most mathematical and optimization models of animal behavior in evolutionary ecology ignore these proximate complexities. By a broad assumption that the phenotypic traits are unconstrained, the tradition in evolutionary ecology has been to model fitness consequences of specific behaviors directly (Emlen, 1966; Fisher, 1930; Houston, Clark, McNamara, & Mangel, 1988, but see Pierce & Ollason, 1987). This assumption, called “the phenotypic gambit” (Grafen, 1984), is justified as a starting point but becomes increasingly unrealistic when studying cognition and behavior in a novel, changing and stochastic environment (Fawcett, Hamblin, & Giraldeau, 2013; Giske et al., 2013, 2014).

In cognitive science, psychology and neurobiology, the approach based on cognitive architecture has experienced an immense success

* Corresponding author at: University of Bergen, Department of Biological Sciences, Postboks 7803, 5020 Bergen, Norway.
E-mail address: sbudaev@gmail.com (S. Budaev).

(Anderson, 1983, 2007; Anderson et al., 2004; Duch, Oentaryo, & Pasquier, 2008; Lucentini & Gudwin, 2015; Samsonovich, 2012). Cognitive architecture is defined as an “overall, essential structure and process of a broadly-scoped domain-generic computational cognitive model, used for a broad, multiple-level, multiple-domain analysis of cognition and behavior” (Sun, 2004). It is a basic framework for the organization of cognition and behavior that can be extended in various ways to build more detailed and specific computational models of these phenomena that focus on particular research problems and purposes (Anderson, 1983; Duch, Oentaryo, & Pasquier, 2008; Newell, 1994; Sun, 2004; Vernon, Von Hofsten, & Fadiga, 2016). Symbolic models that dominate this area, however, are difficult to use for studying animals which lack symbolic processing and language. Thus, there is a need to develop relatively simple holistic yet non-symbolic architectural models suitable for the study of Darwinian evolution of behavior, cognition and mind across a range of species. Such a framework should be really *minimalist* and only include the most essential innate or simplest learning capabilities. It should be *generic*, i.e. not linked to any specific theory, methodology and paradigm (Sun, 2004). In addition, it should be based on simple, relatively low-level neurobiological processes and implement embodied decision making and continuous action selection in real time, as the organism interacts naturally with its environment (Cisek & Pastor-Bernier, 2014; Seth, 2007).

Here we describe an evolutionary modeling framework for cognitive architecture that can be used to study adaptation of behavior, cognition, emotion and mind that we call *Adapted Heuristics and Architecture (AHA)*. This cognitive architecture is non-symbolic, rule-based, grounded in general neurobiological patterns. The AHA framework integrates cognition and behavior in the whole organism with its genome, physiology, perception, emotions, motivation and cognition. This architecture is implemented in the form of a large scale agent-based simulation model (e.g. Budaev, 2018) where artificial organisms evolve by natural selection across many generations (Andersen et al., 2016; Eliassen, Andersen, Jørgensen, & Giske, 2016; Huse & Giske, 1998). The AHA method lets us investigate evolutionary scenarios for the emergence and adaptive role of subjective phenomena. This aim is closely related to the BICA challenge: development of a general-purpose computational equivalent of the human mind (Samsonovich, 2012).

Building blocks of AHA

Overview

The cognitive architecture provides a general meta-model that could be implemented in specific computer algorithms, usually involving a procedural agent-based “machinery” (Sun, 2009). The architecture in AHA consists of simple generic units that are combined and extended to build more specific adaptive evolutionary models. It is implemented at an intermediate level of functional modules and their connecting pathways: it works above the level of individual neurons but below symbolic or other high-level representations. In the [Online Supplement](#) we provide a brief overview of the most important theoretical concepts that provide the major building blocks for the AHA architecture: the survival circuit, emotion, motivation, global organismic state, arousal, attention and prediction-oriented cognition with elementary subjective simulation.

The adaptable cognitive architecture (Fig. 1) is the main feature of an evolutionary agent-based model system—the AHA model—that includes the environment and populations of agents that evolve by natural selection. The environment includes various objects, such as food items that serve as the source of energy and predators that prey on agents. Each agent is a whole virtual organism with a unique genome, simple hormonal system, physiology, neurobiology and behavioral repertoire.

Overall, control of behavior in the organism is achieved through an evolved balanced dialogue between ancient and continuously adapting

programs linking perception, cognition and behavior. The central component of AHA architecture is the *survival circuit* (LeDoux, 2012, see [Online Supplement](#)) integrating multiple sources of perceptual information into a unitary internal state. The internal state depicted by the survival circuit is categorical but also has a continuously graded activation level or *motivation*. The agent has several survival circuits that encapsulate different *emotional states*: fear, hunger, thirst, reproductive drive etc. These internal states trade priority with each other by competing through lateral inhibition so that only one becomes the *global organismic state* (GOS). The level of motivation for this state becomes the *arousal* of the agent. The global organismic state exerts an arousal-attuned top-down attention modulation effect suppressing all perceptions that are not linked to the current GOS.

The life cycle of the agent involves a constant flow of perception → action: adaptive decision making and naturalistic behavior in the model environment. However, most perceptions will not lead to any action. The GOS determines which behavioral action is to be executed. Actions, defined structurally (Hogan, 2015), are selected from a fixed repertoire through various behavior selection paradigms (see below). Agents that survive and successfully secure resources (grow) over the life cycle can pass on their genetic information to the next generation. The offspring genome is subject to genetic recombination and random mutations. Over repeated generations, the genetic algorithm leads to evolutionary adaptation of the agents and the gene pool (Banzhaf & Eeckman, 1995; Huse, Strand, & Giske, 1999; Ruxton & Beauchamp, 2008; Strand, Huse, & Giske, 2002).

The AHA model uses nonparametric functions that are based on nonlinear interpolation (Ioannou & Pitsillides, 2008; Szabados & Vértesi, 1990). This makes it sufficiently generic, free of specific theoretical assumptions, and hopefully well-suited for a range of data-driven models. We have implemented AHA as a programming framework written in object oriented Fortran (see the [Online Supplement](#) for more details) provided as open source at <http://ahamodel.uib.no>.

