Behavior Morphologies of Machine Learning Agents in Media Artworks

SOFIAN AUDRY

Since the 1950s, a range of artists have used artificial agents in their work, in parallel with scientific research in cybernetics, artificial intelligence (AI) and artificial life (AL). In particular, an increasing number of artists work with machine learning and other adaptive systems. Through the author's own engagement with such systems, an analysis of adaptive agents is provided within the broader context of behavior aesthetics. As a result, the author proposes an aesthetic framework for understanding behaviors that accounts for the observer as an adaptive perceiving agent, the unfathomable character of machine learning systems and the morphology of behaviors as a time-based phenomenon.

Contemporary media artworks have employed artificial agents since the early days of computing technologies, including pioneering cybernetic pieces such as Nicholas Schöffer’s CYSP 1 (1956) and Gordon Pask’s Colloquy of Mobiles (1968), artificial life systems like Karl Sims’s Galápagos (1997–2000) and robotic works such as Simon Penny’s Petit Mal (1993) and Driessens and Verstappen’s Tickle (1997). These works do not use computers to produce images and sound; the locus of their aesthetic effect is the real-time behavior of an artificial system as it unfolds in the world. This article focuses on a particular facet of this broader work: agent-based adaptive media artworks, such as Nicolas Baginsky’s self-organizing robotic band The Three Sirens (1992–2005); the curiosity-driven learning robots in Zwischenräume (2014) by Petra Gemeinboeck and Rob Saunders; and the performance/installation N-Polytope: Behaviors in Light and Sound after Iannis Xenakis (2012), by Chris Salter in collaboration with myself, Marije Baalman, Adam Basanta, Elio Bidinost and architect Thomas Spier, in which behavioral patterns of light and sound agents adapt temporally and spatially in real time.

FEELING BEHAVIORS

Human observers of such agent-based artworks interpret the behavioral patterns they generate in ways that are highly context-dependent and subjective. When my collaborators Samuel St-Aubin and Stephen Kelly and I show our work Vessels (2015) (Fig. 1), a robotic installation consisting of groups of water-dwelling robots that adapt to one another and to their environment by evolving their behavior, conversations with passersby often start with a question: “How does it work?” This initial curiosity about the functional dimension of the work reveals cultural preconceptions about technology that feed anxiety. People presuppose that robots are programmed using sets of rationally explainable rules.

I have learned to avoid answering this question by turning it over to the asker: “How do you think it works?” In response, people usually come up with stories that have nothing to do with programming but rather with social and emotional subjects. “Look at these robots: Are they fighting?” “I think they’re making out!” “This one likes to be alone.” Such acts of projection underpin what Simon Penny calls the aesthetics of behavior, an artistic field allowed by the “possibility of cultural interaction with machine systems” [1]. The study of enactive couplings that emerge through the interactions
between audience members and behavioral artifacts is, indeed, a critical topic in contemporary media art studies [2,3].

Such speculations by observers of artificial agents’ demeanors are only one piece of the puzzle. To the question “How does it work?”, I am left to answer that, while I could describe the technical principles with which my collaborators and I built the robots, these explanations would not tell us much about their behaviors as we experience them. There is a gap between the inner workings of a system and how the system’s behavior is perceived by an external observer—including its author.

The projected impressions of “lifelikeness” observed in works such as Vessels are hence underpinned by the unfathomable character of behaviors generated by the adaptive agents in place. The affective perception of behavioral forms that cannot be described logically echoes the experience of other artists working with artificial agents. For example, about his work Eden, in which an ecosystem of artificial audiovisual agents react indirectly to visitors, Jon McCormack notices that, while visitors do not understand the inner functioning of the systems, they describe their experience of the work as “having a sense that it is somehow alive” [4].

This is not surprising, considering how observers themselves are adaptive systems trying to make sense of often complex behaviors. Such agent-based artworks need a particular context to be correctly apprehended: They cannot simply be experienced in the same way as more traditional reactive and/or interactive media artworks. Perhaps more than with nonadaptive media works, one needs to spend time with these agents to get to know them—in other words, to adapt to them phenomenologically. Indeed, it has been my own experience with works such as Vessels that audiences who allow themselves to spend more time with these works tend to enjoy them more.