The survival circuits: Perception and motivation

The perception mechanism allows the agent to obtain information about its own physiological states, physical properties of the environment, as well as various spatial objects. In the simplest implementations perception does not require recognition. In more complex models object recognition can be implemented by connectionist modules mapping fuzzy perceptual patterns into specific sensory signals (see Albantakis, Hintze, Koch, Adami, & Tononi, 2014; Strand, Huse, & Giske, 2002). Sensory signals that are fed from the agent perception are mapped to the neuronal response by the agent's individual genome. The relationship between the perceived sensory stimulus P and the resulting neuronal response R (see Fig. 1) is defined by additive sigmoidal functions (see [Online Supplement, 2.1](#)). Since the work of Cybenko (1989), the sigmoidal function is widely used for input-to-output mappings in the neural networks (Brette et al., 2007; Churchland & Sejnowski, 1992; Lytton, 2002), for example nonlinear perceptual stimulus integration (Gollisch & Herz, 2012), stimulus categorization (Mysore & Knudsen, 2012) and gain modulation (Brozović, Abbott, & Andersen, 2008). AHA extends this by combining several sigmoidal functions, which can represent more complex combinations of sigmoidal, concave, convex or nearly linear relationships (Andersen et al., 2016; Eliassen et al., 2016; Giske et al., 2013, 2014).

Outputs from different neuronal response functions that contribute to the same survival circuit are integrated (Fig. 1), giving a single value of motivation (M) for each circuit. This initial motivation value (primary motivation M_1 , Fig. 1) can be modulated by genetic or non-genetic factors (e.g. hormones influenced by age or life cycle stage of the agent), resulting in the final motivation (M_f , Fig. 1). Reproductive motivation can for example be suppressed in young agents allowing accumulation of an internal hormone-dependent reproductive factor without raising reproductive motivation.

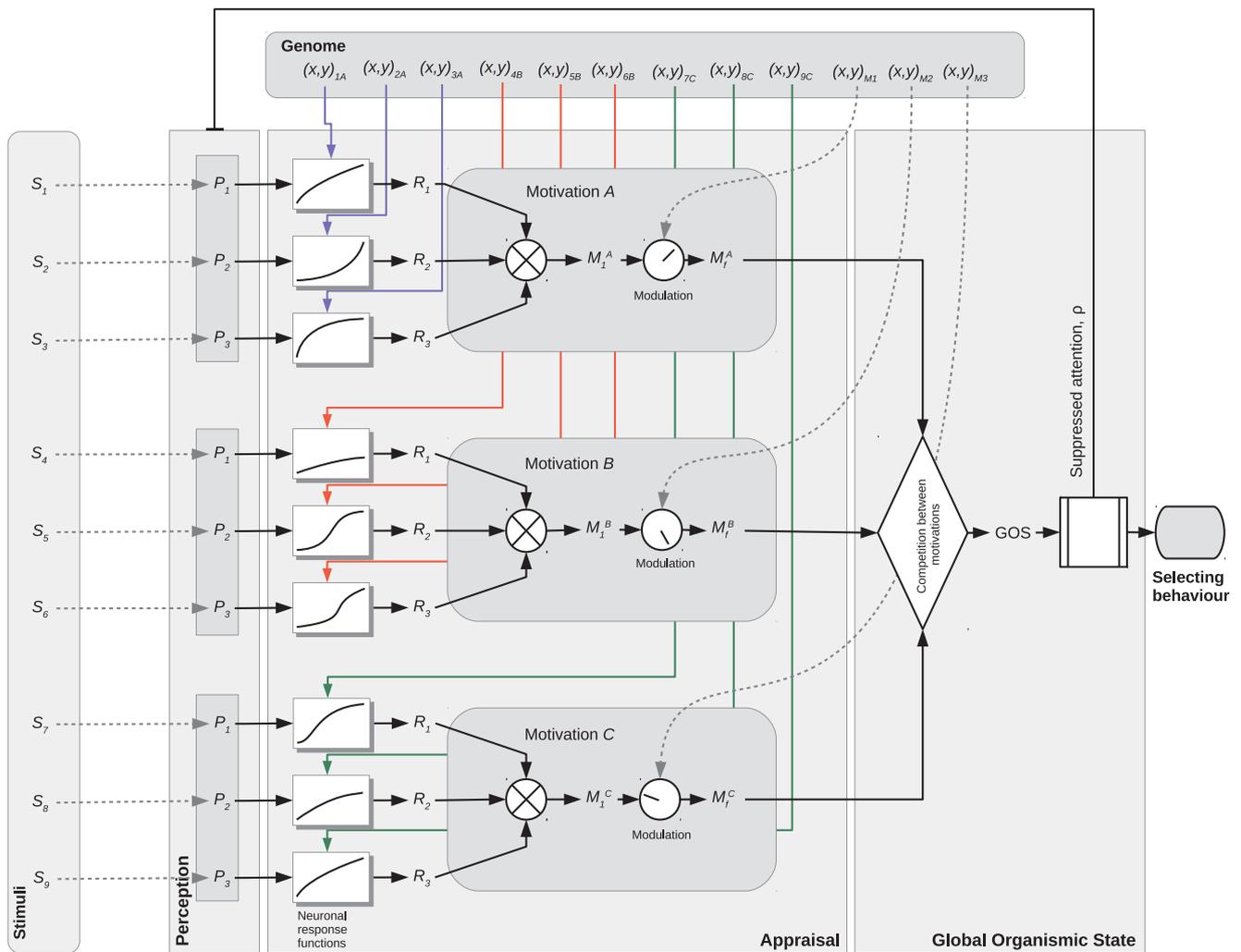


Fig. 1. A simplified scheme of the AHA cognitive architecture assuming three survival circuits and nine stimuli. See text for explanation and symbols.

Affective competition and the global organismic state

Once the final motivation values (M_f , Fig. 1) are calculated for all survival circuits, they compete with each other to become the next GOS of the agent. If no incoming motivation is strong enough, the previous GOS is retained. Competition between the incoming motivations and the current GOS is based on a dynamic threshold (for details see [Online Supplement](#)), which is an inverse convex function of the current arousal level G . Briefly, if the current arousal is low, a relatively *high* competing motivation is required to win competition with the current GOS. Conversely, when the current GOS arousal is *high*, just a marginally higher incoming motivational value is enough to become the next GOS (see [Online Supplement 2.2](#) for details). This makes it relatively easy to switch the internal state when the arousal is high, but more difficult when it is low, providing for continuity of the internal state. The dynamic threshold reduces dithering-like switching between different behavioral actions (see [Bryson, 2000](#); [Lewis & Canamero, 2016](#); [McFarland & Sibly, 1975](#)) which is considered maladaptive for both animals ([Barnard, 2004](#); [McFarland & Sibly, 1975](#)) and artificial agents ([Lewis & Canamero, 2016](#)). This issue is common in decentralized bottom-up architectures ([Bryson, 2000](#)), for example, a robot trying to satisfy several needs simultaneously could keep dithering without guaranteed gratification of any of these needs. In spite of behavioral persistence, the agent should be able to switch its internal state easily when a high arousal GOS is confronted with a high-incentive stimulus,

e.g. if a very hungry or aggressive animal suddenly sees a predator in proximity.

When the current GOS is not outcompeted by an incoming motivation, the old GOS is retained, but the arousal level is reduced (arousal dissipation). Maintaining a high level of neuronal activity is energetically costly ([Ames, 2000](#); [Niven, 2016](#)) and involves synaptic depression due to neurotransmitter exhaustion leading to habituation ([Christoffersen, 1997](#)). This will also prevent the agent from getting stuck on a specific GOS as it will ultimately be encouraged to switch to a different state. The combination of the GOS dynamic threshold mechanism and spontaneous dissipation of repeated arousal state results in substantial continuity of the dominant internal state and can produce agents with various degrees of behavioral flexibilities ([Landsrød, 2017](#)).

Attention control

The GOS at time t modulates the agent's perception at the next time step $t + 1$. Each of the perception signals that is *not* associated with the current GOS is suppressed by the attention modulation factor (see [Online Supplement 2.3](#)). This mechanism provides an effective top-down control of perception that dynamically depends on the current GOS. For example, a frightened agent in a highly aroused state would obtain weakened perceptual signals of food and hence not be distracted from its anti-predator response.

Coping with novelty and uncertainty

Coping with novelty is a special problem in both behavioral ecology and robotics. Although a live organism and robot (Oudeyer, Kaplan, & Hafner, 2007) can use intrinsic exploratory activity to learn from the environment, this would not work when it is suddenly confronted with a totally novel context for which no information—and no expected utility—is available. Confronting an animal with a novel environment is frequently used to test coping strategies and personality because such a test often reveals a wide range of individual differences (e.g. Budaev & Brown, 2011; Sih et al., 2004). In AHA, the agent can adaptively deal with such situations through the so-called *hope function*. Briefly, the expected utility of the novel option is constructed from the pattern of the last perception memory (see Online Supplement 2.4). This can facilitate adoption of the novel option by the agent in situations of diminishing returns. By changing the parameters of the hope function it is possible to produce a range of strategies from novelty avoiding to novelty seeking, which provides a mechanistic tool to model novelty bonuses (see Kakade & Dayan, 2002).

Decision making and action selection

The GOS determines which behavioral action shall be executed. AHA allows various behavior selection paradigms (Bryson, 2000; Seth, 2007; Tyrrell, 1993). In the simplest case the GOS fully dictates the action. For example, if the agent is in the hunger state and there is at least one food item in perception, “eat the nearest food item” action is executed. If the agent perceives no food in its vicinity, a Gaussian random walk can be executed to search for food. Because a long walk puts the agent into a novel environment, the walk size could be determined by the hope function mechanism: If local food availability is diminishing, the expected utility of a longer walk would increase. Thus, the agent could optimize its exploration/exploitation ratio by producing Lévy-like movement patterns known to facilitate finding of rare but clumped food in stochastic environment (Bartumeus & Catalan, 2009; Reynolds, 2015, 2018; Wosniack, Santos, Raposo, Viswanathan, & da Luz, 2017).

In more complex models, the action selection procedure within a GOS can involve competition between several potential actions that best optimize certain criteria (see also Arkin, 1998; Brooks, 1991). This could involve prediction-oriented mechanisms, e.g. predicted reward or arousal and also predicted response in other agents. Each survival circuit can have its own action selection implementation that is activated if this circuit is the GOS. Thus, AHA allows various hierarchical organizations with competing units at each level.

Dual-process framework

The dual process theories of mind postulate two basic modes of operation: autonomous/reflexive and predictive/reflective (Anderson & Lebiere, 2003; Irlbacher, Kraft, Kehrer, & Brandt, 2014; Vaughan et al., 2016). A similar distinction is made in the animal learning and cognition literature, e.g. between associative and cognitive (goal-directed) responses (Dickinson, 1985, 2012; Toates, 1997). Various arbitrary combinations of these two modes can be implemented in AHA, e.g. a fixed stimulus-response reaction for fear and a complex selection of optimal food items to minimize expected hunger.

Elementary self-awareness and neural reuse

One possible prediction-oriented action selection paradigm is based on what we call elementary self-awareness. It is defined here as *the ability of the agent to assess its own internal state and use this information for decision-making and action selection*. In AHA, emotions that are activated by specific survival circuits represent such internal states. The emotional arousal represents a common subjective currency for

decision-making. Thus, the agent selects and executes the action that optimizes the expected emotional arousal (see Online Supplement sect. 2.5 for more details). Consequently, *prediction of the environment by the agent is mediated by prediction of the agent's own internal state*. This reduces a virtually infinite number of possible environmental states and parameters to a small set of internal subjective variables. In effect, decision making is simplified due to grossly diminished combinatorial complexity: now the agent does not need to calculate utility expectations over all possible environmental factors and their effects on all its internal states. In AHA, this computational challenge is moved from individual learning to adaptive evolution of the population gene pool. We propose that subjective prediction mediation may provide one solution to the computational complexity issue in decision making theory (Bossaerts & Murawski, 2017).