Through the design of synthetic agents governed by algorithmic processes, behaviors become an artistic medium of their own. A defining property of behaviors, as opposed to other media such as photo or video, is that they can run for an infinite amount of time while never exactly repeating themselves. Yet despite the inexhaustible nature of their manifestations, they can still be recognizable by a human observer as a definite “thing.”

THE NATURE OF ADAPTIVE BEHAVIORS
How do the behaviors of adaptive/learning agents differ from those produced by other kinds of systems? Machine learning systems are not “explicitly programmed” [5] but rather make up their own knowledge by learning from their sensory data, through a process of trial and error. Indeed, machine learning claims its superiority over rule-based systems precisely because learning algorithms are allowed to self-organize in ways so complex that they cannot be explained using simple logic.

The incommensurability of machine learning systems and of other self-organizing systems distinguishes them from nonadaptive and nonemergent systems, allowing them to produce different behavioral forms. Cybernetician Gordon Pask defines a behavior as an “unchanging form of events” recognized by an external observer [6]. This perceived stability in the “shape” of an agent’s activity comes after a certain period of time that depends on the complexity of the agent producing the behavior and the speed at which it acts, opposed to the capacity of the observer to make sense of that activity in order to perceive the behavior’s morphology. In other words, the observers of a low-complexity system might be able to “get” the behavior in a “how-does-it-work” fashion after a few minutes of observation, whereas they might have to spend more time to “understand” the behavior of a high-complexity system. Furthermore, the nature of adaptive and self-organizing systems will result in a different “kind” of understanding that cannot be explained by logic but rather is “felt” intuitively.

I posit that different categories of system architectures allow for different kinds of behaviors. Existing taxonomies of cybernetics systems have mainly focused on relational and structural aspects of these systems [7,8]. In response, I here propose a flexible taxonomy of embodied systems that focuses on the aesthetics of agent behaviors as their shape unfolds in time.

The “zero-degree” of that categorization is whether the system possesses internal states—which I posit as a necessary condition to consider a system as being able to behave. Stateless systems such as mappings are fundamentally different from stateful systems such as agents, the former involving more-or-less-direct transformations between inputs and outputs, while the latter can change their inner structure in response to events. In other words, blurring an image or compressing a sound is not a behavior; and neither a statue nor a vocoder can be said to behave. This definition slightly raises the bar compared to Pask’s, which grants behaviors even to statues.

By design, stateless devices are incapable of accumulating experience, because their outputs/actions depend entirely on their inputs/observations. Such systems are known in the field of digital media art as mappings and are pervasive in new media art. Mostly devoid of any kind of autonomy and agency, the behavior of a mapping-based device is for the most part exogenous to the system. Whatever sense of aliveness it conveys truly lies outside of the system itself but rather in the systems that generate the inputs.

In his doctoral dissertation, artist Marc Downie attacks the “normative idea” behind mappings: that computer art is about numbers being transformed into numbers [9]. To mappings, he opposes the more fruitful concept of agents. Such stateful devices possess inner structures that are transformed by their interactions with the environment, thus allowing their past inputs to influence their future outputs.

This statefulness can be found in a wide variety of computational systems, for example, in variables and data structures. However, these syntactic components remain in the symbolic and rule-based domain, often resulting in behaviors that are closed rather than open-ended. While the program’s response to sensory data may change depending on context, it will eventually need to “loop back” to a region of the code that was designed by its author: Given enough time, it will, inexorably, come to repeat similar patterns.

I thus refer to such conducts as first-order behaviors—in
other words, behaviors whose form does not change through time (Fig. 2). One example is Erwin Driessens and Maria Verstappen’s *Tickle* (1997), a miniature robot that crawls on the human body, generating a tickling sensation. The robot has a simple behavior implemented using a Finite State Machine (FSM) that makes it navigate effectively on human skin, for example, by steering away when it encounters a steep slope.

Another example is my 2007 interactive installation *Flag*, in which visitors hold small signs in front of a camera to engage in a contemplative experience modeled using an FSM agent. The agent’s default state is one of dormancy. Upon detecting the presence of visitors, it wakes up and starts replacing the word on the signs using augmented reality, cycling through three different phases and eventually returning back to its sleeping state after the visitors leave.

This conception of behavior is insufficient to grasp the kinds of transformative processes involved in systems that change radically through time, such as those present in *Vessels*. Indeed, self-organizing agents such as adaptive and evolutionary agents can produce second-order behaviors or *meta-behaviors* that involve transitions between multiple (first-order) behaviors. They therefore exist in a different, meta-temporal domain than nonadaptive/nonevolutive devices, which affects the overall aesthetic effects they can engender.

To take a real-world example, consider how a cat’s whiskers move under the wind (mapping). Now compare the behavior of a cat preying on a bird (first-order behavior) to the behavior of a house cat throughout its life (second-order) as it evolves from playing frantically as a kitten, to learning how to respond to new conditions (Fig. 2). Typically starting from a state of pure randomness, adaptive agents run through a morphogenetic learning process whereby they progressively modify their behavior. When they reach their final form, they enter a state of morphostasis, exploiting the stabilized, learned behavior that they converged to (Fig. 3). Changes in the environment or in the adaptive agents’ objectives may trigger other metamorphic phases as the agents try to change their behavior to respond to new conditions (Fig. 4).

This convergence is noticeable in works such as *Vessels*. During an indoor exhibition of the work, a media artist who was doing a residency on the premises had the opportunity to experience the work on a daily basis. She reported that as she passed the piece on the way to the coffee machine, she became increasingly familiar with it and started noticing patterns. In particular, she explained that as a given day went by, it seemed that the behaviors became less and less random and seemed to converge and stabilize into more recognizable, interesting patterns.

**BEHAVIOR MORPHOLOGIES**

I propose to use the concepts of morphostasis, morphogenesis and metamorphosis to further characterize the different processes by which behavioral morphologies exist, emerge and/or change over time.

*Morphostasis* refers to the process whereby a behavior hovers around a stable state of being. While the behavioral patterns might appear changing when considered over a certain period of time, morphostatic behaviors quickly exhaust the space of dynamic patterns they can generate and start appearing repetitive. These behaviors are immutable: They stay constant through time.

*Morphogenesis* is the mechanism by which emergent behaviors develop their form in a continuous manner. Only adaptive and evolutionary devices, which are capable of self-organization, support morphogenetic behaviors. The category implies the production of new behavioral morphologies through a system’s interaction with the world.

*Metamorphosis* is intimately related to morphogenesis and refers to the process by which behaviors change from one shape into another. The term should be understood as it is used in common parlance: that is, as an outstanding transformation within a living being or thing. The two main dimensions of metamorphosis are (1) the *metabolys*, that is, the magnitude of the transformation undergone by the behavior; and (2) the *duration* required for the behavior to transition from one form into the next.

These aspects of an agent’s performance should be seen less as hard-set categories than as conceptual tools for describing processes of behavior formation. From this perspective, nonemergent systems such as rule-based programs, simple self-regulated devices and pretrained machine learning algorithms produce morphostatic behaviors.

At the opposite end of the spectrum, some morphogenetic systems freely move from one behavioral embodiment into another, living in a constant state of metamorphosis, as if never fully coming into being. These systems can be broadly referred to as *generative nonadaptive*: They evolve behaviors through time in a nonpurposeful way, as they have no objective relevance criteria such as a fitness or cost function [11].

Adaptive systems, on the other hand, evolve their morphologies in relationship to an “optimal” behavior or process that they try to approach and match. In this, they differ from nonadaptive second-order behaviors. Adaptive systems are intentional and relational devices. They do not simply adapt, they adapt to something.

Typically starting from a state of pure randomness, adaptive agents run through a morphogenetic learning process whereby they progressively modify their behavior. When they reach their final form, they enter a state of morphostasis, exploiting the stabilized, learned behavior that they converged to (Fig. 3). Changes in the environment or in the adaptive agents’ objectives may trigger other metamorphic phases as the agents try to change their behavior to respond to new conditions (Fig. 4).

This convergence is noticeable in works such as *Vessels*. During an indoor exhibition of the work, a media artist who was doing a residency on the premises had the opportunity to experience the work on a daily basis. She reported that as she passed the piece on the way to the coffee machine, she became increasingly familiar with it and started noticing patterns. In particular, she explained that as a given day went by, it seemed that the behaviors became less and less random and seemed to converge and stabilize into more recognizable, interesting patterns.