The notion of elementary self-awareness in AHA is intimately linked to specific action rather than abstract information processing: it involves computationally simple subjective simulations of the affective value of such actions. Conceived in this way, elementary self-awareness would not require complex dedicated neural structures. Instead, it reuses already existing survival circuits that conduct affective computation (neural reuse, see Anderson, 2010). However, it adds a relatively simple component for reentrant¹ activation of the existing survival circuit. Thus, AHA is consistent with the emerging view that components of awareness and consciousness may not require complex dedicated brain machinery and could emerge even in organisms with simple nervous systems (e.g. Barron & Klein, 2016).

Computational example: Can simple self-awareness evolve?

We illustrate the feasibility of the AHA framework by a model of elementary self-awareness. The model is similar to the fish system developed by (Giske et al., 2013, 2014). The source code and detailed documentation are described in (Budaev, 2018). A population of plankton-eating fish lives in a vertically stratified environment where their food performs diel vertical migrations (moving upwards in the water column at night and downwards during the day). Pelagic water masses have strong and predictable vertical gradients of light intensity that affect both prey encounter rate and predation risk. All interactions in this model are local and based on vision. That is, predators and prey have no knowledge outside of their limited visual sensory range. Visibility of an object (visual range) is a function of its size, contrast and illumination level (Aksnes & Utne, 1997). For both the fish and its prey, the risk of being detected by a predator increases with light intensity and prey size. Competition for food and dilution of risk could make opposing effects on the profitability of staying close to conspecifics. This simple system provides physical gradients, environmental dynamics and stochasticity.

To find food, the fish must follow the constantly changing vertical distribution of its prey. The energy gained from feeding is spent on activity and basic maintenance with surplus energy allocated to growth and reproduction. If a fish does not eat, it will quickly die from starvation. The neurobehavioral system of the fish is built according to the above architecture and has three survival circuits: hunger, fear and reproduction. It also has a behavioral repertoire including various movements (including vertical migrations) immobility, escape, long-range migration and feeding on plankton. Behavior selection at each time step minimizes the expected negative emotional arousal by reentrant assessment as described above. All behaviors are selected on this basis, i.e. the agents act based on elementary self-awareness. Trading between feeding, predator avoidance, avoidance of competing conspecifics and attraction to conspecifics to dilute risk would involve

¹ Reentrant processes involve repeated activation loops across neuronal ensembles and are thought to underlay conscious experiences (see Baars, 1995; Edelman, 2004).

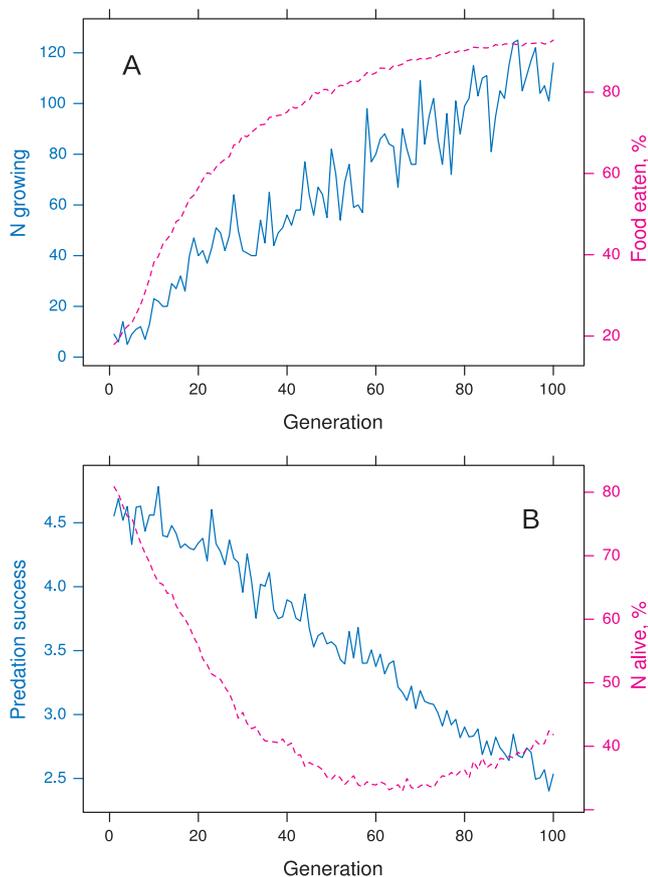


Fig. 2. Results of a computer simulation that demonstrate adaptive evolution of AHA agents implementing behavior selection that minimize the expected emotional arousal by reentrant assessment. A. The number of agents that increased their weight during the life cycle and the proportion of the food that was consumed by the agents during each generation. B. Predation success (i.e. the number of agents killed by the predators per single time step) and the number of agents surviving to the end of their life cycle. See text for more details.

quite a complex computational problem. This is in part due to the flexible and not fully deterministic link between local environmental conditions and the behavior mediated by action selection based on reentrant processing. Nonetheless, our simulation point that evolutionary adaptation is possible in this system. Thus, the population-level “calculations” involving mutations, recombinations, and natural selection across generations produce agents that are making increasingly better decisions, even though they only consider the local and current situation and have no individual learning.

The results in Fig. 2 are based on a single short (100 generations, with population size 10,000 agents) simulation and are used only for illustrative purposes. Not surprisingly, quite few agents succeeded during their life cycle out of the initially random population. However, the AHA cognitive architecture implementing complex behavior selection with reentrant self-simulation can be optimized through adaptive evolution. Indeed, the agents evolving over these generations exhibit steadily increasing growth and capacity to find food (Fig. 2A). Already after about 50 generations, the agents were able to consume more than 80% of all food items available in the environment (Fig. 2A), making the task of finding the increasingly scarce remaining food resource difficult (note that the food resource was not replenished within the generation in this simulation). Predation success also fell over the repeated generations (Fig. 2B). This points that the AHA cognitive architecture implementing elementary self-awareness through reentrant activation mechanism can indeed be optimized by the evolutionary

adaptation. Thus, it represents a viable tool for the study of adaptive mechanisms that shaped cognition and subjective phenomena through evolution.

General discussion

The major paradigm in psychology, ethology and neuroscience is that behavior is just a series of responses to stimuli (Edelman, 2016; Heisenberg, 2014). There is, however, a growing recognition that much of the behavior is in fact generated endogenously (subjectively) by the animal, has intrinsic spontaneity and indeterminacy (Brembs, 2011; Maye, Hsieh, Sugihara, & Brembs, 2007), and could often be based on internal predictive models involving the subjective state (Dickinson, 2011, 2012) even in animals with simple nervous systems (Dyer, 2012; Giurfa, 2013; Haberkern & Jayaraman, 2016; Loukola, Perry, Coscos, & Chittka, 2017; Menzel & Giurfa, 2001; Srinivasan, 2010; Webb, 2012). For example, recent evidence indicates that even seemingly hard-wired phototaxis in *Drosophila* involves quite complex value-driven computations dependent on the internal state of the animal (Gorostiza, Colomb, & Brembs, 2016). This suggests that environmental stimuli “modulate” rather than “elicit” behavioral actions, and that perception is an active process, as emphasized by the Bayesian active inference framework (Clark, 2015; Pezzulo, Rigoli, & Friston, 2015; Seth, 2013; Seth & Friston, 2016). AHA is consistent with such an “agentic” perspective. Indeed, the behaviors of AHA organisms is generated endogenously as the survival circuits trade priority internally, and are modulated (rather than fully caused) by the environmental stimuli.

Internal architecture

A major problem of naturalistic behavior in a stochastic environment is the combinatorial explosion (Bryson, 2000; Railsback & Harvey, 2013): the agent cannot recalculate every possibility at any time and environmental condition, rather sensible architectural constraints on perception, decision making and action must be imposed (Bossaerts & Murawski, 2017). Mapping unconstrained stimuli to a specific response does not scale up to realistic environments and quickly reaches an almost intractable computational complexity, thereby making strictly bottom-up “behaviorist” approach untenable (Tsotsos, 1995). More generally, the frame problem of the computational theory of mind illustrates how non-modular, unencapsulated processes are computationally infeasible (Shanahan & Baars, 2005). A surprisingly similar emphasis on the importance of internal architecture and phenotypic constraints can be found in the study of behavioral ecology and evolution beyond the “phenotypic gambit” (Andersen et al., 2016; Giske et al., 2013, 2014; Rubin, 2016; Springer, Crespi, & Swanson, 2011). The cognitive architecture in AHA mechanistically implements active perception and internal subjective evaluation thereby providing a unifying framework for the study of cognition and behavior in an ecological and evolutionary perspective. Furthermore, it emphasizes the importance of subjective phenomena in behavior of even simple organisms.

Subjective phenomena

Subjective phenomena are internal processes and states of the organism that exist from the first person point of view: their existence is inseparable from and cannot be defined independently of the experiencing subject (Searle, 2002; Zahavi, 2005). Subjective phenomena have long been an anathema in psychology and neuroscience, especially when non-human species were concerned (Baum, 2005). Even though the study of subjective phenomena have been revitalized in psychology and neuroscience (e.g. Barrett, Mesquita, Ochsner, & Gross, 2007; Searle, 2000), they are largely ignored in animal behavior, behavioral ecology and evolution. However, we are now in a position to recognize the importance of subjective phenomena not only in humans, but also

in many non-human species (Baars, 2005; Boly et al., 2013; Butler & Cotterill, 2006; Dawkins, 2015; Edelman & Seth, 2009; Gallup, 1985; Griffin, 1993; Rogers, 1997; Schönfeld, 2006) and even implement them in artificial agents (Goto & Hayashi, 2008; Reggia, 2013).

Elementary self-awareness that builds on reuse of existing neural circuits for simple forms of internal reentrant simulation provides the simplest foundation for subjective experience that can be found in organisms with more complex nervous systems. For example, consciousness is thought to build on reentrant neural activations (Baars, 1995; Edelman, 2003, 2004). Thus, AHA provides the starting point for further models that implement the main properties of conscious experiences such as qualia with unity and continuity (Searle, 2000, 2002). In doing so, it links elementary building blocks thought to produce subjective experiences with fitness and evolutionary adaptation of the integrated phenotype.

Acknowledgements

We thank Bernard Baars, Victoria Braithwaite, Tore S. Kristiansen and Marc Mangel for very helpful discussions. This research is supported by the Research Council of Norway project No. FRIMEDBIO 239834.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.bica.2018.07.009>.