---

Fig. 2. Example temporal evolution of first-order behaviors. The vertical axis represents the behavior of the system, understood as the temporally invariant shape of observable events the system generates. Distance along the vertical axis represents difference in the form of observable events produced by the agent. Recognizable behavior morphologies are represented using labels A, B, C and D. The horizontal axis represents the advance of time. The graphic shows how first-order behaviors remain temporally stable: (1) is unchanging while (2) fluctuates within the boundaries of a morphological kind. (© Sofian Audry)
Artist Nicolas Baginsky, who spent years touring with his machine learning robotic band *The Three Sirens* (1992–2005) (Fig. 5), compares this process to that of aging. In their infancy, the robots tended to play very chaotically. As they moved into adolescence and adulthood, they would start to develop their own melodic structures while exploring interesting patterns in forms akin to jazz improv. Finally, as they grew older, they became more conservative, converging into more static forms of musical expression [12].

The aesthetic experience of these behaviors is dependent on a number of factors. The ratio between the magnitude of change (metaboly) and the time period (duration) necessary to perform the change during metamorphosis—which in the case of machine learning systems is directly related to the learning rate—can be used as a measure of intensity. Abrupt, fast changes can bring a sense of astonishment or angst in the viewer that artists working with interactive media have learned to exploit. In contrast, longer yet steady and noticeable changes can evoke curiosity, anxiety and uncanny sensations. For example, in *Vessels*, the robots are always in a state of flux, which might explain feelings of estrangement described by some members of the audience.

Finally, adaptive behaviors contain a narrative of adaptation and learning. As they unfold before our eyes, we perceive fluctuating stories of trials and errors, of successes and failures, that evoke our own experiences of learning as fallible and imperfect.

The categorization proposed here is not meant to be a systematic classification scheme but rather a frame of reference, a flexible analysis tool provided with the hope that it can be useful to artists and theorists. It provides a perspective, an angle, a way to think about and discuss agent-based systems that accommodates the different kinds of behavior-generating processes found in new media art practice (Fig. 6).
CONCLUSION

This article describes adaptive behaviors generated by machine learning agents in the context of agent-based art and behavior aesthetics, focusing on the morphological aspects of behaviors as they unfold in time. It shows how human observers of self-organizing works such as Vessels and The Three Sirens cannot understand their behaviors rationally since the underlying processes that govern them follow non-logical rules. In contrast, works that are based on mappings and first-order behaviors have the potential to be rationally understood, at least in theory: For example, this photocell triggers that sound effect, such gestures cause such robot to start running in circles, etc. But in order to experience second-order behaviors in all their richness, one needs to “get to know them” phenomenologically, through one’s own body. One needs to adapt, to change oneself until a behavior becomes physically felt.

This necessity is a direct consequence of two important features of adaptive systems. First, the way by which their morphology evolves in time, moving through periods of morphogenesis, metamorphosis and morphostasis, may contribute to their mystifying nature. During their transitive phases, such behaviors lie in a state of flux, making them difficult to grasp. Second, as they increasingly become better at performing the task they are trained for, machine learning models grow into complex, often unintelligible architectures. This is particularly true of artificial neural networks, where large numbers of independent neurons work together, making it near impossible for a human observer to find out which neuron is responsible for what.

For machine learning agents to learn complex interactions, they need huge amounts of information, which requires long periods of time so that they can have access to many different experiences. This often results in long-form, evolution-centric works such as The Three Sirens, Zwischenräume and Vessels, which give value to the durational and contemplative dimensions of the experience, making them akin to performance art practices. Artists working with such systems need to provide appropriate contexts to support the natural unfolding of these new behavioral forms over hours, weeks or even years, allowing audiences to develop a relationship with these systems so that they can get to know them.

Acknowledgments

I thank the Fonds de Recherche du Québec—Société et Culture and Chris Salter for their help and support in carrying out this research. My sincere thanks to Jon Ippolito and Simon Penny for their precious feedback.

References and Notes


Manuscript received 1 October 2018.

SOFIAN AUDRY is an artist, scholar and professor of Inter- active Media within the School of Media at University of Quebec in Montreal (UQAM), working at the intersection between computer art and artificial intelligence.