References

- Aksnes, D. L., & Utne, A. C. W. (1997). A revised model of visual range in fish. *Sarsia*, 82, 137–147.
- Albantakis, L., Hintze, A., Koch, C., Adami, C., & Tononi, G. (2014). Evolution of integrated causal structures in animats exposed to environments of increasing complexity. *PLoS Computational Biology*, 10, e1003966.
- Ames, A. (2000). CNS energy metabolism as related to function. *Brain Research Reviews*, 34, 42–68.
- Andersen, B. S., Jørgensen, C., Eliassen, S., & Giske, J. (2016). The proximate architecture for decision-making in fish. *Fish and Fisheries*, 17, 680–695.
- Anderson, J. R. (1983). *The architecture of cognition*. Harvard, MA: Harvard University Press.
- Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford: Oxford University Press.
- Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences*, 33, 245–266.
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, 111, 1036–1060.
- Anderson, J. R., & Lebiere, C. (2003). The Newell Test for a theory of cognition. *Behavioral and Brain Sciences*, 26, 587–640.
- Arkin, R. C. (1998). *Behavior-based robotics*. Cambridge, MA: MIT Press.
- Arnold, S. J. (1992). Constraints on phenotypic evolution. *American Naturalist*, 140, 85–107.
- Baars, B. J. (1995). *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.
- Baars, B. J. (2005). Subjective experience is probably not limited to humans: The evidence from neurobiology and behavior. *Consciousness and Cognition*, 14, 7–21.
- Banzhaf, W., & Eeckman, F. H. (Eds.). (1995). *Evolution and biocomputation; Computational Models of Evolution*. Berlin: Springer.
- Barnard, C. (2004). *Animal behaviour: Mechanism, development, function and evolution*. Harlow, England: Prentice Hall.
- Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. *Annual Review of Psychology*, 58, 373–403.
- Barron, A. B., & Klein, C. (2016). What insects can tell us about the origins of consciousness. *Proceedings of the National Academy of Sciences*, 113, 4900–4908.
- Bartumeus, F., & Catalan, J. (2009). Optimal search behavior and classic foraging theory. *Journal of Physics A: Mathematical and Theoretical*, 42, 424003.
- Baum, W. M. (2005). *Understanding behaviorism. Behavior, culture, and evolution*. Oxford: Blackwell.
- Bertolero, M. A., Yeo, B. T. T., & D'Esposito, M. (2015). The modular and integrative functional architecture of the human brain. *Proceedings of the National Academy of Sciences*, 112, E6798–E6807.
- Boly, M., Seth, A. K., Wilke, M., Ingmundson, P., Baars, B., Laureys, S., ... Tsuchiya, N. (2013). Consciousness in humans and non-human animals: Recent advances and future directions. *Frontiers in Psychology*, 4, 1–20.
- Bossaerts, P., & Murawski, C. (2017). Computational complexity and human decision-making. *Trends in Cognitive Sciences*, 21(12), 917–929.
- Brembs, B. (2011). Towards a scientific concept of free will as a biological trait: Spontaneous actions and decision-making in invertebrates. *Proceedings of the Royal Society B: Biological Sciences*, 278, 930–939.
- Brette, R., Rudolph, M., Carnevale, T., Hines, M., Beeman, D., Bower, J. M., ... Destexhe, A. (2007). Simulation of networks of spiking neurons: A review of tools and strategies. *Journal of Computational Neuroscience*, 23, 349–398.
- Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, 47, 139–159.
- Brozović, M., Abbott, L. F., & Andersen, R. A. (2008). Mechanism of gain modulation at single neuron and network levels. *Journal of Computational Neuroscience*, 25, 158–168.
- Bryson, J. (2000). Cross-paradigm analysis of autonomous agent architecture. *Journal of Experimental and Theoretical Artificial Intelligence*, 12, 165–189.
- Budaev, S. (2018). sbudaev/AHA-R1: The AHA Model: Reference Model R1, release 1.0. zenodo.org. doi: <https://doi.org/10.5281/zenodo.1215790>.
- Budaev, S. V., & Brown, C. (2011). Personality traits and behaviour. In C. Brown, K. Laland, & J. Krause (Eds.). *Fish cognition and behavior* (pp. 135–165). Cambridge, UK: Blackwell.
- Butler, A. B., & Cotterill, R. M. J. (2006). Mammalian and avian neuroanatomy and the question of consciousness in birds. *The Biological Bulletin*, 211, 106–127.
- Christoffersen, G. R. J. (1997). Habituation: Events in the history of its characterization and linkage to synaptic depression. A new proposed kinetic criterion for its identification. *Progress in Neurobiology*, 53, 45–66.
- Churchland, P. S., & Sejnowski, T. J. (1992). *The computational brain*. Cambridge, Massachusetts: The MIT Press.
- Cisek, P., & Pastor-Bernier, A. (2014). On the challenges and mechanisms of embodied decisions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369, 20130479.
- Clark, A. (2015). Embodied prediction. *Open MIND*, 7, 1–21.
- Clune, J., Mouret, J.-B. J.-B., & Lipson, H. (2013). The evolutionary origins of modularity. *Proceedings of the Royal Society B: Biological Sciences*, 280, 20122863.
- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2, 303–314.
- Dawkins, R. (1986). *The blind watchmaker*. New York: Norton.
- Dawkins, M. (2015). Animal welfare and the paradox of animal consciousness. *Advances in the Study of Behavior*, 47, 5–38.
- Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 308, 67–78.
- Dickinson, A. (2011). Goal-directed behavior and future planning in animals. In y. R. Menzel, & J. Fischer (Eds.). *Animal thinking. Contemporary issues in comparative cognition* (pp. 79–92). Cambridge, MA: MIT Press.
- Dickinson, A. (2012). Associative learning and animal cognition. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 367, 2733–2742.
- Dingemans, N. J., Kazem, A. J. N., Réale, D., & Wright, J. (2010). Behavioural reaction norms: Animal personality meets individual plasticity. *Trends in Ecology & Evolution*, 25, 81–89.
- Duch, W., Oentaryo, R., & Pasquier, M. (2008). Cognitive architectures: Where do we go from here? *Proceedings of the 2008 conference on Artificial General Intelligence 2008: Proceedings of the First AGI Conference* (pp. 122–136). Clifton, VA: IOS Press.
- Duckworth, R. A. (2008). The role of behavior in evolution: A search for mechanism. *Evolutionary Ecology*, 23, 513–531.
- Dyer, A. G. (2012). The mysterious cognitive abilities of bees: Why models of visual processing need to consider experience and individual differences in animal performance. *Journal of Experimental Biology*, 215, 387–395.
- Edelman, G. M. (2003). Naturalizing consciousness: A theoretical framework. *Proceedings of the National Academy of Sciences*, 100, 5520–5524.
- Edelman, G. M. (2004). *Wider than the sky. The phenomenal gift of consciousness*. New Haven: Yale University Press.
- Edelman, S. (2016). The minority report: Some common assumptions to reconsider in the modelling of the brain and behaviour. *Journal of Experimental and Theoretical Artificial Intelligence*, 28, 751–776.
- Edelman, D. B., & Seth, A. K. (2009). Animal consciousness: A synthetic approach. *Trends in Neurosciences*, 32, 476–484.
- Eliassen, S., Andersen, B. S., Jørgensen, C., & Giske, J. (2016). From sensing to emergent adaptations: Modelling the proximate architecture for decision-making. *Ecological Modelling*, 326, 90–100.
- Emlen, J. M. (1966). The role of time and energy in food preference. *American Naturalist*, 100, 611–617.
- Fawcett, T. W., Hamblin, S., & Giraldeau, L.-A. (2013). Exposing the behavioral gambit: The evolution of learning and decision rules. *Behavioral Ecology*, 24, 2–11.
- Fisher, R. A. (1930). *The genetical theory of natural selection*. London: Clarendon Press.
- Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Gallup, G. G. J. (1985). Do minds exist in species other than our own? *Neuroscience and Biobehavioral Reviews*, 9, 631–641.
- Giske, J., Eliassen, S., Fiksen, Ø., Jakobsen, P. J., Aksnes, D. L., Jørgensen, C., & Mangel, M. (2013). Effects of the emotion system on adaptive behavior. *The American Naturalist*, 182, 689–703.
- Giske, J., Eliassen, S., Fiksen, Ø., Jakobsen, P. J., Aksnes, D. L., Mangel, M., & Jørgensen, C. (2014). The emotion system promotes diversity and evolvability. *Proceedings of the Royal Society B: Biological Sciences*, 281, 20141096.
- Giurfa, M. (2013). Cognition with few neurons: Higher-order learning in insects. *Trends in Neurosciences*, 36, 285–294.
- Gollisch, T., & Herz, A. V. M. (2012). The iso-response method: Measuring neuronal stimulus integration with closed-loop experiments. *Frontiers in Neural Circuits*, 6, 1–14.
- Gorostiza, E. A., Colomb, J., & Brembs, B. (2016). A decision underlies phototaxis in an

- insect. *Open Biology*, 6, 160,229.
- Gosling, S. D. (1999). Personality dimensions in nonhuman animals: A cross-species review. *Current Directions in Psychological Science*, 8, 69–75.
- Goto, N., & Hayashi, E. (2008). Design of robotic behavior that imitates animal consciousness. *Artificial Life and Robotics*, 12(1–2), 97–101.
- Grafen, A. (1984). Natural selection, kin selection and group selection. (2nd ed.). *Behavioural ecology: An evolutionary approach*(pp. 62–84). Oxford, UK: Blackwell.
- Griffin, D. R. (1993). *Animal minds*. Chicago, Ill.: University of Chicago Press.
- Haberkern, H., & Jayaraman, V. (2016). Studying small brains to understand the building blocks of cognition. *Current Opinion in Neurobiology*, 37, 59–65.
- Heisenberg, M. (2014). The beauty of the network in the brain and the origin of the mind in the control of behavior. *Journal of Neurogenetics*, 28, 389–399.
- Hogan, J. A. (2009). Causation: The study of behavioral mechanisms. In J. J. Bolhuis, & S. Verhulst (Eds.). *Tinbergen's Legacy. Function and Mechanism in Behavioral Biology*(pp. 35–53). Cambridge: Cambridge University Press.
- Hogan, J. A. (2015). A framework for the study of behavior. *Behavioural Processes*, 117, 105–113.
- Houston, A., Clark, C., McNamara, J., & Mangel, M. (1988). Dynamic models in behavioural and evolutionary ecology. *Nature*, 332, 29–34.
- Huse, G., & Giske, J. (1998). Ecology in Mare Pentium: An individual-based spatio-temporal model for fish with adapted behaviour. *Fisheries Research*, 37, 163–178.
- Huse, G., Strand, E., & Giske, J. (1999). Implementing behaviour in individual-based models using neural networks and genetic algorithms. *Evolutionary Ecology*, 13, 469–483.
- Ioannou, P. A., & Pitsillides, A. (2008). *Modeling and control of complex systems*. Boca Raton: CRC Press.
- Irlbacher, K., Kraft, A., Kehler, S., & Brandt, S. A. (2014). Mechanisms and neuronal networks involved in reactive and proactive cognitive control of interference in working memory. *Neuroscience and Biobehavioral Reviews*, 46, 58–70.
- Kakade, S., & Dayan, P. (2002). Dopamine: Generalization and bonuses. *Neural Networks*, 15, 549–559.
- Landsrød, L. K. (2017). *Decision-making in a proximate model framework: How behaviour flexibility is generated by arousal and attention*. MSc Thesis. University of Bergen <http://bora.uib.no/handle/1956/16385>.
- LeDoux, J. E. (2012). Rethinking the emotional brain. *Neuron*, 73, 653–676.
- Lewis, M., & Canamero, L. (2016). Hedonic quality or reward? A study of basic pleasure in homeostasis and decision making of a motivated autonomous robot. *Adaptive Behavior*, 24, 267–291.
- Lewontin, R. C. (1974). *Genetic basis of evolutionary change*. New York: Columbia University Press.
- Lorenz, D. M., Jeng, A., & Deem, M. W. (2011). The emergence of modularity in biological systems. *Physics of Life Reviews*, 8, 129–160.
- Loukola, O. J., Perry, C. J., Coscos, L., & Chittka, L. (2017). Bumblebees show cognitive flexibility by improving on an observed complex behavior. *Science*, 355, 833–836.
- Lucentini, D. F., & Gudwin, R. R. (2015). A comparison among cognitive architectures: A theoretical analysis. *Procedia Computer Science*, 71, 56–61.
- Lytton, W. W. (2002). *From computer to brain: foundations of computational neuroscience*. New York: Springer.
- Maye, A., Hsieh, C. H., Sugihara, G., & Brembs, B. (2007). Order in spontaneous behavior. *PLoS One*, 2, e443.
- McFarland, D. J., & Sibly, R. M. (1975). The behavioural final common path. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 270, 265–293.
- McGlothlin, J. W., & Ketterson, E. D. (2008). Hormone-mediated suites as adaptations and evolutionary constraints. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363, 1611–1620.
- Menzel, R., & Giurfa, M. (2001). Cognitive architecture of a mini-brain: The honeybee. *Trends in Cognitive Sciences*, 5, 62–71.
- Meunier, D., Lambiotte, R., & Bullmore, E. T. (2010). Modular and hierarchically modular organization of brain networks. *Frontiers in Neuroscience*, 4, 1–11.
- Murren, C. J. (2012). The integrated phenotype. *Integrative and Comparative Biology*, 52, 64–76.
- Mysore, S. P., & Knudsen, E. I. (2012). Reciprocal inhibition of inhibition: A circuit motif for flexible categorization in stimulus selection. *Neuron*, 73, 193–205.
- Newell, A. (1994). *Unified theories of cognition (William James Lectures)*. Harvard, MA: Harvard University Press.
- Niven, J. E. (2016). Neuronal energy consumption: biophysics, efficiency and evolution. *Current Opinion in Neurobiology*, 41, 129–135.
- Oudeyer, P.-Y., Kaplan, F., & Hafner, V. V. (2007). Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, 11, 265–286.
- Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active Inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, 134, 17–35.
- Pierce, G. J., & Ollason, J. G. (1987). Eight reasons why optimal foraging theory is a complete waste of time. *Oikos*, 49, 111–117.
- Rabaglia, C. D., Marcus, G. F., & Lane, S. P. (2011). What can individual differences tell us about the specialization of function? *Cognitive Neuropsychology*, 28, 288–303.
- Railsback, S. F., & Harvey, B. C. (2013). Trait-mediated trophic interactions: Is foraging theory keeping up? *Trends in Ecology and Evolution*, 28, 119–125.
- Reale, D., Dingemans, N. J., Kazem, A. J. N., & Wright, J. (2010). Evolutionary and ecological approaches to the study of personality. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 365, 3937–3946.
- Reggia, J. A. (2013). The rise of machine consciousness: Studying consciousness with computational models. *Neural Networks*, 44, 112–131.
- Reynolds, A. (2015). Liberating Lévy walk research from the shackles of optimal foraging. *Physics of Life Reviews*, 14, 59–83.
- Reynolds, A. M. (2018). Current status and future directions of Levy walk research. *Biology Open*, 7, 1–6.
- Rogers, L. J. (1997). *Minds of their own. Thinking, and awareness in animals*. St. Leonards, NSW: Allen & Unwin.
- Rubin, H. (2016). The phenotypic gambit: selective pressures and ESS methodology in evolutionary game theory. *Biology and Philosophy*, 31, 551–569.
- Ruxton, G. D., & Beauchamp, G. (2008). The application of genetic algorithms in behavioural ecology, illustrated with a model of anti-predator vigilance. *Journal of Theoretical Biology*, 250, 435–448.
- Samsonovich, A. V. (2012). On a roadmap for the BICA Challenge. *Biologically Inspired Cognitive Architectures*, 1, 100–107.
- Schönfeld, M. (2006). Animal consciousness: Paradigm change in the life sciences. *Perspectives on Science*, 14, 354–381.
- Searle, J. R. (2000). Consciousness. *Annual Review of Neuroscience*, 23, 557–578.
- Searle, J. R. (2002). *Consciousness and language*. Cambridge: Cambridge University Press.
- Seth, A. K. (2007). The ecology of action selection: insights from artificial life. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362, 1545–1558.
- Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17, 565–573.
- Seth, A. K., & Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371, 20,160,007.
- Shanahan, M., & Baars, B. (2005). Applying global workspace theory to the frame problem. *Cognition*, 98, 157–176.
- Sih, A., Bell, A. M., Johnson, J. C., & Ziemba, R. E. (2004). Behavioral syndromes: an integrative overview. *The Quarterly Review of Biology*, 79, 241–277.
- Springer, S. A., Crespi, B. J., & Swanson, W. J. (2011). Beyond the phenotypic gambit: Molecular behavioural ecology and the evolution of genetic architecture. *Molecular Ecology*, 20, 2240–2257.
- Srinivasan, M. V. (2010). Honey bees as a model for vision, perception, and cognition. *Annual Review of Entomology*, 55, 267–284.
- Sternberg, S. (2011). Modular processes in mind and brain. *Cognitive Neuropsychology*, 28, 156–208.
- Strand, E., Huse, G., & Giske, J. (2002). Artificial evolution of life history and behavior. *The American Naturalist*, 159, 624–644.
- Sun, R. (2004). Desiderata for cognitive architectures. *Philosophical Psychology*, 17, 341–373.
- Sun, R. (2009). Theoretical status of computational cognitive modeling. *Cognitive Systems Research*, 10, 124–140.
- Szabados, J., & Vértesi, P. (1990). *Interpolation of functions*. Singapore: World Scientific.
- Toates, F. (1997). The interaction of cognitive and stimulus–response processes in the control of behaviour. *Neuroscience and Biobehavioral Reviews*, 22, 59–83.
- Toates, F. (1998). *Control of behavior*. Berlin: Springer.
- Tsotsos, J. K. (1995). Behaviorist intelligence and the scaling problem. *Artificial Intelligence*, 75, 135–160.
- Tyrell, T. (1993). The use of hierarchies for action selection. *Adaptive Behavior*, 1, 387–420.
- Vaughan, S. L., Mills, R. F., Peterson, G. L., Grimaila, M. R., Rogers, S. K., Oxley, M. E., & Patterson, R. E. (2016). A dual-process Qualia Modeling Framework (QMF). *Biologically Inspired Cognitive Architectures*, 17, 71–85.
- Vernon, D., Von Hofsten, C., & Fadiga, L. (2016). Desiderata for developmental cognitive architectures. *Biologically Inspired Cognitive Architectures*, 18, 116–127.
- Wagner, G. P., Pavlicev, M., & Cheverud, J. M. (2007). The road to modularity. *Nature Reviews Genetics*, 8, 921–931.
- Webb, B. (2012). Cognition in insects. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 367, 2715–2722.
- West-Eberhard, M. J. (1989). Phenotypic plasticity and the origins of diversity. *Annual Review of Ecology and Systematics*, 20, 249–278.
- Wilson, D. S., Clark, A., Coleman, K., & Dearstyne, T. (1994). Shyness and boldness in humans and other animals. *Trends in Ecology & Evolution*, 9, 442–446.
- Wosniack, M. E., Santos, M. C., Raposo, E. P., Viswanathan, G. M., & da Luz, M. G. E. (2017). The evolutionary origins of Lévy walk foraging. *PLoS Computational Biology*, 13, e1005774.
- Zahavi, D. (2005). *Subjectivity and selfhood. Investigating the first person perspective*. Cambridge, MA: Bradford Books.